Experimental and Data-driven Workflows for Microstructure-based Damage Prediction

Zur Erlangung des akademischen Grades eines
DOKTORS DER INGENIEURWISSENSCHAFTEN (Dr.-Ing.)
von der KIT-Fakultät für Maschinenbau des
Karlsruher Instituts für Technologie (KIT)

genehmigte
DISSERTATION
von

M.Sc. Ali Riza Durmaz

Tag der mündlichen Prüfung:	31. Januar 2022
Referent:	Prof. Dr. Peter Gumbsch
Korreferent:	Prof. Dr. Christoph Eberl
Korreferent:	Prof. Dr. Christoph Kirchlechner

Kurzfassung

Materialermüdung ist die häufigste Ursache für mechanisches Versagen. Die Degradationsmechanismen, welche die Lebensdauer von Bauteilen bei vergleichsweise ausgeprägten zyklischen Belastungen bestimmen, sind gut bekannt. Bei Belastungen im makroskopisch elastischen Bereich hingegen, der (sehr) hochzyklischen Ermüdung, bestimmen die innere Struktur eines Werkstoffs und die Wechselwirkung kristallografischer Defekte die Lebensdauer. Unter diesen Umständen sind die inneren Degradationsphänomene auf der mikroskopischen Skala weitgehend reversibel und führen nicht zur Bildung kritischer Schädigungen, die kontinuierlich wachsen können. Allerdings sind einige Kornensembles in polykristallinen Metallen, je nach den lokalen mikrostrukturellen Gegebenheiten, anfällig für Schädigungsinitiierung, Rissbildung und -wachstum und wirken daher als Schwachstellen. Daher weisen Bauteile, die solchen Belastungen ausgesetzt sind, oft eine ausgeprägte Lebensdauerstreuung auf. Die Tatsache, dass ein umfassendes mechanistisches Verständnis für diese Degradationsprozesse in verschiedenen Werkstoffen nicht vorliegt, hat zur Folge, dass die derzeitigen Modellierungsbemühungen die mittlere Lebensdauer und ihre Varianz in der Regel nur mit unbefriedigender Genauigkeit vorhersagen. Dies wiederum erschwert die Bauteilauslegung und macht die Nutzung von Sicherheitsfaktoren während des Dimensionierungsprozesses erforderlich.

Abhilfe kann geschaffen werden, indem umfangreiche Daten zu Einflussfaktoren und deren Wirkung auf die Bildung initialer Ermüdungsschädigungen erhoben werden. Die Datenknappheit wirkt sich nach wie vor negativ auf Datenwissenschaftler und Modellierungsexperten aus, die versuchen, trotz geringer Stichprobengröße und unvollständigen Merkmalsräumen, mikrostrukturelle Abhängigkeiten abzuleiten, datengetriebene Vorhersagemodelle zu trainieren oder physikalische, regelbasierte Modelle zu parametrisieren. Die Tatsache, dass nur wenige kritische Schädigungen bezogen auf das gesamte Probenvolumen auftreten und die hochzyklische Ermüdung eine Vielzahl unterschiedlicher Abhängigkeiten aufweist, impliziert einige Anforderungen an die Datenerfassung und -verarbeitung. Am wichtigsten ist, dass die Messtechniken so empfindlich sind, dass nuancierte Schwankungen im Probenzustand erfasst werden können, dass die gesamte Routine effizient ist und dass die korrelative Mikroskopie räumliche Informationen aus verschiedenen Messungen miteinander verbindet.

Das Hauptziel dieser Arbeit besteht darin, einen Workflow zu etablieren, der den Datenmangel behebt, so dass die zukünftige virtuelle Auslegung von Komponenten effizienter, zuverlässiger und nachhaltiger gestaltet werden kann. Zu diesem Zweck wird in dieser Arbeit ein kombinierter experimenteller und datenverarbeitender Workflow vorgeschlagen, um multimodale Datensätze zu Ermüdungsschädigungen zu erzeugen. Der Schwerpunkt liegt dabei auf dem Auftreten von lokalen Gleitbändern, der Rissinitiierung und dem Wachstum mikrostrukturell kurzer Risse. Der Workflow vereint die Ermüdungsprüfung von mesoskaligen Proben, um die Empfindlichkeit der Schädigungsdetektion zu erhöhen, die ergänzende Charakterisierung, die multimodale Registrierung und Datenfusion der heterogenen Daten, sowie die bildverarbeitungsbasierte Schädigungslokalisierung und -bewertung. Mesoskalige Biegeresonanzprüfung ermöglicht das Erreichen des hochzyklischen Ermüdungszustands in vergleichsweise kurzen Zeitspannen bei gleichzeitig verbessertem Auflösungsvermögen der Schädigungsentwicklung. Je nach Komplexität der einzelnen Bildverarbeitungsaufgaben und Datenverfügbarkeit werden entweder regelbasierte Bildverarbeitungsverfahren oder Repräsentationslernen gezielt eingesetzt. So sorgt beispielsweise die semantische Segmentierung von Schädigungsstellen dafür, dass wichtige Ermüdungsmerkmale aus mikroskopischen Abbildungen extrahiert werden können. Entlang des Workflows wird auf einen hohen Automatisierungsgrad Wert gelegt. Wann immer möglich, wurde die Generalisierbarkeit einzelner Workflow-Elemente untersucht. Dieser Workflow wird auf einen ferritischen Stahl (EN 1.4003) angewendet. Der resultierende Datensatz verknüpft unter anderem große verzerrungskorrigierte Mikrostrukturdaten mit der Schädigungslokalisierung und deren zyklischer Entwicklung. Im Zuge der Arbeit wird der Datensatz wird im Hinblick auf seinen Informationsgehalt untersucht, indem detaillierte, analytische Studien zur einzelnen Schädigungsbildung durchgeführt werden. Auf diese Weise konnten unter anderem neuartige, quantitative Erkenntnisse über mikrostrukturinduzierte

plastische Verformungs- und Rissstopmechanismen gewonnen werden. Darüber hinaus werden aus dem Datensatz abgeleitete kornweise Merkmalsvektoren und binäre Schädigungskategorien verwendet, um einen Random-Forest-Klassifikator zu trainieren und dessen Vorhersagegüte zu bewerten. Der vorgeschlagene Workflow hat das Potenzial, die Grundlage für künftiges Data Mining und datengetriebene Modellierung mikrostrukturempfindlicher Ermüdung zu legen. Er erlaubt die effiziente Erhebung statistisch repräsentativer Datensätze mit gleichzeitig hohem Informationsgehalt und kann auf eine Vielzahl von Werkstoffen ausgeweitet werden.

Abstract

Materials fatigue poses the most frequent cause of mechanical failure. The degradation mechanisms that dictate component life at comparatively large cyclic loads are well known. In contrast, when loads in the macroscopically elastic range are concerned, the so-called (very) high-cycle fatigue regime, the inner structure of a material, and crystallographic defect interactions therein govern the service life. Under these circumstances, the internal degradation phenomena at the microscopic scale are largely reversible or do not culminate in the formation of critical damage instances which can undergo continued growth. However, some grain ensembles in polycrystalline metals depending on their local microstructural character are vulnerable to initial damage formation, crack initiation and growth and therefore act as weak points. Therefore, components exposed to such loads, often exhibit pronounced service life scatter. The fact that a comprehensive mechanistic understanding of these early degradation processes in various materials is yet to be unveiled, entails that current modeling efforts typically predict the mean life and its variance with unsatisfactory accuracy. This in turn hampers the component design and necessitates the application of safety factors during the dimensioning process.

Remedy can be provided by addressing the lack of data capturing initial fatigue damage evolution and its influence factors. The data shortage continues to adversely affect data scientists and computational modeling experts attempting to derive microstructural dependencies, train predictive data-driven models, or parameterize physical rule-based models from small sample size data with incomplete feature representations. The fact that only a few critical damage instances occur referred to the whole specimen volume and high-cycle fatigue is characterized by a variety of distinct dependencies, placing some demands on the data acquisition and processing routine. Most importantly, the measurement techniques should provide the sensitivity to capture nuanced fluctuations in the specimen state, the whole routine should be efficient, and correlative microscopy should link spatial information from different measurements.

The primary objective of this work is to establish a workflow to address the data shortage such that future virtual component design can be rendered more efficient, reliable, and sustainable. To this end, this work proposes a combined experimental and data post-processing workflow to generate multimodal fatigue damage data sets. Particularly, the focus is placed on the emergence of slip markings, crack initiation, and microstructurally short crack growth. The workflow unifies fatigue testing of mesoscale specimens to increase damage detection sensitivity, complementary characterization, multimodal registration and data fusion to address and exploit data heterogeneity, and image processing-based damage localization and evaluation. Mesoscale bending resonant cyclic testing enables reaching the high-cycle fatigue regime in comparatively short time spans while maintaining unprecedented time resolution of damage evolution. Depending on the complexity of the individual computer vision tasks and data availability, either rule-based image processing or representation learning methods are purposefully incorporated. For example, semantic segmentation of damage instances ensures retrieval of important fatigue features from micrographs. Emphasis is placed on a high degree of automation. Whenever possible, the generalizability of individual workflow elements is explored. This workflow is applied to a ferritic steel (EN 1.4003) case study. The resulting data set links amongst others large distortion-corrected microstructure data with damage localization and its cyclic evolution. The data set is explored with respect to its information content by performing detailed, analytical studies on individual damage instances. This provided quantitative insights with respect to microstructural plastic deformation and crack arrest mechanisms. Furthermore, grain-wise feature vectors and binary damage categories derived from the data set are utilized to train and evaluate a random forest classifier. The proposed workflow holds the potential to lay the foundation for future data mining and data-driven modeling of microstructure-sensitive fatigue by efficiently providing statistically representative data sets with a high information content extendable to a wide range of materials.

Acknowledgements

First, I want to formulate my gratitude to Prof. Dr. Chris Eberl for inviting me into this exciting research field after supervising my master's thesis. Aside from this, I am grateful for the encouragement I received from you repeatedly in the course of this journey. In countless discussions, your thinking out of the box was rather contagious and not only pointed me in the right direction but also helped me to grow professionally and personally.

I want to express my appreciation to Prof. Dr. Peter Gumbsch who enabled this research in the first place, amongst others by acquiring the funding. Despite your involvement in numerous associations and your responsibilities as director of institutions, you managed to make time for supervising me, which I highly cherish. Your quick perception and analytical thinking rendered our discussions helpful at all times and made me critically examine my approaches.

Prof. Dr. Chris Eberl, Prof. Dr. mont. Christoph Kirchlechner, and Prof. Dr. Peter Gumbsch deserve my deep gratitude for critically reviewing this work.

Furthermore, I want to thank my group leader and friend, Dr. Thomas Straub. Your Ph.D. efforts set the groundwork for my study and I am holding our sessions dear in which you (patiently) introduced me to the experimental setup and the new work environment. You did not only contribute significantly to motivating me but also covered me which allowed me to focus and delve deep into my research interests. It will still take some time to get over you not being interested in the city of Essen but I am positive that we can stay friends.

Next, I want to thank the Bosch research foundation who funded this research. I always appreciated the freedom you provided to follow my own research ideas and the interdisciplinary exchange during the yearly meetings. This did not only help me to learn to speak in front of interdisciplinary audiences but also to befriend them. At the same time, I was provided with two excellent scientific advisors from Bosch research, Dr. Manfred Bacher-Höchst and Dr. Tobias Kennerknecht. Through lively exchange with Dr. Petra Sonnweber-Ribic, I got to collaborate with and befriend many Bosch Ph.D. students, namely, Nikolai Arnaudov, Benjamin Schäfer, and Erik Natkowski.

Prof. Dr. Christian Elsässer, Dr. Michael Luke, and Dr. Silke Sommer also supported me throughout my research by not only providing me the necessary encouragement and time in the final period of my thesis but also the opportunity to develop freely.

Next, I want to phrase my thankfulness towards my research partner in crime, Akhil Thomas. While our common interest in research at the interface between material science, fatigue, computer vision, and data-driven methods initiated our collaborations, our different angles of approach and previous knowledge make our discussions and numerous literature review, brainstorming and lab sessions especially enjoyable. I am looking forward to collaborating with you in the future, supporting you during your Ph.D. wherever I can, and finally eating some beef ularthiyathu over some board games. Also, Edward Kreutzarek, Aurèle Goetz, Muhammad Kashan Karim, and Sai Teja Potu were or are part of our machine learning 'team'. Our similar working ethos and interests rendered collaborating and celebrating a pleasure at all times. Thank you.

Another person I want to dedicate my gratitude to is Nadira Hadzic who supported me greatly in the course of this work by contributing many pondering hours and lines of code. You being the first master's thesis student I was allowed to supervise was a joy and accompanied by many educational aspects for me. I wish you the best in the somewhat new research field and am thrilled to see you continue developing.

Moreover, I would like to thank all past and present group members of the micro- and meso mechanics group, at the Fraunhofer institute for mechanics of materials. All of you contribute to the balanced, enthusiastic, ambitious, and easy-going

working atmosphere. The flat hierarchies that our executives set an example for in our daily work are accompanied by mutual helpfulness which rounds it off.

In particular, my deep gratefulness is directed to my parents, Hoşi and Halis Durmaz. Their love and selflessness are difficult to put into words, nothing short of admirable and enabled this in the first place. Thank you for your open-mindedness towards other cultures making it possible for me to find a home, become independent, and pursue my research even though not all did comply with your ideas initially. My older siblings, above all my brother Vedat Durmaz, helped me a lot when I was a young straggler not overly interested in education. Teşekkürler!

Freiburg im Breisgau, November 2022 *Ali Riza Durmaz*

Table of contents

List of abbreviations and symbols

Symbol		
$\mathbb{I}$	Image	–
$\vec{b}$	Burgers vector	m
$\vec{l}$	Line vector	–
A	Pixel area	–
A_z	Zener elastic crystal anisotropy	-
d	Grain size	m
DD	Kullback-Leibler divergence	–
g	Crystallographic orientation matrix	rad
H	Area-averaged accumulated intensity	–
h	Slip band width	m
I	Accumulated pixel intensity	–
k	Friction stress	MPa
L_{CE}	Cross-entropy loss	–
M	Taylor factor	–
m	Schmid factor	–
p	Cyclic slip irreversibility	–
R	Stress ratio	–
t_{exp}	Exposure time	s
t_i	Time between images	s
t_{st}	Pulse duration	s
W_s	Specific slip band fracture energy per unit area	J/m^2
ε	Uniaxial strain	–
μ	Statistical mean	–
σ	Uniaxial stress	MPa
σ^2	Statistical variance	–
a	Crack length	m
c	Surface crack length of a semi-circular crack	m
$f_{res,0}$	Resonant frequency at unfatigued specimen state	Hz
f_{res}	Resonant frequency	Hz
N	Cycle number	–
N_f	Cycle number to failure	–
N_i	Cycle number to crack initiation	–

N_{ro}	Cycle number to fatigue run-out (run-out criterion)	–
Q	Quality factor	–
Δg	Crystallographic misorientation matrix	rad
$\Delta\tau$	Shear stress range	MPa
$\gamma_{p,cum}$	Cumulative plastic shear	–
$\gamma_{p,irr}$	Irreversible plastic shear	–
κ	Growth rate	–
μ_s	Shear modulus	MPa
ν	Poisson's ratio	–
ω	Crystallographic disorientation	°
Φ	Second Bunge orientation angle	°
ϕ	Phase shift	°
ϕ_1	First Bunge orientation angle	°
ϕ_2	Third Bunge orientation angle	°
ϕ_p	Phase shift between piezoelectric actuators	°
ρ_{GND}	Geometrically necessary dislocation density	$1/m^2$
ρ_{SSD}	Statistically stored dislocation density	$1/m^2$
τ_{CRSS}	Critical resolved shear stress	MPa
τ_{RSS}	Resolved shear stress	MPa

Indices

bg	Background
C	Crystal
c	Crack
e	Extrusion
fg	Foreground
i	Image
o	Overall
p	Pixel
S	Specimen
te	A test portion of a data set
tr	A train portion of a data set
x	Component in x-direction
y	Component in y-direction
z	Component in z-direction

Constants

π	Circle constant	–

Abbreviations

AFM	Atomic force microscopy
AHSS	Advanced high-strength steels
API	Application programming interface
APT	Atom probe tomography
AUC	Area unter curve
bcc	body-centered cubic
CI	Confidence index
CNN	Convolutional neural network
CPFEM	Crystal plasticity finite element method
CS	Coordinate system
CSL	Coincident site lattice
CV	Computer vision
DCT	Diffraction contrast tomography
DIC	Digital image correlation
DL	Deep learning
DSIC	Data set imbalance correction
EBSD	Electron backscatter diffraction
EBSP	Electron backscatter patterns
ERF	Effective receptive field
FE	Finite element
FIP	Fatigue indicator parameter
GAN	Generative adversarial network
GB	Grain boundary
GND	Geometrically necessary dislocations
GOS	Grain orientation spread
GUI	Graphical user interface
HCF	High cycle fatigue
HEDM	High energy X-ray diffraction microscopy
HR-DIC	High-resolution digital image correlation
HR-EBSD	High-resolution electron backscatter diffraction
IoU	Intersection over union
IPF	Inverse pole figure
IQ	Image quality
KAM	Kernel average misorientation
LARS	Least-angle regression
LED	Light emitting diode
LOM	Light optical microscopy
MDA	Mean decrease in accuracy
MDF	Misorientation distribution function
MDI	Mean decrease in impurity

ML	Machine learning
MOR	Multiples of random
NLP	Natural language processing
ODF	Orientation distribution function
PSB	Persistent slip band
PSD	Position-sensitive detector
RMS	Root mean square
ROC	Receiver operating characteristic
SC	Sampling correction
SE2	Secondary electron Everhart-Thornley detector
SEM	Scanning electron microscopy
SIFT	Scale-invariant feature transform
SP	Slip plane
SWC	Sample weight correction
SWSC	Sample weight and sampling correction
TNR	True negative rate
TPR	True positive rate
TRF	Theroretical receptive field
TRIP	Transformation-induced plasticity
TWIP	Twinning-induced plasticity
wt	weight

1 Introduction

1.1 Motivation

A substantial portion of the current technological advances can be ascribed to materials development and improvement rendering materials science a key technology [1]. For instance, steels whose worldwide production was estimated 1259 million tons in the year 2005 [2] can adopt a vast range of properties depending on their composition and underwent thermo-mechanical treatments. Therefore, steels with their generally good formability and machinability can serve optimal structural and magnetic characteristics to a wide range of sectors such as automotive, construction, energy, and plant engineering. However, long development cycles are required for introducing tailored materials into production. This can be ascribed to necessary adjustments in processing, which are nowadays largely addressed by trial-error-based approaches, as well as to materials and component safety assessment. Furthermore, sustainability regulations such as decarbonization of production pose constraints to manufacturers [3].

There are multiple national efforts addressing the digital transformation of materials aiming to reduce the time to market. For instance, the Materials Genome Initiative (MGI) [4] in the United States, Materials Genome Engineering [5] in China, and MaterialDigital [6] as well as NFDI MatWerk [7] in Germany are aiming to achieve this through structuring and connecting materials data [8] as well as transcribing materials knowledge into machine-readable representations. In the future, digital twins of materials, e.g. predictive machine learning models, potentially can provide a remedy for time-consuming experimentation and process optimization. Moreover, instances of such models could monitor the material's state during operation. By combining materials knowledge and data to develop hybrid artificial intelligence-based digital twins, potentially generalization to a variety of materials and high fidelity can be achieved simultaneously.

Materials fatigue testing often poses the time limiting factor in materials assessment but is crucial considering that over 80% of materials' mechanical failure can be accounted to fatigue [9]. Bearing in mind the current scarcity of structured fatigue data, collecting and linking faithful fatigue data and knowledge according to the FAIR data principles (**F**indable, **A**ccessible, **I**nteroperable, and **R**euasble [10]) can not only reduce the time spent during testing and design but in the long run, could also culminate in component longevity and increased sustainability. Furthermore, product recalls due to service failure could be circumvented.

Fatigue describes the continuous degradation of a component exerted to cyclic loading. In the course of fatigue, damage mechanisms take place that range from dislocation structure formation over the arrangement of slip bands and extrusions to different crack states, ultimately culminating in terminal component failure. Therefore, fatigue phenomena encompass the atomic scale through crystallographic defect interactions or point defect diffusion, the mesoscopic scale where interactions between microstructural units in polycrystals modulate externally imposed loads, and the continuum level. This inherently renders fatigue a multi-scale process.

When external cyclic loading amplitudes are concerned which result in a high number of withstood loading cycles ($> 10^5$), initial fatigue states such as damage evolution, crack initiation and short crack growth govern most of the fatigue life. Another attribute of this so-called high cycle fatigue (HCF) regime is that the damage and crack initiation sites are rather few and localized to regions of critical microstructure (pronounced microstructure-sensitivity). However, neither sufficient data nor exhaustive mechanistic understanding is available on the microstructure dependency of these early fatigue states. This leads to the unsatisfactory lifetime prediction accuracy of current analytical and computational knowledge-driven models in the HCF regime and beyond.

The material composition and processing can culminate in a variety of distinct microstructures, intrinsic stress states, and surface properties collectively altering the material response to aforementioned cyclic loads and the underlying mechanisms [11]. In addition, during service life, mechanical loading characteristics along with environmental conditions affect the fatigue process severely.

Therefore, advancing the understanding in the field of HCF fatigue requires multimodal feature representations which take the aforementioned dependencies into account. However, to date, despite the vast availability of distinct imaging techniques, there is no individual characterization technique that can comprehensively capture all relevant features. Holistically, full-field deformation and microtexture information in three dimensions would be preferred to derive a mechanistic understanding. Simultaneously, in the HCF scenario, where typically elevated loading frequencies are utilized, sufficiently high time resolution would be beneficial to resolve individual cycles. Additional requirements posed are non-intermitted testing and spatial resolutions allowing the extraction of microscopic defects, without relying on electron microscopy in vacuum environments which alter the fatigue properties. While there are powerful aspiring experimental techniques able to collect data on the three-dimensional microstructure evolution under cyclic loading such as in-situ near-field and far-field high energy X-ray diffraction microscopy (HEDM) or diffraction contrast tomography (DCT) [12–14], these suffer from limited availability and are not applicable for in-situ HCF testing [15, 16] on a wide variety of materials. Especially when sheet beams rather than scanning pencil beams are utilized, these methods typically fall short of conventional electron backscatter diffraction (EBSD) in spatial resolution by approximately one order of magnitude. While these X-ray-based approaches are predestined to retrieve sub-surface information in a non-destructive fashion, nuanced intragranular deformation fields are comparatively difficult to extract. In contrast, the vast majority of experimental works attempt to unify multimodal information from several consecutive surface measurements [17–21].

The multimodal and multiscale character of the problem along with the diversity of the materials impede the data post-processing required to convert the raw data into interpretable information. Conventional computer vision (CV) pipelines are sometimes used to spatially align such data (registration [21]), detect damage, or infer deformation fields [22] but often rely on extensive, manual feature engineering and user experience [23]. Adopting such rule-based CV approaches to another material requires fine-tuning many model parameters. Owing to the fact that HCF damage formation only occurs in a few regions of critical microstructure or defects and therefore poses rare events, data sets are often accompanied by severe data imbalance. This scarcity of damage instances, in turn, raises the requirement for a large number of tests to collect statistically relevant data. This is aggravated by the fact that most studies are built on incomplete feature representations without information about sub-surface characteristics or local surface roughness. The increased uncertainty coming with it leads to a higher data demand to infer causal relations in the high dimensional data. A significant portion — often even the majority — of time allocated to research efforts is spent during data pre-processing stages, which highlights the importance of generalizable data post-processing pipelines. However, due to the constraints introduced in this paragraph such standard post-processing routines are at the moment unavailable.

Currently, machine learning (ML) approaches find increasing acceptance in the field of materials science. For problems such as fatigue, where the mechanistic understanding is incomplete and the feature space is vast, ML can facilitate learning complex, unknown relations from data to achieve higher fidelity in fatigue predictions. A criticism frequently posed to such models is that the decision process of deep learning variants is not traceable [24], that representation learning methods require a large number of input-output pairs due to their often limited data-efficiency [25], and that their generalization capability is limited [25–27]. These limitations are in conflict with the pronounced expense associated with fatigue experimentation, the costly data annotation through experts, and materials diversity. However, these limitations were the subject of much research in the computer science field in the past decade. Nowadays, there are many novel techniques to bypass former limitations. For instance, few-shot learning techniques address learning from few examples [28], domain adaptation models tackle transferring between domains to achieve model transferability [29], and knowledge graph embedding [30] as well as physics-informed models [31] achieve better generalization through the incorporation of existing knowledge. Extensive knowledge-driven modeling efforts in the materials domain, from various computational techniques covering all scales, can supplement synthetic data [32, 33]. Due to increased industrial automation and data acquisition rates, the discrepancy between input data and annotated data is continuously rising. Therefore, the demand for semi- or unsupervised ML methods which

require only a portion of the data to be annotated will inevitably gain relevance. Along with the digital transformation of the materials science field improving the systematic collection and integration of disjoint materials data, these new techniques can assist in unraveling unresolved problems such as fatigue.

1.2 Objectives and approach

The primary objective of this work is to establish an efficient workflow to collect and fuse multimodal experimental and computational HCF fatigue data as well as complementary specimen characterization data. Thereby the expense to acquire and process statistically relevant fatigue data should be reduced. Specifically, data on the early stages of fatigue damage evolution on the microstructure level is addressed. Fatigue characterization of mesoscale specimens leveraging a bending resonance setup presented in [34] and [35] represents the foundation of the workflow. The bending resonance characterization of mesoscale specimens unifies a few beneficial properties. On one hand, the load is predominant in the surface near regions. These regions are small enough to be characterized by time-consuming techniques such as EBSD providing crucial microtextural features for fatigue. On the other hand, higher frequency loading permits accumulating many loading cycles within maintainable testing time spans. Moreover, the sensitivity is pronounced enabling the detection of initial damage. In terms of data processing, multimodal data registration, segmentation, clustering, and filtering are core features of the workflow.

An overview of the workflow addressing experimental, computational and data analysis steps is depicted in figure 1.1 and elaborated briefly in the following.

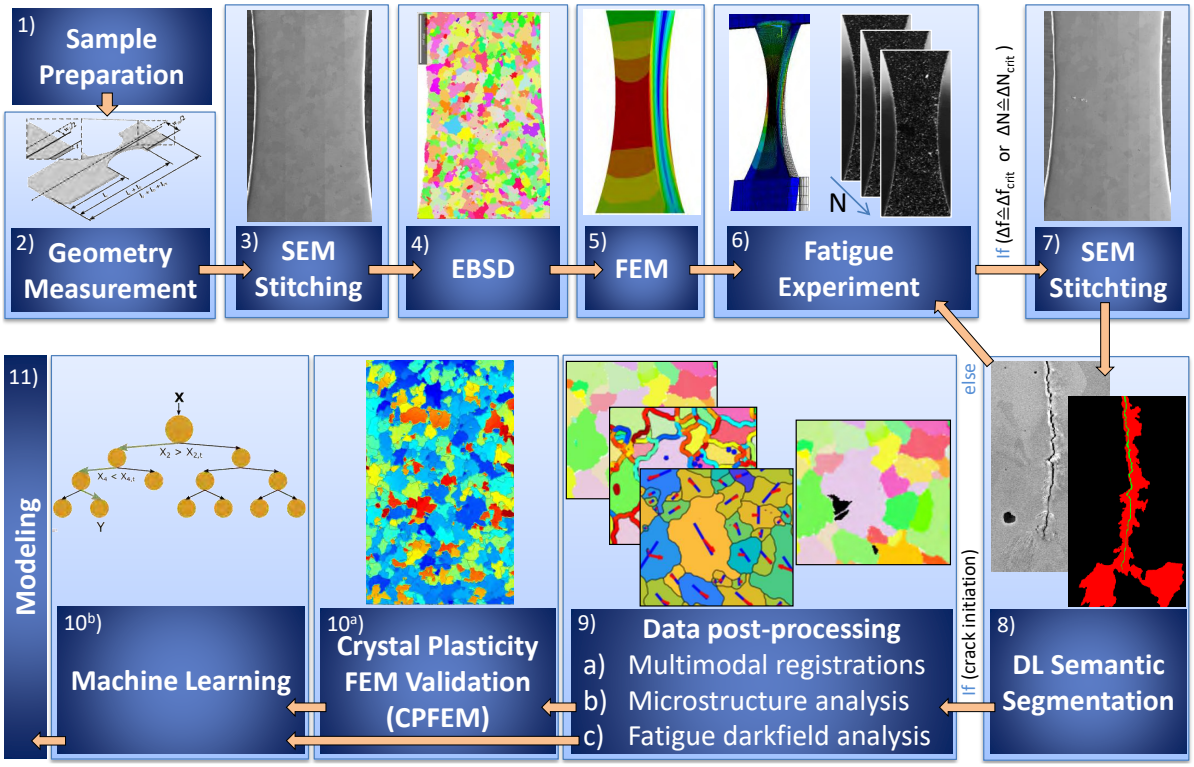

Figure 1.1: Overview of the workflow including the experimental approach and the data handling [36].

After specimen preparation (1) and geometry assessment (2), automated mosaic image acquisition in a scanning electron microscope is performed with a topography-sensitive detector (3). The subsequently stitched image serves the purpose of providing information about the initial state (e.g. pore defects) in the whole highly loaded specimen area without sacrificing resolution. Moreover, this image acts as a reference for registration due to the absence of deformation and its reduced imaging-induced spatial distortion. Analogously, for some specimens, EBSD (4) is acquired for microtexture information (pixel-wise

sampled crystallographic orientation). Upon later spatial alignment of the different image types, this facilitates attributing fatigue damage instances to certain microstructural entities such as grains or their interfaces. A homogenized linear elastic FE analysis (5) provides an initial estimate of the specimen geometry-induced stress distribution and determines the target amplitude during testing. Subsequently, a bending resonance fatigue experiment (6) is performed. This experiment employs a frequency drop criterion along with in-situ imaging to obtain information on integral and local damage formation. Another SEM stitching image acquisition (7) aimed at attaining comparatively high-resolution damage information completes the experimental procedure. By comparison with the initial SEM stitching, the degree of deformation and whether prior defects were responsible for damage initiation can be deduced.

All aforementioned data undergoes extensive post-processing (8/9). This includes the following aspects:

1. Filtering, smoothing, and clustering of EBSD data to extract potentially relevant microstructural features
2. Registration of the in-situ image series correcting for slight specimen slip during fatigue
3. Multimodal registration of all image data sets
4. Semantic segmentation of damage instances, namely protrusions and cracks, from the post-mortem stitched SEM image
5. Damage segmentation in the light optical in-situ image series by utilizing the SEM-derived damage mask
6. Image processing for a quantitative damage evolution analysis from the registered in-situ image series

All functionalities are integrated into a graphical user interface (GUI) to facilitate user input at different stages, for instance during the point correspondence selection for multimodal data registration.

Since the post-processing culminates in displacement boundary conditions, specimen geometry, and microstructure being fused and the relevant, highly-loaded domain size is manageable in terms of computational expense, the data sets provide straightforward integration into computational modeling efforts (10^{a}). For instance, these data sets facilitate the parametrization and validation of computational micromechanical fatigue models such as crystal plasticity finite element method (CPFEM) or related fast Fourier transform (FFT)-based approaches. A small portion of the work at hand is dedicated to comparing hotspots in CPFEM-derived so-called fatigue indicator parameter (FIP) maps with experimental damage masks regarding their spatial correlation. In opposite direction, the data space can be further enriched by supplementing features derived from crystal plasticity.

Through this workflow, a data foundation for HCF damage prediction appropriate for training ML models shall be created. Here the aim is to produce highly quantitative data sets for data-driven fatigue damage evolution modeling or data mining to find sensitive features. A case study is conducted where different microtextural, micromechanical, topological, morphological, and loading features are extracted from the fused data to predict damage initiation locations. Therefore, a binary classification algorithm is applied considering grains as instances. Exemplary, decision tree and random forest-based approaches are tested (10^{b}). Thereby, the interplay and relevance of such quantitative features in determining damage-prone grain ensembles are explored. The derived knowledge can then provide the groundwork for future modeling purposes (11).

Moreover, the work at hand aims to provide methodological insights into fatigue characterization and data post-processing techniques. Both classical computer vision and machine learning elements are integrated into the workflow to achieve a high degree of automation. Within this workflow, in particular, the appropriate usage of deep learning models (DL) for the segmentation of initial damage is investigated. In this regard, the focus lies on convolutional neural network (CNN) models for segmentation tasks. Such models are typically composed of an encoder and decoder portion. The encoder extracts relevant image textural features and the decoder recovers the spatial resolution for the dense prediction task. The influence of model weight initialization, image data pre-processing, and data augmentation on the final model performance is explored. Emphasis is placed on methodologies reducing the data demand and improving generalizability. In conclusion, major shortcomings of DL methodologies (introduced in section 1.1) preventing the applicability in materials science are tackled.

Prospectively, this workflow is supposed to address the present lack of microstructure-informed fatigue data for a wide range of metals. Therefore, the generalizability of essential workflow elements to alternate materials is explored and discussed. While the majority of the work is performed on a highly chromium alloyed ferritic steel, for the generalization study of segmentation models, martensitic steel and high-purity copper are considered additionally. Ideally, the data registration routines should enable spatial correlations between further relevant modalities and the linking of data on distinct length scales.

The present doctoral thesis draws from four previously published journal articles. Primarily, it is based on a publication that presented the machine learning-assisted experimental workflow to generate data on initial fatigue damage efficiently [36]. The contents of that publication were created by the author by conducting experiments, data processing, analysis, and writing. Another effort that is fundamental to the doctoral thesis presented here is the one on deep learning segmentation of damage instances [37]. In that work, major contributions were made by conceptualizing the study, data analysis, and writing, and minor contributions to writing the training code. Furthermore, a few contents of two other journal publications are included here. Both address the validation of crystal plasticity fatigue simulations performed by industrial partners. While one introduces the validation methodology and utilizes it to compare predicted hotspot maps with segmented damage maps [38], the other assesses the fidelity of microstructural short crack growth paths predicted by a phenomenological CPFEM model against experimental data [39]. In the former case, the validation methodology and metrics were proposed and implemented by the candidate and major contributions were made in terms of data analysis and writing. For the microstructural crack growth paper, minor contributions were made in terms of conceptualization and writing.

2 Fundamentals

2.1 Microstructure and fatigue mechanisms

2.1.1 Microstructure of ferritic steels and crystallographic descriptors

There is a vast variability of steels resulting in substantial customization potential and tailored steel materials for specific applications. The thermo-mechanical processing history and alloying composition of steel strongly affect its microstructure and thus the properties of steel. There are various steel categories distinguished by their alloying element content (e.g. carbon) into grade standards such as those of the Society of Automobile Engineers (SAE). Application fields are amongst others construction, transportation, plant engineering, energy, and automotive which each requires certain combinations of mechanical, magnetic, and corrosion resistance properties. Despite steels' longstanding usage in many applications, there is significant progress in the development of advanced high-strength steels (AHSS). For instance, dual-phase steels combining ferrite and martensite phases or twinning/transformation-induced plasticity (TWIP/TRIP) steels [40, 41] belong to this group. The development of these is driven by the demand for steels that combine both, outstanding strength and toughness [42]. Another branch of steel research attempts to attain competitive properties from recycled materials despite the high contents of tramp elements [43].

In polycrystalline materials, microstructural units, so-called grains, are defined as volumetric regions with a regular lattice composed of periodic elementary unit cells (see figure 2.1a). In reality, even within grains, lattice rotations, so-called intragranular misorientations, exist which can be ascribed to defects present in the individual grains. The shown elementary cell with atoms situated at the corners and the center (single atom basis) is referred to as body-centered cubic (bcc). In mildly-alloyed steels, this atomic configuration is reflected in the thermodynamically stable low-temperature phase *ferrtite*.

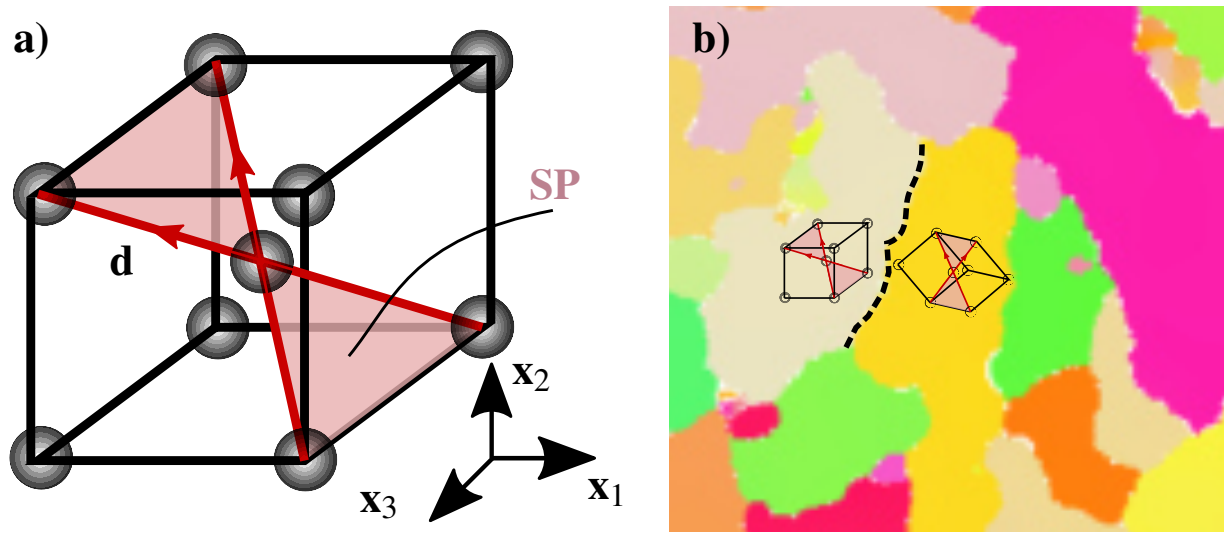

Figure 2.1: a) Body-centered cubic elementary cell with indicated slip planes (SP) and slip directions (d). **b)** Microstructure map in inverse pole figure color coding with two unit cells illustrating the orientation distribution. The dashed line represents the grain boundary.

The unit cell exhibits octahedral voids situated at the centers of faces and edges. The unit cells' symmetry is characterized by its point group *m3m*, corresponding to 24 symmetry operations to which the unit cell is invariant. Present asymmetries in this unit cell give rise to anisotropic (=direction-dependent) materials properties in individual grains. For instance, the elastic anisotropy factor proposed by Zener [44] for cubic crystals describes the anisotropy in elastic modulus using individual tensor elements of the second-order stiffness:

$$A_z = \frac{2c_{44}}{c_{11} - c_{12}}. \tag{2.1}$$

In the case of iron, pronounced elastic anisotropy arises ($A_z \approx 2.4$ [45]).

There are multiple ways to describe the crystallographic orientation of grains with respect to a specimen coordinate system (CS_S) including the 3×3 orientation matrix g. This orientation matrix maps CS_S into the crystal coordinate system (CS_C) and is composed of the direction cosines between each of the axes.

$$CS_C = g \cdot CS_S = \begin{pmatrix} cos\alpha_{1X} & cos\alpha_{1Y} & cos\alpha_{1Z} \\ cos\alpha_{2X} & cos\alpha_{2Y} & cos\alpha_{2Z} \\ cos\alpha_{3X} & cos\alpha_{3Y} & cos\alpha_{3Z} \end{pmatrix} \cdot CS_S \tag{2.2}$$

Here the subscripts denote the axes between which the angles are measured, where the first refers to crystal axes (1 = [1 0 0], 2 = [0 1 0], 3 = [0 0 1]) and the second to specimen axes X, Y, Z. From this general representation, different other means to describe crystallographic orientations, directions, or planes can be derived including the Euler angle, quaternion, Rodrigues vector, or Miller index representations. When orientations are encoded in Miller notation [46], the last and first columns are extracted from g, posing the specimen Z and X axis directions in crystal coordinates, respectively. Then both vectors are multiplied with their least common multiple $N_{Z/X}$ to obtain integer values:

$$(h,k,l)[u,v,w] = (N_Z cos\alpha_{1Z}, N_Z cos\alpha_{2Z}, N_Z cos\alpha_{3Z})[N_X cos\alpha_{1X}, N_X cos\alpha_{2X}, N_X cos\alpha_{3X}]. \tag{2.3}$$

Accordingly, directions, planes, and their symmetrically equivalent families can be described in Miller's notation. This notation provides an intuitive understanding of which crystallographic axis/plane is aligned with relevant specimen axes (i.e., the tensile axis or the rolling direction for sheet materials). In Miller notation, square, round, curly, and triangular brackets are used to refer to crystallographic directions, planes, sets of equivalent planes, and set of equivalent directions, respectively. The Miller indices of a plane are computed by finding its intersections with the crystal coordinate system in terms of multiples of lattice spacing, computing the reciprocals, and subsequent multiplication with the least common multiple.

An alternative and commonly used way of encoding orientations is the Euler representation, where arguably the Bunge convention is the most widespread in the domain of crystallography. This convention applies three consecutive rotations with angles ϕ_1, Φ and ϕ_2 around the specimen Z axis, the transformed X axis, and finally the transformed Z axis to map CS_S onto CS_C. These angles are defined in the ranges $0° \leq \phi_1, \phi_2 \leq 360°$ and $0° \leq \Phi \leq 180°$.

When considering full polycrystals, rather than individual grains, pole figures are a common tool to visualize the distribution of specific crystal axes/planes in the specimen coordinate system. This is achieved by computing the intersections of the target plane normals with a unit sphere and mapping them into the equatorial plane by stereographic projection. For more details, the author refers to [47]. For instance, rolling can result in non-uniform orientation distributions. This so-called *macroscopic texture* can be measured by pole figures and more extensive Euler or Rodrigues space representations.

On the microscopic scale, when transitioning into a neighboring grain, an orientation discontinuity is present at the grain boundary (GB), see the dashed line in figure 2.1b. GBs are two-dimensional defects that form networks and are affecting materials properties to a large extent. They can be described by five macroscopic parameters [48], three accounting for the crystallographic misorientation Δg across the boundary, and two for the GB plane orientation. Especially when considering its additional microscopic degrees of freedom such as the curvature, internal dislocation structure, or GB segregation, the parameter space becomes vast [49, 50]. A change in orientation between grain 1 (reference) and grain 2 can be expressed by

$$\Delta g = g_1^{-1} \cdot g_2. \tag{2.4}$$

The angle-axis representation poses a frequently-used descriptor, where an axis, given in Miller indices, and a rotation around it, given as a *disorientation* angle, maps the reference crystal coordinate system into the second crystal coordinate system. Sometimes the disorientation angle is used to categorize GBs into small-angle GBs ($< 15°$) and large-angle GBs ($> 15°$) which are associated with distinct characteristics. While absent in single-phase materials, phase boundaries are the analogous concept separating two different phases. Moreover, there are so-called twin boundaries, where the atomic configuration on both sides is mirrored. In coincident site lattice (CSL) boundaries the adjoining grains share potential lattice sites, where the reciprocal of the Σ-parameter measures their portion. Such boundaries in some cases exhibit special properties [51].

The crystallographic planes characterized by the highest planar packing density of atoms and relatively large plane distances (small Miller indices) are referred to as slip planes. Equivalently, slip directions are defined as directions within the slip planes which possess the highest linear atom packing. An individual $(1\,1\,0)$ slip plane (SP) of the bcc crystal along with its slip directions (d) $[\bar{1}\,1\,1]$ and $[\bar{1}\,1\,\bar{1}]$ are shown in figure 2.1a. However, owing to the symmetry of the bcc unit cell, six equivalent $\{1\,1\,0\}$ slip planes with two $\langle 1\,1\,1\rangle$ slip systems each exist, amounting to twelve *slip systems*. As the name implies, slip systems compared to arbitrary directions and planes, require lower resolved shear stresses in order to permit dislocation slip and hence glide of crystallographic planes against each other. Moreover, there are thermally-activated slip planes (12 $\{2\,1\,1\}$ type and 24 $\{3\,2\,1\}$ type) with one $\langle 1\,1\,1\rangle$ direction each. The activity of these systems depends on chemical composition, strain rate, strain amplitude, and temperature mainly [52].

2.1.2 Dislocations and their role in fatigue

Dislocations are often portrayed as infinite atom half-planes that are inserted into a crystal. They are considered line defects since the lattice irregularity is localized to the dislocation core (tip of the inserted plane defined by the line vector $\vec{l}$). The dislocation core is surrounded by an elastic stress field. Dislocation line defects nucleate beyond a critical loading at all types of defects, such as grain boundaries, precipitates, inclusions, and the free surface. The dislocations' density and velocity in the crystal determine the amount of plastic deformation observed in the material. According to [53], dislocation slip is activated once the resolved shear stress τ_{RSS} on a dislocation exceeds the critical resolved shear stress τ_{CRSS}. The mobility of dislocations as well as τ_{CRSS} show a dependence on temperature, strain rate, material, and the slip plane family. While the so-called Burgers vector ($\vec{b}$) defines the direction and magnitude of the caused atomic displacement, the line vector defines its geometry. Depending on the relative orientation of those two vectors to each other, the literature distinguishes between edge dislocation ($\vec{b} \perp \vec{l}$) and screw dislocation ($\vec{b} \parallel \vec{l}$).

Dislocation *glide* or *slip* is a mechanism, which describes the motion of dislocations on crystallographic densely-packed planes and directions (slip systems). Due to their characteristics ($\vec{b} \parallel \vec{l}$), screw dislocations can operate on all slip planes which contain the dislocation line. The mechanism by which screw dislocation segments can change the slip plane is denoted as *cross-slip* [54]. This phenomenon allows screw dislocations to circumvent obstacles. The process of cross-slip is thermally activated and, when activated on a large scale, can cause dynamic recovery due to the annihilation of screw dislocations with opposing signs. In contrast, edge dislocations glide only on slip planes that contain both, the line and Burgers vector. However, when vacancies interact with an edge dislocation it can give rise to a displacement of the dislocation out of the slip

plane, so-called dislocation *climb*. This process is largely dependent on the environmental temperature as vacancy mobility is affected strongly.

Despite the conceptual view of edge and screw types, dislocations observed in transmission electron microscopy studies are typically a mixture of both types. Typical are so-called dislocation loops or dislocation dipoles [55]. As the name implies, in dislocation loops the dislocation line vector is a closed loop. In contrast, dipoles represent two dislocations on parallel planes with identical Burgers vectors but opposing line vectors. These dipoles are usually composed of edge portions as cross-slip in screw components results in annihilation. Dislocation loops are typically formed by Frank-Read sources, which represent partial, sessile (i.e., immobile) dislocations with both ends pinned causing dislocation multiplication. In some materials, when energetically favorable, dislocations can dissociate into dislocation segments with different characters. Also, the interaction of dislocations often leaves sessile partial dislocations behind, referred to as dislocation debris. This debris functions as obstacles for glissile dislocations and thereby contributes to cyclic hardening.

Dislocations in fatigue. Aforementioned orientation differences between adjacent crystalline regions in conjunction with elastic modulus anisotropy cause microstructural stress concentrations at grain boundaries. Such elastic incompatibilities, along with mismatched plastic properties at intermediate strains, cause microstructure-based modulation of the specimen geometry-induced stress distribution [56]. This gives rise to a local increase in dislocation density in the vicinity of grain boundaries [57]. Especially in such regions, irreversibilities in dislocation motion occur and a portion of the total cumulative plastic shear $\gamma_{p,cum}$ within one loading cycle remains. This portion is the irreversible plastic shear $\gamma_{p,irr}$. Their ratio, the cyclic slip irreversibility p is defined in the range $0 \leq p \leq 1$ and depends amongst others on the loading amplitude [58].

$$p = \frac{\gamma_{p,irr}}{\gamma_{p,cum}} \tag{2.5}$$

According to Mughrabi [58], phenomenons in fcc materials causing irreversibilities are cross-slip of screw dislocations, the annihilation of screw or edge dislocations, and to-and-fro glide of dislocations. The mechanisms of microplasticity in fcc materials are comparatively well understood and will be introduced briefly in the following as they lay the foundation for bcc materials.

In the course of further cycling, dislocation multiplication and the formation of low-energy dislocation structures takes place. Dislocation-rich regions (*veins*) form consisting of regions with clustered edge dislocation dipoles, and dislocation loops all oriented perpendicular to the primary slip plane and the slip band. The stress field surrounding individual straight dislocations decays with 1/r, while the decay in dislocation loops and dipoles is more pronounced. Therefore, dislocation loops and dipoles represent low energy configurations which also permit higher dislocation densities in veins as the repulsive forces, which dipoles exert onto each other are comparatively less. The vein regions are separated by dislocation-depleted *channels*, together forming a (persistent) slip band (PSB). This alternating configuration in the bulk is coined *ladder* structure. Depending on the material and load, these ladder structures come in different manifestations and are enclosed by the matrix in which typically veins exist as well. Irreversible plastic strain is localized in PSBs after their formation. During cyclic loading an increase of vacancies is observed due to the annihilation of predominant vacancy-type dislocation dipoles [59]. Vacancies are produced in the channels of the PSB and flow into vacancy sinks (edge dislocation-rich veins in PSB) or the matrix. Therefore, atoms are flowing from the matrix into the channels and from the veins into the channels. In Polák's extension [59] of the EGM model [60], the continuous production of vacancies in PSBs and their flow into sinks cause mass transport into the PSBs. Therefore, in contrast to the EGM model, [59] explains not only the extrusion relief formation at the free surface of the PSB but also the intrusion formation where the PSB-matrix interface meets the surface.

Dislocations in bcc materials. As this work addresses fatigue in ferritic (bcc) steel, in the following a brief survey of microscopic plasticity in such materials is presented and peculiarities are emphasized. In the low-temperature regime of bcc materials, the dynamics of dislocations differ from that of fcc materials. It is defined as the regime where higher yield stresses occur due to the low mobility of screw dislocations. The transition temperature depends on the strain rate, where

higher strain rates effectively increase the transition temperature [61]. Since plasticity-governing cross-slip of $1/2\langle 1\,1\,1\rangle$ screw dislocations is inhibited at low temperatures, bcc materials behave brittle. At low stresses, slip is facilitated by thermally-activated nucleation of in-plane steps on the dislocation line (*kink pairs*) lowering the required load to move the dislocation [62]. The immobility at low temperatures suppresses the Frank-Read dislocation multiplication mechanism strongly [61]. Dissociation of dislocation cores was reported to be absent in bcc materials since there are no related metastable stacking faults [63]. Furthermore, it is observed that also non-shear components couple with dislocations (non-Schmid behavior), which can be ascribed to the fact that the $1/2\langle 1\,1\,1\rangle$ dislocation core extends across multiple planes of the $\langle 1\,1\,1\rangle$ zone [63]. This extended core is also considered the origin of the pronounced lattice friction (*Peierls stress*) resulting in comparatively higher yield stresses in bcc materials and the temperature/strain rate dependence. Another anomaly encountered is the influence on the sign of shear (*twinning-antitwinning asymmetry* or *slip plane asymmetry*) on slip activation. This mechanism causes irreversibilities in slip due to distinct screw dislocation paths during tension and compression.

In a fatigue setting (often at high strain rates), the reversible to-and-fro glide of edge dislocation segments in mixed dislocations or individual edge dislocations accommodates the plasticity, when small strains are concerned [58]. Irreversibilities are caused due to the interaction of such dislocations with defects and predominantly the free surface. The elastic strain energy is minimized when a dislocation moves to the free surface. To accommodate the cyclic plasticity, multiple slip is initiated which upon cycling is accompanied by hardening and dislocation cell formation confined to some surface near regions. As the loads are increased, these regions extend further into the bulk. While classical dislocation structures, e.g. ladder structures as in fcc are typically absent in bcc *metals*, extrusions and intrusions still form at the surface.

As opposed to this, in *alloyed* materials, the dislocation slip behavior is amongst others altered by dynamic strain aging. This mechanism refers to an increase in pinning strength of the otherwise mobile edge dislocations, which effectively assimilates the mobilities of screw and edge segments and results in dislocation dynamics more similar to fcc. This was hypothesized to contribute to the formation of PSB structures [61] otherwise atypical for pure bcc metals. In an extensive slip marking formation study, Man et al. [64] found ladder structures in ferritic X10CrAl24 steel similar to those in fcc materials with edge dislocations-enriched walls and channels containing few screw dislocations. In general, such ladder structures were only found in alloyed bcc microstructures, e.g. [65, 66]. As opposed to typically reported fcc cases, the ladder-type PSB structures reported in [64] and [66] are not surrounded by high dislocation density regions (veins) in the matrix. This is argued to affect the mass redistribution (diffusion of vacancies) [64]. Wavy slip is observed on the surface and argued to be typical for bcc metals due to the importance of dislocation cross-slip. Cracks are formed from intrusions which are accompanying individual extrusions. Comparatively high loads in bcc materials facilitate slip markings across the whole grain area with a slight localization towards grain boundaries. In individual instances, slip markings were observed to penetrate through grain boundaries [64]. Crack initiation at the location where PSBs impinge on GB was reported to be another dominant mechanism [67].

As the contribution of screw dislocations to plastic deformation increases (through alloying, higher loads, or higher temperatures), the impact of slip plane asymmetry becomes relevant. Effectively, this results in shape changes of crystals, especially when they are weakly constrained such as in surface regions. This has been shown to favor intercrystalline crack initiation [68]. When screw dislocation glide is activated, cross-slip occurs readily in bcc materials due to their multitude of intersecting slip planes. For iron activation of both $\{1\,1\,0\}$ and $\{1\,1\,2\}$ slip was reported [69].

Furthermore, intergranular crack growth was reported to be dependent on the grain size, where larger grain bcc materials tend towards intergranular crack initiation as observed in the case of low carbon steels [70]. Typically, grain boundaries orthogonal to the loading axis connecting grains with a substantially different plastic response were affected.

Previously, GBs and ledges in them were argued to be the dominating source for dislocations [57]. Therefore, high GB densities are associated with high dislocation densities [70]. The importance of the grain boundary plane for the grain boundary energy was highlighted by [71] in bcc metals using first principle methods. By molecular dynamics simulations, the static grain boundary energy was shown to have an inverse relationship with the dislocation nucleation energy from GBs as well as the energy barrier for dislocation-GB transmission [72]. Therefore, GBs affect the local hardening response of the material. Depending on the grain sizes and dislocation densities present in the material, the influence of slip transmission on the

mechanical properties varies [73]. In pioneering transmission electron microscopy studies [74–76] different criteria for slip were determined. In [77], mechanisms, influence factors and models of slip transmission are reviewed extensively. The ideal transfer of dislocations across GBs requires identical Burgers vectors of the adjoining slip systems. Otherwise, a residual dislocation remains in the boundary plane [78]. If slip transmission is inhibited, dislocation pile-up arises. These insights have led to the formulation of a multitude of slip transmission metrics. In nanoindentation experiments, Soer and De Hosson [79] as well as Britton et al. [80] found, that slip transmission was preceded by dislocation pile-up exerting stress at the tip dislocation. Lattice interstitial atoms (carbon) were pointed out to affect slip transmission substantially due to the formation of Cottrell atmospheres pinning the dislocations. Confirming previous observations, large slip plane and Burgers vector misalignments led to increased critical stresses to conquer the boundary. Furthermore, grain boundary segregation was previously reported to affect slip transmission [81].

2.1.3 Fatigue – A multi-scale problem

Fatigue refers to the degradation of materials under cyclic loading, which eventually culminates in failure. Typically, fatigue properties are characterized in S-N curves by imposing constant cyclic load amplitudes and measuring the cycle number to failure N_f, where failure is often defined in terms of a load drop. An exemplary S-N diagram is depicted in figure 2.2. The phenomenon of fatigue has been observed over a large range of stress or strain amplitudes, even at amplitudes well below the macroscopic elastic limit. In the regime of small cyclic load amplitudes, and particularly pronounced in ferritic materials, the S-N curve exhibits a plateau (the so-called high cycle life endurance limit) indicating infinite life in the absence of internal non-metallic inclusions or pores [82]. The amount of plasticity and dissipated energy in an individual cycle can be quantified by the area within the stress-strain (σ-ϵ) hysteresis loop. While loading amplitudes below the classical endurance limit result in elastic behavior, the limit itself can be perceived as the threshold to crack propagation, where plasticity, damage accumulation and crack initiation can still occur.

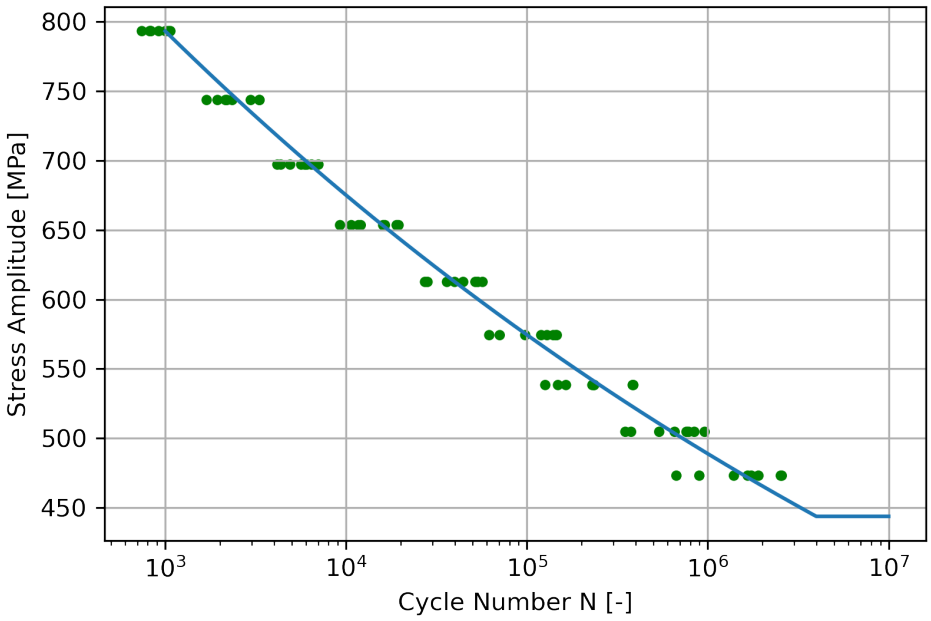

Figure 2.2: Schematic of a S-N diagram.

Fatigue is a multi-scale problem, which ranges over multiple length scales from dislocation interactions to macroscopic crack growth and failure. The amplitude, amongst other influence factors, strongly affects the underlying mechanisms of fatigue

and its sensitivity to the internal local structure of the material. While in and beyond the so-called high cycle fatigue (HCF) regime under low cyclic loads the accumulated cycle number N_f depends strongly on specimen intrinsic influence factors, the opposite holds true for high cyclic loads. Therefore, for low loading amplitudes generally, a larger scatter in life is observed (see figure 2.2). Under these conditions, the global lifetime is to the largest extent determined by the stages up to crack initiation [83, 84] as highlighted in figure 2.3.

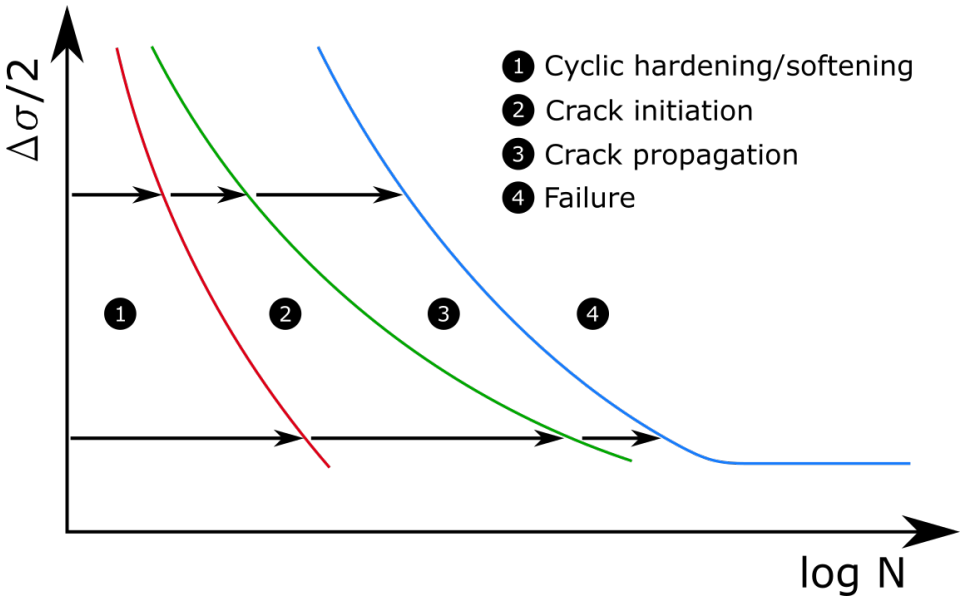

Figure 2.3: S-N diagram (blue) with additional curves indicating cycles to crack initiation (green) and cyclic saturation (red), after [85].

In literature, depending on the scientific domain, multiple ways of subdividing fatigue life into hierarchical stages were proposed, which renders comparisons difficult. The categorization in figure 2.4 is used throughout this work. Initially, in the course of cycling, dislocations are moving on preferably-oriented slip systems and interact with each other, grain boundaries, the free surface, and other types of defects. This renders dislocations immobile and causes cyclic hardening. Usually, the dislocations, carriers of plasticity, localize in the vicinity of grain boundaries, which pose barriers to dislocation motion, and there form dislocation structures and slip bands in the bulk. These regions of high dislocation density are accompanied by irreversible plastic slip steps at internal interfaces and, most evidently, at the free surface. Through the accumulation of irreversible micro plasticity, pronounced surface topographies form. Depending on their shape, such topographies are referred to as protrusions, extrusions, intrusions, and combinations thereof. The manifestation of such surface damage is strongly related to the underlying microstructure and loading. As a result of such topography and the stiffness mismatch between the matrix and the dislocation enriched (i.e., hardened) regions, cracks are likely to nucleate at the corresponding interface.

As opposed to figure 2.3, crack propagation is distinguished into microcrack growth, short crack growth, and long crack growth due to their inherently distinct mechanisms. This partitioning is in line with [84], where microcrack and short crack growth are referred to as microstructurally small crack growth and physically small crack growth, respectively. Microcrack growth, also referred to as stage I growth, is typically confined to individual grains and slip systems that are favorably oriented for shear-activated slip (around 45° to the loading axis). The crack is dominated by fracture mode II (in-plane shear) and mode III (out-of-plane shear) [86]. In the (physically) small crack growth regime, the crack length exceeds that of multiple microstructural units but the plastic zone at the crack tip is comparable to characteristic microstructural length scales (e.g. grain size) affecting the fracture mechanisms. As a consequence of the higher stress intensity at the crack tip, multiple slip (i.e., activation of multiple slip systems) can occur in the plastic zones at the crack tip. In both aforementioned regimes, linear elastic fracture mechanics are invalid. Upon further crack growth, the long crack growth stage is reached, where the crack path deviates from slip planes and aligns orthogonally to the load axis (normal stress-dominated crack opening mode I). Depending on whether the crack transitions through grains or coincides with grain boundaries, they are distinguished into trans- and intercrystalline cracks.

In the initial stages of fatigue, the microstructure-sensitivity is pronounced, as dislocation slip is localized to a few grains, and grain boundaries pose barriers to dislocation motion and micro-/short cracks.

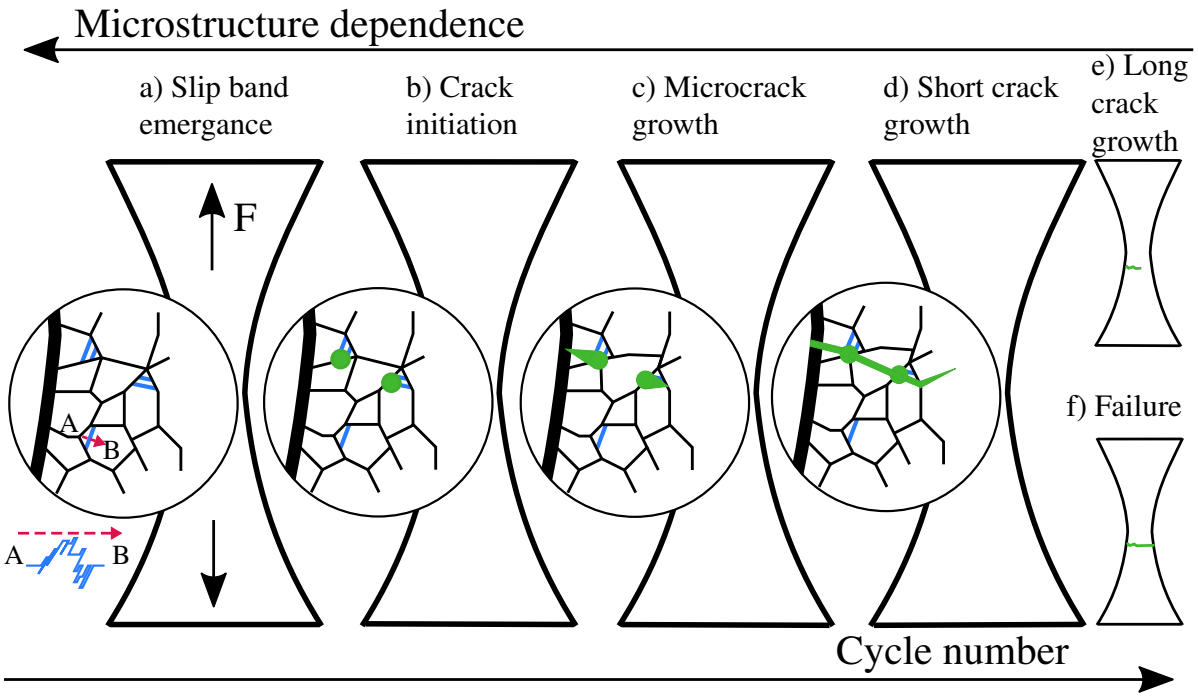

Figure 2.4: Schematic representation of the multi-scale characteristic of fatigue process covering microstructure sensitive (**a–d**) and insensitive (**e–f**) states. The blue color indicates persistent slip bands and slip markings. Closed green circles and lines represent crack initiation and growth, respectively.

A non-exhaustive overview and categorization of influence factors are given in the following:

- **Geometric**: The macroscopic stress state is substantially affected by the component or specimen geometry. Specimen surface topography such as waviness and microscopic roughness represent notches acting as stress concentrators on different scales. Moreover, the volume of the specimen exposed to high loads can alter the fatigue properties substantially. In particular, in brittle and high-strength materials, where manufacturing-induced critical defects or inclusions determine the fatigue life, the volume affects the likelihood of encountering a critical defect [87]. This results in a larger scatter in small highly loaded volumes (*size effect*). The same applies to mesoscale specimens, depending on the ratio between specimen dimension and size of microstructural units (grains) in polycrystalline materials. In this case, the specimen orientation distribution (macroscopic texture) often is not representative of the overall material orientation distribution. In mesoscale specimens, microscopic plasticity influences the overall specimen deformation more sensitively. When the scale of the specimen approaches that of grains, *scale effects* become relevant. These are associated with altered physical mechanisms governed by the free surface or interfaces, e.g. pronounced dislocation irreversibility due to free surface interactions in thin film materials [88].
- **Loading**: The loading amplitude plays a crucial role as it affects all stages and mechanisms of fatigue. It determines whether plasticity can be accommodated by the quasi-reversible glide of dislocations [58], whether strain hardening due to dislocation structure formation occurs, and whether cracks can transition through obstacles such as grain boundaries [89]. Especially body-centered cubic metals and alloys show strain rate effects. Due to the character of screw dislocations in these materials, these are usually immobile (=sessile) at room temperatures and high excitation frequencies used for fatigue [58]. This affects the plasticity onset and irreversibility. Mean offset stresses applied aside from the cyclic amplitude affect crack opening or closure and thereby can alter the sustained life substantially. Residual stresses introduced through fabrication are often detrimental to fatigue life.

- **Ambient conditions**: Changes in temperature alter the defect mobility of dislocations and different point defects. Therefore, thermally-activated processes such as diffusion are emphasized. Furthermore, the tendency of oxidation and corrosion is increased at higher temperatures and specific gas compositions. Thomson et al. [90] and Boyce et al. [91] investigated the influence of oxidation at slip bands. The mechanism found was that metal atoms exposed during the slip in the forward cycle undergo oxidation and are then incorporated into the bulk during the reverse cycle. This results in plasticity-induced oxidation at the free surface of slip bands where oxides often substantially thicker than the native oxide form [91]. Under vacuum conditions often longer lifetimes and smaller cyclic crack growth rates are reported, e.g. for chromium and molybdenum-alloyed martensitic steel [92]. Atmosphere composition can be an important influence factor as well. For instance, the presence of hydrogen can affect the mobility of defects or enhance local plasticity and thereby cause material embrittlement [93].
- **Material-intrinsic**: The internal structure of the material i.e. the microstructure and the defect distribution governs HCF characteristics to a large extent. The interaction of different lattice defects is the origin of cyclic irreversibilities which culminate in failure. For instance, the size of grains plays a decisive role in grain boundary strengthening described by the Hall-Petch relation. [94, 95]. Grain boundaries, depending on their exact character, can pose barriers to dislocation motion resulting in so-called dislocation pile-up at boundaries. The amount of dislocations in a pile-up is proportional to the grain size and since the dislocations in such a configuration typically exert a repulsive force onto each other, grain boundaries pose a smaller barrier in large grains. Grain boundaries also pose obstacles for crack growth where the Schmid factor difference with the most favorably oriented slip system in the grain ahead of the crack tip is attributed significance [89]. The alloying composition, aside from obvious changes in the atomic structure modifying dislocation slip, changes the oxidation tendency and propensity for dynamic strain aging. Latter describes the interaction of interstitial or substitute solutes in the crystal with pinned or relatively immobile dislocations. For instance, provided time to diffuse, carbon can form so-called *Cottrell atmospheres* around dislocations hampering their initial mobility. Furthermore, in highly-alloyed materials in particular, when the solubility limit of the host lattice is exceeded inclusion formation, precipitation as well as grain boundary segregation [96] can occur.

Depending on the material, the fatigue mechanisms and thus the sites for crack initiation within the microstructure differ. While in single-phase materials the PSB-matrix interfaces often pose favorable sites, in brittle and highly-alloyed materials inclusions represent the critical elements. In this work, the focus is set on the initial stages of fatigue, covering slip band formation, crack initiation, and the propagation of microcracks. For these stages, a comprehensive mechanistic understanding is still lacking.

2.1.4 Damage modeling

In the past decades, a variety of damage models were developed focusing on different stages of fatigue. Moreover, different approaches were developed, where analytical models were adopted in computational efforts [97]. Computational methods that account for microstructure-defect interactions range from first principles simulations at the atomic length scale (e.g. to model dislocation-GB interactions [98, 99] or cracking [100, 101]) over the defect scale (e.g., discrete dislocation dynamics of double kink formation in bcc [102]) to phenomenological or physics-based crystal plasticity simulations to predict fatigue damage and crack growth in polycrystals [103, 104]. Additionally, there are data-driven approaches emerging that attempt to predict crack paths or the crack driving force [105, 106], however, are hampered by the lack of data.

In the following, analytical models to estimate crack initiation are introduced, which extend upon Basquin or Coffin-Manson type relations by considering microstructural entities or resolved stresses.

Tanaka and Mura crack initiation model. The microstructure-based fatigue crack initiation model proposed by Tanaka and Mura [107] predicts crack initiation life N_i in favorably-oriented surface grains and is based on irreversible motion and pile-up of dislocation dipoles on two coupled parallel slip planes within planar slip bands. Each of the slip planes is assumed to accommodate either forward or backward dislocation motion in an irreversible fashion. This model relies on the energy

balance assumption, that crack initiation occurs once sufficient planar cyclic slip or stored strain energy is accumulated to exceed the slip band fracture energy.

$$(\Delta\tau - 2k)N_i^{1/2} = \left(\frac{8\mu_s W_s}{\pi d}\right)^{1/2} \tag{2.6}$$

In this equation, $\Delta\tau$ is the shear stress range on the slip system, and k is the corresponding friction stress that needs to be overcome by a dislocation. The shear modulus μ_s, specific slip band fracture energy per unit area W_s, and the grain size d complete the formulation. This model postulates an inverse relationship between the grain size of the critical grain and crack initiation life. The degree of slip reversibility is not considered explicitly in this formulation. On the contrary, slip is considered to be fully irreversible by utilizing a higher frictional stress for the slip plane responsible for reverse motion.

The parameter W_s can be used interchangeably with the specific grain boundary fracture energy per unit area W_G if intercrystalline crack initiation is the relevant mechanism. Subsequently, the model was adapted to incorporate crack initiation at inclusions [108]. The dependence of friction stress on strain rates or temperature is not considered in these models.

Chan crack initiation model. Chan [109] expanded upon the crack initiation model proposed by Tanaka and Mura for crack initiation at slip bands and inclusions. The objective was to include the crack length at nucleation into a closed-form crack initiation model such that it can be linked with crack growth models. In order to achieve this, the Gibbs free-energy change formulation for the nucleation of a fatigue crack from a double pile-up of dislocation dipoles according to Mura and Nakasone [110] was employed.

$$(\Delta\sigma - 2Mk)N_i^{\alpha} = \left(\frac{8\mu_s^2 M^2}{\lambda\pi(1-\nu)}\right)^{1/2}\left(\frac{h}{d}\right)\left(\frac{c}{d}\right)^{1/2} \tag{2.7}$$

In contrast to the model of Tanaka and Mura, here the exponent α is generalized to the range $0 < \alpha \leq 1$ and proposed to be dependent on the stacking fault energy [111] and slip irreversibility. The parameter M represents the Taylor factor which is the conversion factor between normal stresses and shear stresses in polycrystalline materials. Furthermore, the Poisson's ratio ν, slip band width h, surface crack length of a semi-circular crack c, and $\lambda \approx 0.005$ [112] is employed. As a threshold for the cyclic accumulation, the factor $2 \cdot M \cdot k$ is used and in practice substituted by the fatigue limit of the material [109]. However, slip and crack initiation can occur below the macroscopic fatigue limit, which can rather be perceived as a limit to crack growth. Both models are based on the notion of planar slip materials, where dislocation cell and PSB formation are inhibited [109]. Therefore, its mechanistic validity in materials exhibiting multiple slip is debatable. Moreover, the grain boundary character at which the parallel slip bands impinge is not considered.

While these works represented major advances, the analytical description of all microstructural and extrinsic influence factors is a challenge that remained unsolved. Instead, models were gradually adopted in computational approaches to tackle the extensive complexity.

Computational modeling. State-of-the-art in modeling of fatigue crack initiation in polycrystalline materials are computational methods such as the crystal plasticity finite element method (CPFEM). These approaches resolve externally imposed loads on a microstructural level and are able to capture stress concentrations arising from elastic and plastic incompatibilities between adjacent grains. While phenomenological crystal plasticity approaches utilize stress and strain tensors as state variables, physics-based approaches consider dislocation densities. In order to estimate the vulnerability to fatigue, these state variables are translated into so-called fatigue indicator parameters (FIPs) [97]. Among the most frequently used FIPs there is the dissipated energy [113], and the Fatemi-Socie critical plane metric [114]. Recently, a so-called *stored energy density* FIP was proposed and its potential demonstrated [19, 115]. This FIP is based on the idea that the portion ζ of the plastic deformation energy per unit volume and per cycle $\dot{U} = \oint_c \sigma : d\varepsilon^p$, which is stored in dislocation structures (and not dissipated

by heat) causes crack nucleation when accumulated over multiple cycles. It relates the stored energy portion per cycle to a storage volume, which is computed by the reference area and it's computed (or measured) statistically stored ρ_{SSD} and geometrically necessary dislocation density ρ_{GND} [115].

$$\dot{G} = \frac{\dot{U}\Delta V_s}{\Delta A_s} = \oint_c \frac{\zeta\sigma : d\varepsilon^p}{\sqrt{\rho_{SSD} + \rho_{GND}}} \tag{2.8}$$

While this metric computes the stored energy density per cycle, it is multiplied by the number of cycles to measure it against a critical value associated with crack nucleation. It predicted fatigue crack nucleation life in ferritic steel accurately and its localization at inclusions correctly [19].

2.2 Characterization techniques

2.2.1 Electron backscatter diffraction

Electron backscatter diffraction (EBSD) is a characterization technique typically employed in a scanning electron microscope (SEM) in order to gain knowledge on the so-called microtexture of a specimen. Microtexture involves spatially-resolved crystallographic orientations from which grains can be reconstructed to estimate their shape, GB misorientations, and crystallographic phase. Especially in the context of a fatigue study, such full-field information is crucial to infer microstructural influence factors.

Electrons are accelerated as a high-energy incoherent beam towards the specimen which is tilted by typically 70° for a rather small electron incidence angle. A schematic illustration of the typical setup used for EBSD is depicted in figure 2.5. The high incidence angle results in a superficial electron interaction volume which renders the EBSD technique highly surface-sensitive. Upon interacting with the specimen surface, electrons are diffracted. As a matter of fact, low miller index lattice planes, for which the Bragg condition is fulfilled, are channeling the incident electrons partially. A phosphor screen collects the impinging channeled electrons which form the so-called electron backscatter patterns (EBSPs). Imaging of the phosphor screen with a sensitive camera results in image representations of the EBSP. During the EBSD measurement, the specimen surface is sampled with a regular grid culminating in a distinct EBSP pattern at each grid position. These EBSPs are affected by the grain orientation, the present allotrope, residual strain, local defect density, and surface condition. Latter dependency results in the requirement of a highly polished specimen surface absent of deposited residuals. In practice, this prerequisite is achieved by preceding an EBSD scan by electropolishing or chemo-mechanical polishing. Depending on the surface quality, material, and present lattice distortions, a specific exposure time is prescribed which determines the time for which the electron beam remains at each grid position.

In EBSPs there are straight bands present referred to as Kikuchi bands (see figure 2.5) which are related to the gnomonic projection of the various lattice planes. The width of these bands is inversely-proportional to the lattice distance.

Based on these EBSPs, different image post-processing routines can be utilized to derive crystal orientation or further information. Typically, an edge detection algorithm followed by a Hough transformation is performed in order to localize the bands and transform them into the Hough space [116]. A set of phases assumed to be contained in the material is prescribed. During so-called *indexing*, dictionaries of these phases and their Hough peaks are matched with the experimentally derived EBSP in Hough space to obtain orientation measurements. Typically this procedure achieves an absolute orientation error and relative disorientation error down to 2° and 0.5°, respectively. Extending upon this approach, more elaborate techniques were introduced. For instance, HR-EBSD [117] relying on cross-correlation functions between two adjacent EBSP were employed to increase the orientation measurement accuracy and therefore enable the measurement of lattice rotations, elastic strains,

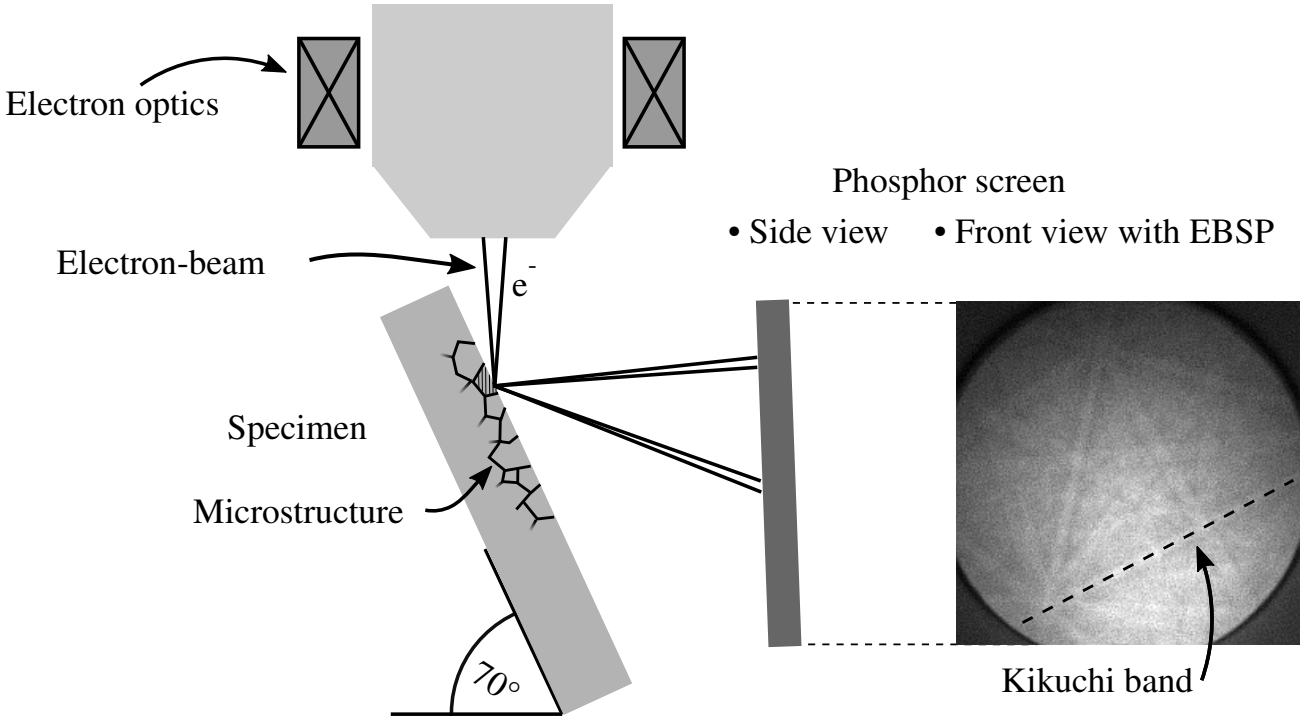

Figure 2.5: Schematic EBSD setup and an EBSP.

and geometrically necessary dislocations (GND) densities. Moreover, convolutional neural networks operating on EBSP have been employed to measure crystal orientations [118].

Consequently, a file is obtained which specifies three Euler angles in the Bunge convention ϕ_1, Φ and ϕ_2 defining the grain orientation with respect to the specimen coordinate system at each sampled position in the scanning coordinate system. Depending on the vendor further metrics, such as phase identifiers, an image quality metric (IQ), and a confidence index (CI) are supplied. In [119], the capability of different image quality metrics to provide grain boundary, strain, phase, and topography contrast is discussed. While each considered IQ metric resulted in distinct contrasts, an image quality metric computed as the average Hough peak height showed the best description of strain and grain boundaries and hence is frequently found in commercial systems. Therefore, the image quality metric can be perceived as a measure of EBSP band quality. The confidence index quantifies the confidence in the selection of a specific orientation based on a voting scheme [120]. Since for some materials containing substantial deformation and internal stresses poor pattern quality can be obtained, different techniques such as neighbor pattern averaging schemes were introduced [121]. In these cases, prior to data analyses, another post-processing step is often required to clean and reconstruct the data and grain structure. The highly quantitative data set can then be analyzed to derive the macroscopic texture of a material, local orientation-related metrics, and various morphological microstructure metrics. As an example, the EBSD map of a martensitic steel shown in figure 2.6 depicts the sample coordinate system normal direction (ND) in the *local* crystal coordinate system in inverse pole figure color-coding.

Furthermore, in terms of phase distinction, phases with an identical or similar crystal structure are difficult to distinguish. This applies amongst others to ferrite, martensite, and bainite phases commonly occurring in steels. The distribution of these phases and the local morphology affects the crack initiation processes and thus is relevant information. Bainite is a phase mixture that consists of ferrite and a phase exhibiting a high carbon concentration, the latter being either cementite or austenite. While the distinction of ferrite and the retained or reverted austenite is state of the art due to the distinct crystal symmetry, it is not trivial and part of ongoing research to differentiate between ferrite, martensite and bainite with EBSD. Approaches to differentiate martensite and bainite from ferrite utilized grain average values of another EBSP quality parameter referred to as *band slope* [122, 123]. Typically grains of bainite and in particular martensite display a smaller band slope due to the contained lattice distortions. Nevertheless, since EBSP quality metrics additionally show a pronounced dependence of surface quality and contamination the discrimination between phases based on quality metrics can be ambiguous. An alternate approach

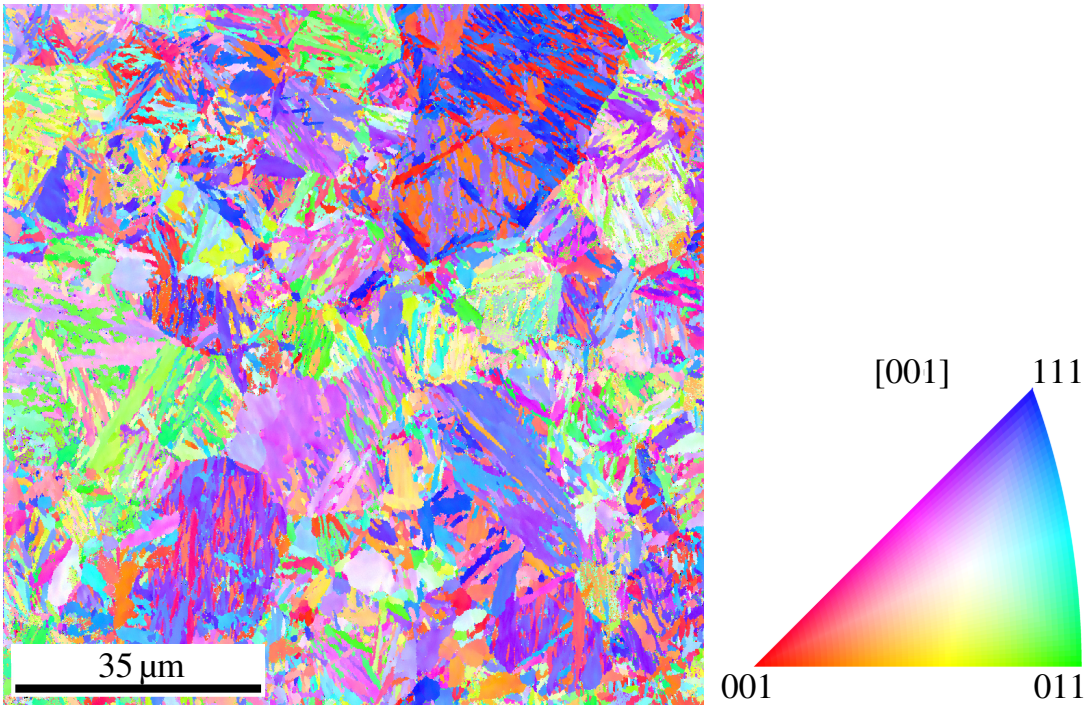

Figure 2.6: Exempary hierarchical microstructure of a martensitic steel. The pixel color in the map describes the orientation of the sample normal direction in the measured crystal coordinate system of each pixel according to the depicted reference triangle. Therefore, each pixel in the map is derived from a single EBSP.

proposed in [124] uses the kernel average misorientation (KAM) metric. This metric relies on orientation measurements rather than on raw EBSP and describes the average of local disorientations between the center pixel and each surrounding pixel contained in a kernel of size $N \times N$.

$$KAM_{i,j} = \frac{1}{N^2 - 1} \sum_{(k,l) \in N_N(i,j)} \omega(g_{i,j}, g_{k,l}) \tag{2.9}$$

Here $\mathrm{N}_N(i,j)$ denotes the subset of neighbor pixels for the center pixel at position (i, j) and $\omega(g_{i,j}, g_{k,l})$ the disorientation between the center pixel and another pixel within the kernel. The KAM metric has been employed to characterize dislocation structures within grains. In order to avoid that sub-grain boundaries, with their comparatively high disorientation, conceal dislocation structures, typically an upper bound for the disorientation value is applied.

A metric that characterizes grain-level average disorientations is represented by the grain orientation spread (GOS). This metric leverages the mean grain orientation $\overline{g_n}$ by averaging the disorientation at each pixel contained in the grain with respect to it $\omega(\overline{g_n}, g_p)$.

$$GOS_n = \frac{1}{N_p} \sum_{p=1}^{N_p} \omega(\overline{g_n}, g_p) \tag{2.10}$$

The subscripts n and p are indices for the grain and pixel, respectively.

Since EBSD is a surface-sensitive technique, information such as the grain boundary orientation and the 3D grain shape can not be derived unless serial sectioning is applied. Serial sectioning refers to the sequential removal of material which

is combined with conventional EBSD measurements to sample along the third dimension. The removal is often induced by (plasma) focused ion beam (FIB), laser ablation [125] or chemo-mechanical polishing [126].

2.3 Statistical methods and machine learning

Recently, machine learning (ML) and deep learning (DL) is being introduced to material science in order to solve various problems as summarized in numerous surveys, e.g. [127–129]. Amongst others, problems approached by data-based approaches comprise microstructure classification [130, 131], synthetic microstructure generation [132], steel phase segmentation [133], crystallographic defect segmentation [134], crack detection [135], prediction of materials properties from atomistic structures [136] or from microstructure [137–139], predicting mechanical load distribution [140] or crack paths [105, 141] in polycrystals, and identifying relevant microstructural features for crack propagation [106].

ML algorithms attempt to learn representations from training data. The process of model optimization is referred to as *training*. In research questions where complexity is pronounced and the knowledge is incomplete, ML approaches can outperform knowledge-driven models in terms of achieved accuracy by large margins. For instance, deep learning variants superseded rule-based computer vision (CV) models in many areas [142]. The reason for the superiority of ML in such scenarios lies in their pronounced expressivity, i.e., the capability of describing highly nonlinear relations in high dimensional feature spaces which facilitates taking account of even nuanced correlations.

In so-called supervised learning, a mapping between input variables X (*features*) and a supervisory signal Y (*target variable*) is extracted from a training set of X-Y pair examples (*instances*) by applying dedicated algorithms. Aside from supervised learning approaches, partially supervised methods (semi-supervised learning), unsupervised learning, and self-supervised learning exist. These methods attempt to learn regularities and patterns in the data despite the (partial) absence of the supervisory signal Y. This is motivated in the fact that target variables in many cases are difficult to obtain and sometimes require extensive manual annotation efforts. Therefore, by avoiding the necessity of target variables, the available data is increased severely. Self-supervised learning attempts to learn inherent regularities in the data through artificially constructed supervisory signals from the data itself, so-called *pretext tasks*. For instance, predicting words that were removed from a sentence is a classical pretext task in natural language processing (NLP) that facilitates learning a linguistic syntax. Models *pre-trained* on such pretext tasks are subsequently *fine-tuned* in another subsequent training routine to a *downstream task*. The general procedure of re-training a pre-trained network to benefit from the prior learning is commonly referred to as *transfer learning* and was shown to increase model generalizability, accelerate convergence, and improve performance in case of low data availability for downstream task [143]. During the past decade, applying large annotated data sets such as ImageNet [144] has become a standard procedure in pre-training. In contrast, fine-tuning of self-supervised pre-trained models was recently demonstrated to culminate superior NLP and CV downstream tasks performance in comparison to fine-tuning their supervised counterparts or direct supervised learning on the downstream task [145–147]. This can be attributed to the access to substantially more training data in conjunction with appropriate model capacity [147], and the capability of extracting superior feature representations from strong supervisory signals [146, 148].

As hinted at in the aforementioned materials science applications, there is a range of tasks that can be addressed through machine learning approaches, including the following, to name a few.

- **Regression**: Prediction of a continuous quantity (from the set of real numbers), e.g. number of fatigue cycles to failure.
- **Clustering**: Clustering of data instances in an unsupervised fashion depending on specific distance metrics.
- **Classification**: Prediction of the semantic category of each data instance X. Special cases include binary and multi-class classification.
- **Segmentation**: Assignment of a semantic category to each pixel in an image. Special cases include binary, semantic, instance and panoptic segmentation.

- **Detection**: Object detection refers to the task of locating objects of defined semantic categories in images or higher order tensors by retrieving their bounding boxes.

The models applied to address these tasks differ strongly in their complexity (i.e., model capacity) and thus interpretability [149]. There are models such as linear regression or decision trees which offer algorithmic transparency, comparatively high degree of decomposability and even simulatability for smaller variants [150]. The latter two notions refer to the comprehensibility of individual model parts (e.g. nodes in a decision tree) and the whole model, respectively. In contrast so-called deep learning (DL) methods, which represent a sub-class of ML, are often characterized by their high expressivity at the expense of poor interpretability and are therefore commonly referred to as black-box models. However, there are some fairly translucent [151] approaches to analyze DL methods' internal operating principle to achieve decomposability such as GradCAM [152] or NetDissect [153]. As the name implies, deep learning methods such as multi-layer perceptrons (MLP) and convolutional neural networks (CNN) often exhibit high network depth, i.e. apply a series of transformations to the data, see section 2.3.4. In the past two decades, a manifold of highly specialized DL network architectures was tailored to address specific tasks [154]. The network size, training approach, and problem complexity dictate the amount of necessary training data. Common applications of DL methods comprise classification, detection, and semantic segmentation.

Aside from model interpretability, there are two major remaining technological challenges impeding the increased utilization of deep learning in the material science field:

- **Limited data-efficiency**: Human cognition permits learning new concepts through the observation of a few examples. Deep learning models, in contrast, require a vast amount of instances [25]. The limited data efficiency adversely affects domains such as materials science in particular owing to the expense associated with acquiring and annotating data. This challenge is amplified by the inherent imbalance of some problems. For instance, as a consequence of fatigue with low loading amplitudes, the plasticity onset is exceeded only in a few grain ensembles culminating in a scarcity of damage instances. The limitation in data efficiency has led to the emerging research field of few-shot learning. Moreover, aforementioned transfer learning concepts in ML [155] and DL [156] alleviate this problem to some extent.
- **Poor generalization capability**: Another critical drawback of supervised learning techniques is their limited transferability to alternate domains, i.e. to achieve acceptable performance on out-of-distribution instances [25–27, 150]. An example brought up frequently in this context is the failure of CNNs to detect mammals out of their usual habitat (e.g. cow on a beach) [25]. This problem is associated with so-called *spurious correlations* that the network learns (grass $\rightarrow$ cow). Analogously, this raises the question of whether a classifier trained to detect martensite in dual-phase steel can discern martensite from other phases in complex-phase steels. However, annotating the same entities in distinct environments is infeasible considering the variety of materials that exist. Within the broad concept of transfer learning, there are transductive approaches [156] which specialize in transferring between data domains despite the unavailability of target domain annotations. This sub-field is referred to as domain adaptation [157].

In the work at hand, both ML and DL are applied for four tasks. While addressing these tasks, emphasis is put on tackling the aforementioned challenges. The data set creation for most of these tasks is a central aspect of this work.

1. Semantic segmentation of fatigue damage, namely protrusions and cracks, in secondary electron SEM images of single-phase steel. Therefore, a CNN architecture, designed for dense prediction tasks called U-Net was trained in a supervised fashion.
2. Identification of relevant microstructural and micromechanical features affecting the localization of fatigue damage nucleation. Since the objective lies in extracting relations within the data, conventional ML with lower model complexity and increased interpretability such as decision trees are suited particularly. Decision tree algorithms have internal feature importance metrics and allow for analyzing the interdependence of features in a multivariate feature space by analyzing the structure of the learned tree, see sections 2.3.1 and 2.3.2.

3. Prediction of damage sites within the microstructure given a microtexture and load distribution. Therefore, the balanced random forest algorithm which extends upon the decision tree algorithm is used. Similar approaches were previously used by [137, 138] to predict twinning events and by [139] to predict stress hotspots in crystallites.

2.3.1 Decision tree algorithm for classification

A decision tree is a learning concept that takes data in a tabular setting. Data instances are arranged in rows while features as well as the respective labels (target variable) constitute columns. The features span an M-dimensional orthonormal basis and can either be of categorical or continuous data type. Depending on whether the target variable Y is categorical or continuous, the tree can be distinguished into classification and regression type. During training, the decision tree picks some of the M features and performs orthogonal splits in the feature space in order to optimize the information gain. Thereby, the algorithm classifies the data in different groups and a decision tree is constructed. Figure 2.7 depicts a schematic decision tree.

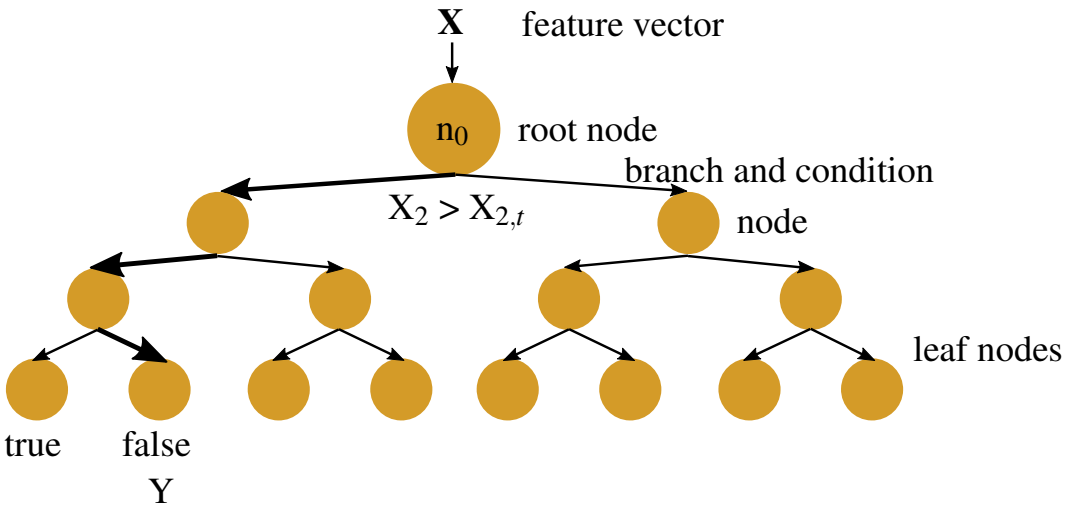

Figure 2.7: Schematic decision tree composed of nodes, branches, and leaf nodes. Depending on how the branching conditions are met, the instance of a feature vector can result in different leaf nodes.

A decision tree consists of a root, nodes, branches, and leaves. The root is the node, where the first decision (split) takes place. In figure 2.7, when feature two X_2 of an instance exceeds the threshold value for feature two $X_{2,t}$, the instance is passed to the left branch. During the training procedure, the algorithm decides on which feature and which respective threshold to split at each node. Thus, nodes are junctions present in the decision tree, at which the instances are separated into different groups (branches) depending on whether the branching condition is met or not. The terminating nodes, after which no branching occurs, are denoted as leaf nodes. The algorithm is recursive since the branching conditions need to be set for every node except the leaf node. In the case of binary classification, the outcome at each leaf could be true or false. If a certain feature for splitting appears close to the root node and/or is used frequently, it indicates substantial relevance of that feature with regard to the target variable.

There are two widely used algorithms for decision trees called Gini-Index and ID5. In the latter case, the decision, of which features and conditions to utilize for the data splits, happens by information gain maximization. Therefore, the Shannon entropy

$$H(Y) = \sum_{c \in C_Y} -p(c) \cdot log_2 p(c) \tag{2.11}$$

is considered with c, C_Y, and $p(c)$ denoting a class, a set of classes in Y, and the probability of that class to occur in the label, respectively. The entropy is a measure of the impurity of a label and amounts to unity if the possible classes in it (e.g., true and false) are distributed evenly and zero if there is no variance in the label. The conditional entropy

$$H(Y \mid X_i) = \sum_{c \in C_X} -p(Y|_c) \cdot H(Y \mid X_i|_c) \tag{2.12}$$

is computed for every feature X_i in the data set and every split on feature X_i. Here, $H(Y \mid X_i|_c)$ indicates the post-split entropy and C_X the classes after a split. In the case of continuous feature data types, typically midpoints between the ranked features' numeric values are evaluated as split candidates in equation 2.12 to determine the ideal splitting threshold.

In order to compute the information gain, also referred to as Kullback-Leibler divergence $DD(Y \parallel X_i)$, characteristic for the feature, the entropy of the labels $H(Y)$ before the split is subtracted by the conditional entropy $H(Y \mid X_i)$ for each feature and split on it.

$$DD(Y \parallel X_i) = H(Y) - H(Y \mid X_i) \tag{2.13}$$

The Kullback-Leibler divergence is a measure of mutual information between the target variable Y and a feature X_i and measures the distance between two density distributions [158]. The boundary at which the feature is separated is the one at which the information gain is maximal. The whole procedure is performed recursively to define multiple branching conditions.

By default, decision trees are rather prone to so-called *overfitting*. Overfitting describes the state when a model learns features in the data that are specific to the training data set (e.g., noise characteristics) and hence is unable to generalize to other data sets. This problem is often associated with models that have too high expressivity with respect to task complexity and data quantity. Different strategies for preventing overfitting are available. In decision trees, hyperparameters such as the maximum tree depth, the minimum required data instances for a node split, and the number of leaf nodes can be specified. Moreover, there are pruning methods in which certain branches, which are based on features having a low impact on the prediction, are discarded.

Apart from pruning, different concepts have been proposed utilizing ensembles of decision trees to achieve better performance and generalization, eventually resulting in the random forest algorithm.

2.3.2 Random forest algorithm

Ensemble learning methods rely on the concept, that by unifying many weak learners, a good predictive performance can be achieved [159]. Based on this, in [160], an algorithm named AdaBoost was introduced, relying on the concept of *boosting*.

However, the random forest algorithm as it is utilized in most implementations exploits a slightly different concept. As the name implies, the random forest algorithm [161] *aggregates* predictions of multiple decision trees (ensembles), see figure 2.8. Rather than constructing a multitude of decision trees on the complete data, subsets of the data are used to grow weak classifiers. Namely, features and instances are randomly sampled with replacement (*bootstrap samples*) and provided to each decision tree, which is the other reason for the designation *random* forest. The combined principle of aggregating the predictions of weak classifiers, which were trained with random data subsets, is denoted as bootstrap aggregating (bagging). Random forest models typically surpass decision trees in terms of prediction accuracy for data sets, where the number of instances exceeds a few hundred [162]. Furthermore, bagging mitigates the issue of overfitting. In the case of classification forests, a majority vote of all trees is performed. On the other hand, regression forest models average the values obtained by the individual trees to predict continuous variables.

In the case of M total features, m features with $m \ll M$ are sampled at each node, and the best split among them is computed. The value of m is constant for all nodes in every tree. Contrary to individual decision trees, the trees within a random forest are grown deeper and their pruning is omitted.

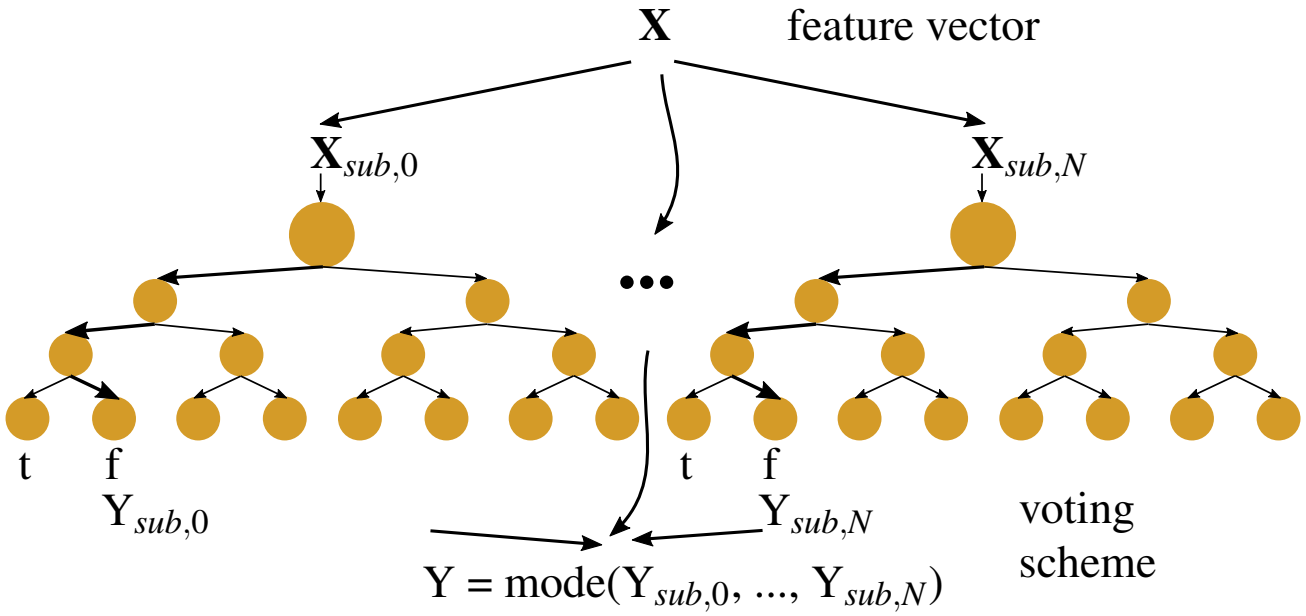

Figure 2.8: Schematic random forest composed of multiple decision trees fed with subsets of the data.

Models have been developed that are capable of handling imbalanced data sets, where the classes among labels are distributed unevenly [163]. This is required since typical classifiers achieve substantially higher accuracy on the majority compared to the minority class [164]. This skewed behavior can be attributed to the characteristic of classification algorithms to minimize the overall error rate. Therefore, there is demand for algorithms that put emphasis on the minority class, because they are frequently of high importance (e.g., rare instances of diseases or scarce microstructure-scale fracture in materials fatigue). Common approaches to tackle this issue comprise artificially balancing the data set by over-/ under-sampling the minority/majority class or applying increased class weights to minority class instances to take them into consideration more during training. For instance, the SMOTE algorithm, inspired by data augmentation techniques, oversamples the minority class by synthesizing minority instances [164]. A random forest incorporating the aforementioned balancing concepts was proposed in [165]. An implementation thereof is provided by the *imbalanced learn* python package [166], where under-sampling of the majority class for each bootstrap sample is performed and optional class weights can be applied.

For model training, the available data is split into training, validation, and test subsets. The validation set facilitates monitoring the model performance during training on unseen instances and can give indications of overfitting (e.g. through an increase in the validation loss). The test set is necessary since the validation set is used during model optimization and hyperparameter tuning. When data scarcity and pronounced imbalance play a role, drawing conclusions from ablation studies is hampered. A commonly applied practice when training models is so-called k-fold cross-validation. The complete data set is sampled into k partitions out of which k training and test sets are constructed. This allows investigation of the influence of sampling when training the model to recognize selection bias. *Stratified* k-fold cross-validation attempts to sample the data such that each partition contains equal class label distribution.

2.3.3 Feature importance and feature selection

Aside from data curation, feature engineering and feature selection are the initial steps in designing a machine learning model. Feature engineering describes the process of deriving features using existing domain knowledge or hypotheses to facilitate a more descriptive feature space. This step is often followed by a feature selection step which serves the purpose of obtaining a feature subset with the highest relevance and lowest redundancy (avoiding correlated variables) regarding a target variable. It attempts to avoid unnecessary model complexity by discarding irrelevant features, which ultimately increases model interpretability and prevents overfitting. While their objective is similar, so-called dimensionality reduction techniques, in contrast, to feature selection approaches, transform the data into a lower dimensional space. For instance, in

principal component analysis (PCA), a multivariate feature space and the variance therein are expressed efficiently by an orthonormal basis with lower dimensionality than the original feature space. However, criticism posed to these techniques is that the projections into the principal component space hamper physical interpretability [167]. Moreover, importance metrics exist that aim to deduce an insight into data or the model relations.

2.3.3.1 Tree-based feature importance measures

Many commonly used decision tree or random forest implementations contain embedded feature importance measures, which are summarized in [168]. For instance, the mean decrease in impurity (MDI) is frequently utilized [169, 170]. For calculating the MDI of a certain feature, every node where the feature is used for a split is assessed. For each of these nodes, the product of the sample fraction reaching the node with the impurity decrease at that node is computed. Finally, the sum of these products represents the MDI for that feature. In case of tree ensemble models, the MDI is either averaged or cumulated over all trees depending on the implementation.

Another feature importance metric incorporated in tree-based models is referred to as permutation accuracy or mean decrease in accuracy (MDA) importance [169]. As the names imply, this measure relies on random permutations of single features X_j, such that its relation to the target variable Y is nullified. When this permuted feature along with the other unpermuted features are used for prediction, a reduction in prediction accuracy representing the MDA is expected, depending on the prior correlation of X_j and Y. In contrast to the MDI, the MDA measure is typically less biased towards features exhibiting a pronounced cardinality. Another advantage over the MDI is, that the MDA is typically evaluated on the test set rather than during training. However, the MDA measure was shown to overestimate the importance of correlated input variables [171].

2.3.3.2 Feature selection methods

This section aims to provide a brief overview of different feature selection techniques. These techniques attempt to identify the smallest feature subset in the data set that contains relevant information with respect to the target variable, i.e. the active predictors, before model selection and training. The appropriate selection method depends on the number of features, their types, correlations in the data set, noise levels as well as the number of instances [172]. All subsequently mentioned feature selection methods struggle when complicated, non-linear functional dependencies exist and combinations of variables are particularly relevant. In this regard, the aforementioned tree-based feature importance metrics are typically better suited. Since the following heuristic feature selection algorithms were designed to identify relevant features for comparatively less expressive model types such as linear regression, caution is required to avoid the exclusion of relevant features when combining them with stronger models that can extract intricate relations in the data. When feature selection is to be performed in advance, especially for intricate tasks, reviewing and adjusting selected feature subsets based on domain knowledge is crucial.

In the first instance, frequently, a metric denoted as Pearson correlation coefficient is utilized. It poses means to quantify how linear the correlation between a variable X_i and Y is. The Pearson correlation coefficient is computed by dividing the covariance between both variables by the product of their standard deviations. It ranges from negative to positive unity, the boundaries representing either positive or negative perfect linear correlation, respectively. Another common feature selection tool is the least-angle regression (LARS) algorithm. This iterative algorithm determines coefficients to all features in order to compute a linear combination of features that is descriptive of the target variable [173]. These coefficients give an indication of which variable or feature to include into the model. Initially, the algorithm identifies the feature variable X_1 which exhibits the strongest correlation with respect to the target variable Y and adjusts its coefficient β_1 while keeping track of the residual $r = Y - \hat{Y}$ where $\hat{Y} = \beta_1 \cdot X_1$. As soon as the residual is correlated with another feature as much as with the initial feature, both of their coefficients are tuned jointly. This procedure is repeated until all features are considered. While the LARS algorithm is computationally efficient, there are drawbacks when dealing with data sets containing noise or outliers [174] owing to the iterative minimization of the residual. It is argued that the LARS algorithm is less aggressive than other forward selection approaches when it comes to discarding somewhat correlated features that contain useful information [173]. However, this

can lead to the inclusion of highly correlated features [175]. The so-called Lasso algorithm [172] differs from LARS in that regularization of the covariate coefficients is performed by setting the condition that the sum of absolute values of all regression coefficients (i.e. their L_1 norm) is less than a predefined constant. This effectively leads to the suppression of unimportant covariates by setting their coefficients to zero. However, this limits the number of selectable features (limited to the number of data instances) [176]. The choice of the regularization parameter and the data set size can affect the robustness i.e. consistency of selected feature sets. Feature grouping information is not respected by any of the aforementioned feature selection techniques which poses an issue when considering grouped variables such as genes or crystallographic orientation representations. Different extensions have been proposed to tackle these shortcomings [176, 177]. Another extension of Lasso is the so-called FeaLect algorithm [178]. FeaLect applies the Lasso algorithm multiple times on bootstrap samples of the data set. Then a score is computed for each feature that indicates the frequency of how often Lasso incorporates a feature in the relevant subset. This scoring approach differs from [177] where features are required to occur in all bootstrap runs to be considered relevant. Moreover, in contrast to Bolasso [177], Fealect takes the whole regularization path into account. Different empirical studies on real, i.e., non-simulated data sets, indicate the superiority of FeaLect for feature selection [178], amongst others, a study on stress hotspot classification in microstructures [179].

2.3.4 Deep learning

In contrast to previously discussed conventional machine learning approaches, which typically require manual feature engineering, a sub-domain of ML referred to as deep learning (DL) is used for *representation learning*. Here, feature extraction is performed automatically by the algorithm, rather than by the user. Images share comparatively more commonalities (edges or corners are present irrespective of the task) than data sets with highly specific feature vectors. This fact is frequently exploited in transfer learning which finds use in DL particularly. The efficient feature extraction in DL is enabled through the interplay of the key concepts — loss functions, backpropagation, and optimizers, which will be addressed subsequently. Therefore, for example, by supplying examples of input and desired output data to a network in a supervised learning setting, it can learn to put emphasis on certain features while discarding others. Furthermore, these DL methods are prevalent in image or higher-order tensor processing, rather than in tabular numerical and categorical data. The tasks that are commonly tackled with deep learning range from classification over detection to segmentation. Challenges such as object detection in autonomous driving and segmentation of carcinoma in magnetic resonance images rank among the most confronted ones.

Generally, the instances where DL has been applied to material scientific problems and particularly mechanics of materials or fatigue-related challenges are comparatively sparse. In the context of this work, DL has amongst others been utilized to achieve automatic fatigue damage semantic segmentation from scanning electron microscopy images. Semantic segmentation represents a dense prediction task and, in the context of images, describes the assignment of a class label to every pixel. While there are classical rule-based computer vision techniques to perform this task, recent DL models perform substantially better in challenging scenarios where pixel intensities are not sufficient to discern classes. Hence, in the following, the focus is placed on DL fundamentals and concepts that enable accurate segmentation from images. A typical DL workflow includes data acquisition, systematic labeling, model selection, data pre-processing, data augmentation, training, and model evaluation.

2.3.4.1 Semantic segmentation

With the advent of data-driven techniques, neural networks called multilayer perceptrons (MLP) were utilized for segmentation tasks. These MLPs rely on serial connections of building blocks referred to as layers which are in turn composed of multiple so-called perceptrons. Perceptrons map their inputs to an output by computing a weighted sum with its internal weights and its bias. Typically, these feedforward neural networks are *fully connected*, which means perceptrons of a specific layer are linked to each perceptron of the prior and subsequent layer in the network. This along with so-called activation functions permits these models to approximate non-linear relations in the data. However, with the introduction of convolutional neural networks (CNNs), MLPs were superseded in the field of image processing. This can be attributed to two characteristics of CNNs.

- **Memory and training time efficiency:** The analogy to a neural network's perceptrons in CNNs are convolution operations between the input and convolution kernels. Since the comparatively small convolution kernels are slid across the whole input, the weights, now part of the convolution kernel, are shared. The shared weights result in a pronounced memory and training time efficiency as fewer weights need to be optimized. Therefore, during CNN training, multiple images (*batch size*) can be passed to the network simultaneously. Especially as image sizes increase, the amount of parameters to optimize in fully connected perceptron architectures increases rapidly — the fully-connected networks do not scale well with input size.
- **Invariance to image transformations:** Since in CNNs, the same kernels are applied iteratively to all regions of an image, an invariance to translations and other image transformations is attained. This represents an inductive bias introduced by the convolution operation. Such models are referred to as being spatial-agnostic and channel-specific (i.e., each channel in the input to a convolution is treated distinctly) [180].

In the following, the key concepts of convolutional neural networks and CNN semantic segmentation models are introduced.

Convolutional layers: The name CNN implies that convolutional layers are fundamental to the functionality of these networks. These layers allow the extraction of features such as edges or corners from the input images and at later stages of the network the identification of relevant structures in the intermediate tensors. Each convolutional layer includes multiple convolution filters. The size of these filters is a fundamental hyperparameter of a CNN affecting to which extent contextual information is considered at each layer. In figure 2.9a the operation is illustrated.

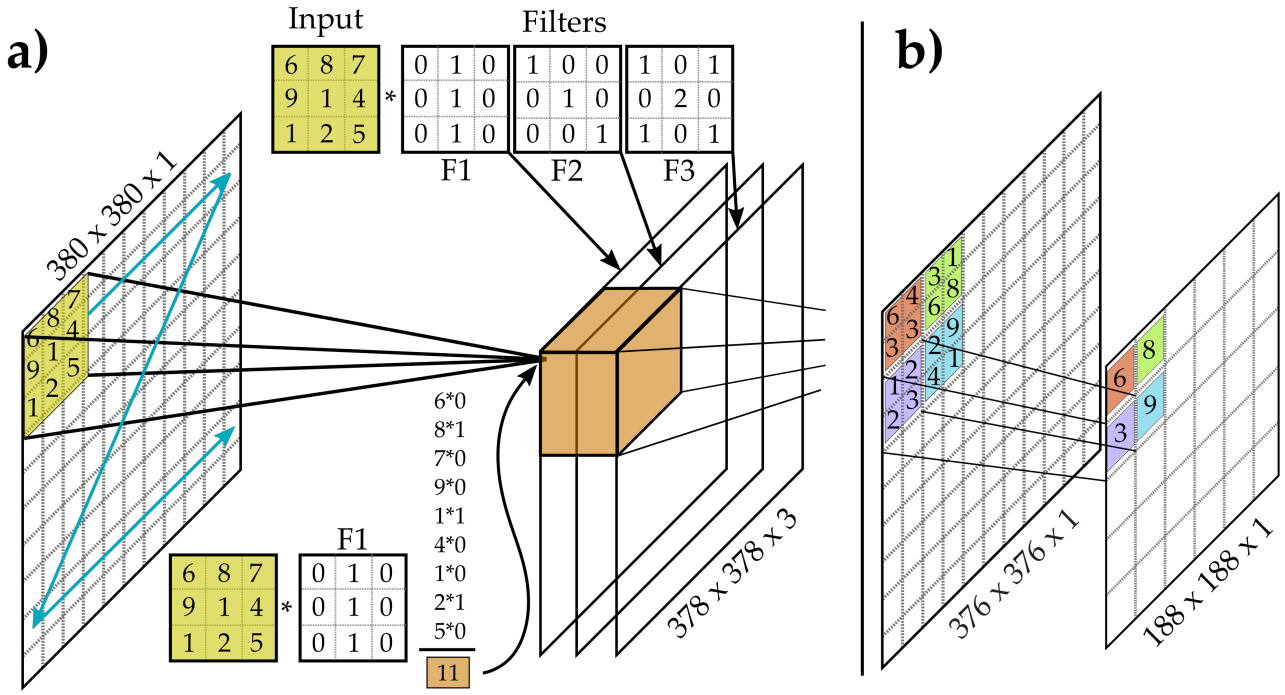

Figure 2.9: Schematic description of **a)** Convolution operation and **b)** Max pooling operation

In a convolutional layer, multiple convolution kernels (filters) are swept across the whole image. A hyperparameter referred to as *stride* indicates how large the pixel step size is at which the filter is moved. This parameter directly affects the feature resolution. The computation of each element in the output volume during the convolution process amounts to the Frobenius inner product of the filter with the region of the input which is swept plus a bias term, see figure 2.9a. When a 3×3 filter is applied to an input image, the lateral dimensions of the output volume are reduced relative to the input. This can be ascribed to the missing information when the center of the filter is applied to the border of the input. Frequently, this problem is alleviated by applying *padding*. Padding describes the artificial expansion of the input through repetition of the border value, appending a constant value, or mirroring boundary regions. Each filter F1 – FX is applied across the whole channel dimension of the

input tensor. Hence, the extent of the channel dimension in the output feature volume is determined solely by the number of filters in the convolutional layer.

Different types of convolution operations were introduced in literature including atrous convolutions [181] and separable convolutions [182]. Atrous convolutions utilize an additional hyperparameter, the *dilation rate* which spreads the filter to a wider range while maintaining the number of trainable parameters. This spreading results in gaps in the filters. In particular, segmentation tasks can benefit from replacing conventional convolutions through atrous convolutions as shown in [183]. Networks are able to produce denser feature maps when pooling steps are discarded and atrous convolutions are employed to compensate for the loss in the receptive field. In order to receive contextual information on multiple scales, multiple atrous convolutions with different dilation rates are applied to achieve an improved segmentation performance. Separable convolutions are interesting when mobile applications where efficiency plays a significant role are concerned. For a more comprehensive description and visualization of convolution operations, it is referred to [184].

Activation functions: Activation functions are utilized subsequently to convolutions in order to enable describing complex relationships better by introducing non-linearity into the network. Typically, activation functions map elements in the feature tensor to an output element-wise. Arguably, the most commonly used activation function is the **Re**ctified **L**inear **U**nit (ReLU) [185], which shows beneficial characteristics compared to hyperbolic tangent or sigmoid activation functions. A common problem, which is encountered when training a deep CNN with latter activation functions, is the vanishing gradient problem. This problem describes the vanishing of error gradients during the training stage leading to diminishing weight updates and therefore slow training. It applies in particular to the initial layers as these activation functions are bound to a range between 0 and 1 successively decreasing the update rates when propagating backward. On the other hand, ReLu mimics the behavior of cortical neurons as these are typically not saturated as pointed out by Glorot et al. [186]. The absence of an upper saturation and of a limitation in terms of the range of values results in ReLu being comparatively less prone to the vanishing gradient problem. Aside from these reasons, training with the ReLU activation is computationally efficient and leads to a sparse activation output resembling the behavior of the human brain. This gives the network robustness against noise [186].

Batch normalization: The batch normalization operation was introduced comparatively recently in [187] to accelerate the training process. The fact that model parameters in each layer are updated simultaneously during training and their adjustments are interdependent impedes the efficient computation of an optimal model state. This manifests in discontinuities and flat regions present in loss functions [188]. Batch normalization renders the inputs to each activation function more stable by normalizing the input distribution. When batch normalization is not applied, input distributions evolve together with prior layer parameters in the course of training. This change in the input distribution of the network activations is referred to as *internal covariate shift* and its reduction was considered the reason for the undeniable effectiveness of batch normalization. However, a recent study [189] suggests that the impact of batch normalization can be ascribed to a smoothing effect on the loss function and gradient of the loss function. Through batch normalization, the vanishing gradient problem is circumvented since the likelihood of an activation to remain saturated is reduced, resulting in comparatively faster learning. In practice, batch normalization facilitates the acceleration of the training process by allowing for higher learning rates. The original implementation [187] utilizes the statistics of mini-batches to compute the mean and variance for each feature channel.

Consider a four-dimensional feature tensor where the dimensions i, j, k, and l represent the batch B, spatial X, spatial Y, and channel C dimensions, respectively. Such a feature tensor exhibits the size $[N_B, N_X, N_Y, N_C]$ and is the output of an arbitrary convolution layer. The computation of the mean μ and variance σ^2 is performed along all dimensions except the channel dimension. Therefore, this operation results in a one-dimensional vector with extent $[\mathrm{N}_C]$ for the mean μ, variance σ^2 as well as for the learnable shift β and scale γ parameters. In equations 2.14–2.17, $x^{(l)}_{i,j,k}$ represent an element of the feature tensor at position i, j, k and (static) l. Hence, this set of computations is performed for each channel in N_C.

$$\mu^{(l)} = \frac{1}{N_B \cdot N_X \cdot N_Y} \sum_{i=1}^{N_B} \sum_{j=1}^{N_X} \sum_{k=1}^{N_Y} x_{i,j,k}^{(l)} \tag{2.14}$$

$$\sigma^{2,(l)} = \frac{1}{N_B \cdot N_X \cdot N_Y} \sum_{i=1}^{N_B} \sum_{j=1}^{N_X} \sum_{k=1}^{N_Y} \left(x_{i,j,k}^{(l)} - \mu^{(l)} \right)^2 \tag{2.15}$$

$$\hat{x}_{i,j,k}^{(l)} = \frac{x_{i,j,k}^{(l)} - \mu^{(l)}}{\sqrt{\sigma^{2,(l)} + \varepsilon}} \tag{2.16}$$

$$y_{i,j,k}^{(l)} = \gamma^{(l)} \cdot \hat{x}_{i,j,k}^{(l)} + \beta^{(l)} \tag{2.17}$$

The shift and scale parameters β and γ per channel fulfill the purpose of driving the output distributions $y_{i,j,k}^{(l)}$ into the saturation range of the activation function for specific channels, whenever beneficial for the learning process. Further, batch normalization introduces robustness to different learning rates and parameter initialization [187]. There are several derivatives of batch normalization in literature including layer [190] or instance [191] normalization.

Pooling layers: In order to keep the network size and thus computational expenses maintainable despite introducing many feature channels in the convolution step, a non-learnable pooling operation can be utilized. Additionally, these layers help to generalize and to attain the translation invariance [185]. There are different types of pooling layers that are commonly utilized in CNNs. In comparison to average pooling, max pooling has been shown to improve performance and result in faster training convergence [192]. Therefore, the max pooling operation depicted in figure 2.9b is applied most commonly. This operation downsamples the feature map in the spatial dimensions (applied equally to all channels without downsampling the channel dimension). In the illustration, the omnipresent 2×2 max pooling is shown. Generally, the pooled regions are directly adjoining without any overlap. The maximum value from within each region is considered in the output in order to preserve characteristic edges and textures in the image. The concept of the *receptive field* describes the input image region that is considered at a specific layer. The extent of image context taken into consideration is significantly affected by the architecture. In particular, the amount of repeated consecutive convolution and max pooling operations in CNN architectures affects it distinctly. Therefore, pooling operations help the network to take image features at different scales into account. However, the contents inside the receptive field are typically not taken into account equally but the emphasis is placed on the center of the receptive field. In literature, this concept was referred to as an effective receptive field [193].

Upsampling layers: In segmentation models, upsampling layers are frequently utilized after multiple convolutions and pooling stages to achieve dense predictions by retrieving the spatial resolution of the input image. A summary of different upsampling strategies can be found in [194]. Here, it can be observed that for the segmentation task no substantial differences arise from the choice of the upsampling strategy. The most commonly utilized upsampling operation is the *transposed convolution*. As opposed to other upsampling strategies such as image interpolation [195], transposed convolution is a learnable operation. Rather than performing an element-wise multiplication between the kernel and input region as in convolutional layers. Each element of the input is multiplied with all kernel elements (learnable), which results in multiple matrices. The positions and connectivity of the inputs are preserved when positioning the matrices in the output. Subsequently, potentially overlapping regions of the matrices are summed which can result in checkerboard patterns in the spatial coordinates of the feature space [196] depending on the stride and kernel size hyperparameters. Similar to the pooling operation, the upsampling operation layer does not affect the channel dimension but operates on all channels. So-called *skip connections* are a common tool used in dense prediction tasks to recover the spatial resolution after the successive compression through max-pooling operations. Such skip connections essentially bypass larger portions of the network to concatenate largely unprocessed feature volumes where spatial information is unperturbed with reconstructed ones [197].

Loss functions: The loss function is computed in the course of model training in order to evaluate the difference between the prediction probability distribution and the supplied ground truth segmentation. For segmentation problems, frequently the pixel-wise cross-entropy loss [198] L_{CE} has been employed whose definition matches equation 2.11. In segmentation

problems, however, this operation is computed and averaged over each pixel. This implies equal weighting for all pixels. In some cases, a pixel or class weighting scheme can be of interest. For instance, when a pronounced class imbalance occurs where the foreground class constitutes only a small portion of the image. The imbalance is accompanied by the issue that the typically dominant negatives do contribute to learning only to a small extent. Furthermore, despite the diminishing loss contribution from straight-forward negative pixels, the sheer mass can distract the model from learning based on the positive and difficult negative samples. Therefore, different approaches have been proposed to tackle the imbalance problem including weighting each class channel in the output distinctly [199].

Furthermore, focal loss [200] extends upon the weighted cross-entropy loss by introducing a modulating factor $(1-p_t)^\gamma$. It is described by the authors as a "dynamically scaled cross-entropy loss", which puts emphasis on difficult samples while reducing the impact of simple instances.

$$L_{FL} = -\alpha_t (1-p_t)^\gamma \log_2 p_t \tag{2.18}$$

Here $\alpha_t \in [0,1]$ describes a weighting factor to address the class imbalance and the focusing parameter $\gamma_t \geq 0$ a weighting factor describing how much emphasis is put on difficult examples compared to trivial examples. However, the standard implementation of the focal loss does not address the need for multiple α values in the case of semantic segmentation where multiple foreground classes contain relative imbalance to each other.

Another question concerning loss functions is pixel-based loss weighting for accentuating image proximity-related weighting. In [197] the problem of close foreground objects separated by a small gap was tackled by applying higher pixel weighting at the border pixels. To be precise, pixel weights were sampled from a Gaussian distribution where the independent variable was the sum of the euclidean distances between a specific pixel and the closest two foreground object borders.

Backpropagation: An indispensable contribution to the functionality of neural networks is made by backpropagation as it enables the iterative optimization of network parameters. Similar to the loss function, it is rather a methodology than a core part of the network architecture. It utilizes the computed loss and propagates the error gradients backward by applying the chain rule to compute partial derivatives with respect to the trainable network parameters in all layers. Obviously, this necessitates that every operation implemented by the network architecture is differentiable. On the basis of the computed partial derivatives, an *optimizer* tweaks the trainable parameters and optionally hyperparameters such as the *learning rate* in the multidimensional optimization space to reduce the loss. The learning rate is a parameter that specifies to which extent the parameters in the network are adjusted in each iteration. In practice, the gradient descent and more elaborate approaches such as the Adam algorithm [201] are used as optimizers. The Adam optimizer is a stochastic optimizer, which relies on first-order derivatives. During learning, learning rates for specific parameters are adapted by consulting the first moments (mean) and second moments (uncentered variance) of their corresponding derivatives. It was devised to be applicable to sparse gradients, to have update rates that are invariant to gradient rescaling, and to perform step size annealing intrinsically [201]. In different empirical studies, its effectiveness as compared to other optimizers was demonstrated [202].

2.3.4.2 Network architectures

From this building blocks, many sophisticated architectures were devised in literature, including U-Net [197], 3D U-Net [203], DeepLab [204], or fully-convolutional DenseNets [205], to name a few. The architectural decision depends on the complexity of the task, computational capabilities, quantity of data, its annotation type, and data characteristics. An example of the latter is the distribution of features within the image, which affects the decision of whether the architecture requires a larger receptive field. In order to address this, Yu and Koltun developed a backbone with dilated convolutions [206]. However, the concept of fully-convolutional networks and the convolution operation specifically involves an inductive bias and hence constraints the achievable performance. Namely, the assumption that only the vicinity of each pixel contributes to the decision of assigning a specific class to a pixel might not be justified for every dense task. Recently, architectures that reduce the

inductive bias have shifted into the focus of research. These include transformer-based models which rely on so-called *attention heads* and were introduced in natural language processing initially. Such networks can potentially consider wide-range features, while also applying its weighting dynamically rather than the static weights in convolution kernels [207, 208].

2.3.5 Performance metrics

The selection of appropriate performance metrics is central for evaluating models and prevents drawing incorrect conclusions. Depending on the task and data properties, different metrics are advised. While dense prediction tasks for images have dedicated metrics such as intersection over union or average precision for semantic and instance segmentation, respectively, classification models often are assessed using accuracy, F_1 score, or the area under the receiver operator characteristic curve (ROC-AUC). Most classification metrics are based on the so-called confusion matrix. The confusion matrix relates the classifier predictions (columns) with the actual labels (rows) for all instances, see table 2.1.

Table 2.1: A schematic structure of a confusion matrix for binary classification.

		Prediction	
		Positive	Negative
Actual	Positive	TP	FP
	Negative	FN	TN

Here, TP, TN, FP, and FN are the amount of true positive, true negative, false positive, and false negative classified instances, respectively. In multi-class scenarios, the confusion matrix expands according to the number of classes present. There is a multitude of derived metrics that attempt to measure specific properties of the classifier or its performance as a whole. For instance, the *sensitivity* or true positive rate (TPR) expresses the proportion of correctly predicted positives among all actual positives. Similarly, the false positive rate (FPR) is defined as the ratio of the falsely positive predicted samples to all actual negative samples. Their mathematical definitions are provided along with those of a few more relevant metrics in Table 2.2.

Table 2.2: Different relevant evaluation metrics based on the confusion matrix.

True positive rate	**False positive rate**	**Precision**
$TPR = \frac{TP}{TP+FN}$ (2.19)	$FPR = \frac{FP}{FP+TN}$ (2.20)	$PPV = \frac{TP}{TP+FP}$ (2.21)
F1 score	**Accuracy**	**Intersection over union (IoU)**
$F1 = 2 \cdot \frac{TPR \cdot PPV}{TPR+PPV}$ (2.22)	$Acc = \frac{TP+TN}{TP+TN+FP+FN}$ (2.23)	$IoU = \frac{TP}{TP+FP+FN}$ (2.24)

All metrics are defined in the range from zero to unity (or 0–100%), where the upper bound corresponds to an ideal model prediction. The accuracy and IoU metrics are used in this work to evaluate image segmentation models. While the accuracy metric measures the correctly predicted pixel percentage, the IoU measures the ratio between the intersection and union of predicted and labeled pixel areas. The accuracy metric suffers from limited sensitivity in case of notable class imbalance when the negative class is stronger represented. Note that TN occurs in additive terms in the nominator and denominator of Equation 2.23. Nevertheless, due to its intuitiveness, it is provided in this work whenever appropriate. In contrast, the IoU captures the model differences more adequately for data sets skewed towards the negative class, which is why we focus on it for the comparison between the individual models. In the case of semantic segmentation where multiple foreground classes

are supposed to be distinguished from the background and each other, the IoU is evaluated for each foreground class and background distinctly and subsequently averaged. For non-dense classification tasks, and in the presence of data imbalance, G-mean, weighted accuracy, and the F1 score are often employed [209].

Many classifier models provide probabilities or their unnormalized variants, i.e., logits as their output. The decision threshold in such cases needs to be selected optimally and affects the confusion matrix. One way of selecting a favorable decision threshold is to evaluate the so-called receiver operator characteristic (ROC). It represents the true positive rate (TPR) plotted against the false positive rate (FPR) for varying decision thresholds. An exemplary ROC curve is illustrated in figure 2.10.

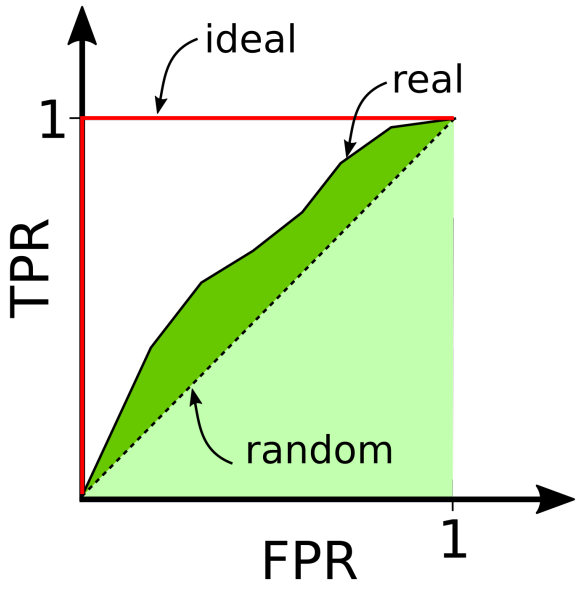

Figure 2.10: A schematic receiver operator characteristic plot with different depicted scenarios.

The angle bisector of the ROC plot, represents no expressivity of the model with respect to the target variable, i.e., an entirely random model (dashed line in figure 2.10). An ideal ROC curve would be a step function. Typically, real models exhibit an intermediate ROC curve progression. An ideal decision threshold would then coincide with the point where the distance to the random model is maximal. The area under the curve (ROC-AUC) is a measure of model performance as it describes separatability between two or multiple classes. As the name implies it computes the integral under the ROC curve. It is an aggregate metric that takes all decision thresholds into account and indicates the degree of overlap in the class probability distributions. The ROC-AUC ranges from zero to unity, where unity represents ideal separability, 0.5 no separability, and zero inverted prediction. The ROC-AUC metric has the characteristics of scale invariance and decision threshold invariance. In cases where either false negatives or false positives are substantially more relevant than the other, additional metrics should be involved since the ROC-AUC, due to its integral character, does not provide the means for such an optimization. In general, it is good practice to consider multiple of the aforementioned metrics. Especially, when imbalanced data is concerned and different models are compared, some of the metrics such as ROC-AUC can produce a deceptive ranking of the models [210].

3 Method development

3.1 Fatigue setup development

The resonant fatigue setup presented here expands upon [211] and [35]. It achieves a sensitivity that enables the detection of early damage states starting with the formation of persistent slip bands (PSBs). The necessary signal-to-noise ratio is achieved through miniaturization, specific loading conditions, and the sensitive control mechanism.

As depicted in figure 3.1a, the mesoscale planar hourglass-shaped specimen (blue) is mounted in a specimen holder. Two piezoelectric actuators induce a deflection on the specimen holder, which excites a resonant state in the specimen.

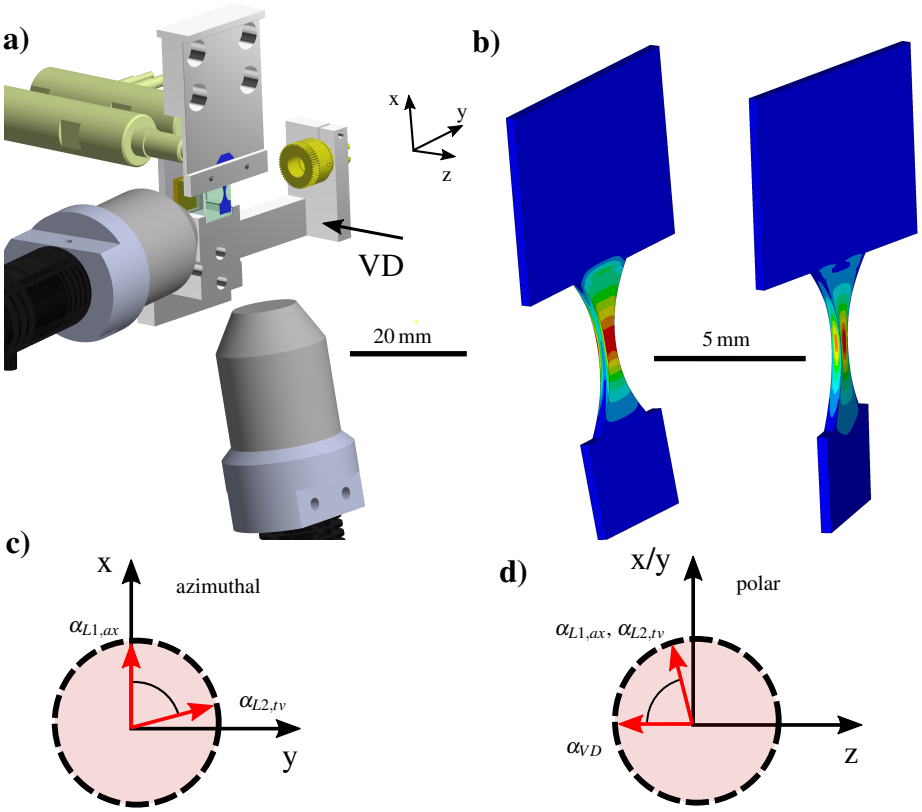

Figure 3.1: a) Micro fatigue setup for sensitive detection of early fatigue states composed of two piezoelectric actuators (khaki), a laser (yellow), a beam splitter (transparent blue), position-sensitive device (olive) LED spot lights (black), and a specimen (blue). The viewing direction (VD) describes the orientation of the imaging optics. **b)** The von Mises stress distribution induced in mesoscale specimens for bending and torsion scenarios. **c–d)** Angular arrangement, i.e., azimuthal and polar orientation, of the LED spot lights for stroboscope illumination. Adapted from [36].

Depending on the phase between the two piezo actuators ϕ_p either a bending, torsional or multi-axial resonance can be induced. The bending and torsion loading states are depicted in figure 3.1b. In the following, the bending case is addressed mainly, since it represents the central loading scenario in this work. The out-of-plane bending resonance mode is induced by applying a cyclic bending deflection of the corresponding frequency at the specimen holder. Since the specimens are designed such that their relevant resonance frequencies and higher-order harmonics are far off those of the specimen holder, the deflection of the latter is comparatively negligible. By the virtue of this configuration, only fully reversed loading ($R = -1$)

is attainable. Aside from specimen miniaturization, the first-order bending loading further confines the highly loaded volume to the specimen surface. This fosters the high sensitivity of the setup since changes in this small volume define the specimen response.

The bending slope is acquired by illuminating the unconstrained end of the gauge section with a laser and measuring the displacement of the reflected laser spot on a position-sensitive detector (PSD). If the specimen is bent, the laser spot reflected from the specimen is deflected up or downwards. Underlying geometrical considerations tying the specimen deflection with the laser shift on the PSD, are explained in detail in [211]. Since the displacement detected on the PSD is used as a control parameter for adjustments in the piezo displacement and frequency, the setup can be considered bending slope-controlled. In order to estimate the stress state for a given PSD amplitude, continuum simulations are performed in advance of the experiment.

The sinusoidal piezo actuator signal and the sinusoidal PSD signal represent an input and output signal, see figure 3.2. In the resonant state, the relative phase shift between these signals corresponds to $\phi = 90\,°$. In the course of cyclic loading, resonant frequency changes occur due to softening/hardening, damage evolution, or oxidation in the whole loaded region (integral measurement). The frequency of the actuation signal is being controlled to maintain the 90 ° phase shift. Hence, the resonant frequency changes are traced by continuously altering the actuation frequency to match the change in specimen resonant frequency. This contributes to the pronounced sensitivity since the phase response exhibits a steep slope in the vicinity of the resonant frequency. Additionally, the piezo input amplitude is controlled using a proportional-integral-derivative (PID) control to preserve a constant specimen bending slope during the test. Since amplitude and frequency are correlated through dampening phenomena, both, the frequency and amplitude control are acting simultaneously. This control mechanism, and specifically the control parameter being measured at the end of the gauge section come with a few implications. When a crack exceeding a critical length is present, the bending and stress state deviates from the simulation. This applies especially to the section of the beam between the crack and the gauge-mass transition (where the laser measures the bending slope) in which the load diminishes. Therefore, this setup is typically applied to characterize fatigue states ranging from PSB development up to physically short-crack growth.

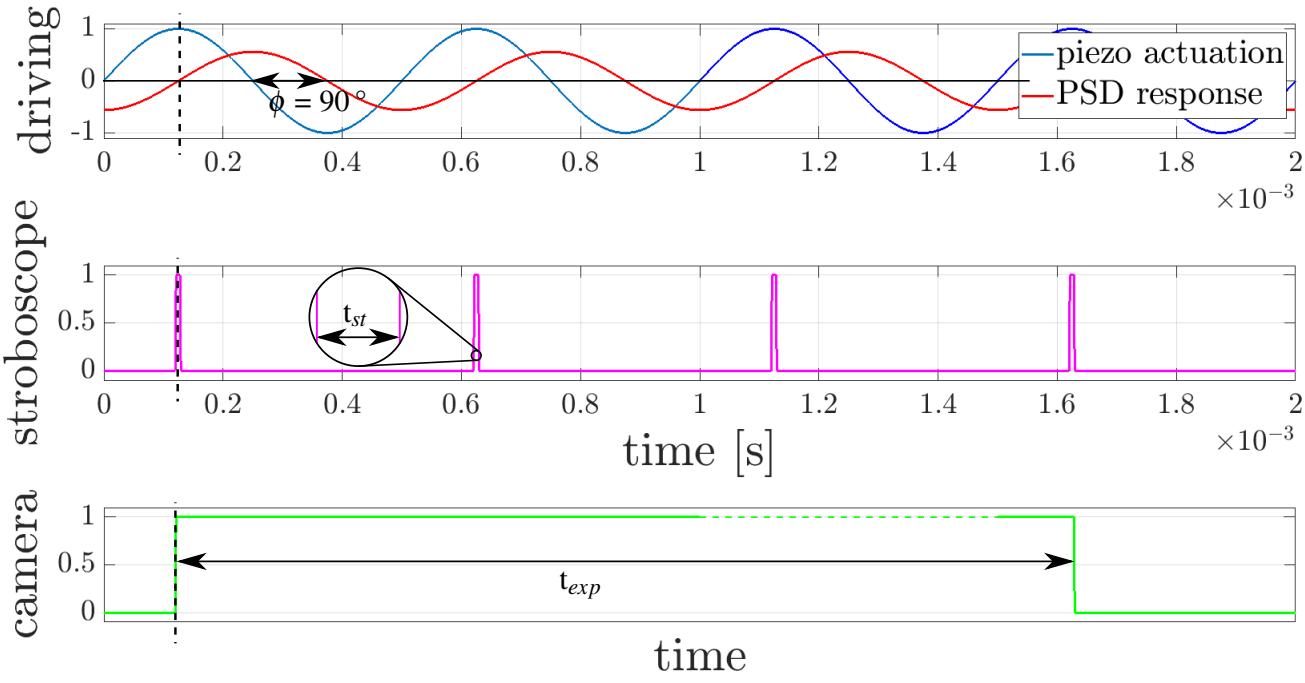

Figure 3.2: The top diagram illustrates the cyclic piezo actuation and the response on the position-sensitive detector schematically. In the center, the timing of the LED stroboscope illumination is displayed. The camera remains open to accumulate the light retrieved from the specimen surface over several fatigue cycles.

While the actuation frequency signal gives integral information on the specimen damage state, an additional methodology to capture local changes based on [35] is introduced in the following. Retrieval of spatial information was achieved by acquiring images of the specimen surface. This has the objective of observing the cyclic evolution of local damage instances and correlating it with the underlying microstructure. Therefore, the setup was extended with a stroboscope illumination and camera system (Basler® beat beA4000-62km monochrome camera and Canon® MP-E 65mm f/2.8 1–5× macro objective) capturing a series of in-situ light optical images. The monochrome camera variant was selected due to its superior quantum efficiency, i.e., more efficient conversion of photons into electrons which facilitates shorter exposure times. In terms of illumination, a bidirectional oblique illumination by two light emitting diode (LED) spot lights (2× CCS® HLV2-22BL-1220_3W LED spot light, see figure 3.1a) in the blue wavelength range with its peak at 465 nm was utilized. Previously, superior contrast and resolution obtained by oblique illumination techniques was reported [212]. The image formation is based on light scattered at damage emerging from the specimen surface facing the camera. Therefore, surface defects such as protrusions and cracks appear as regions of high intensity.

The orientation of both LED spots lights in the azimuthal and polar plane of the spherical coordinate system, see figure 3.1c and 3.1d, was chosen such, that the azimuthal angles differed by approximately 90 ° to capture surface defects independent of their surface topography. Moreover, it was empirically ensured that the optical yield on the camera is maximized and that the interference on the position-sensitive device is negligible. An influence of the illumination orientation on the image formation is shown in [213]. Due to the fairly flat light incidence angle and the camera viewing direction, it is assumed that both, specular reflection and predominantly diffusive reflection (scattering) contribute to the image formation. The scattering portion strongly depends on the ratio between surface roughness (or scale of surface features) and the wavelength [214].

Both distinctly oriented LED spot lights illuminate the specimen at the inflection point of the fully reversed bending motion (i.e., the non-defected state) as illustrated in figure 3.2. Note that the fast motion of the specimen at typically 1–2 kHz in bending resonant mode raises the demand for accurate timing and short duration of the stroboscope light pulses. Hence, short pulses with t_{st} in the range of 5–10 µs are utilized ensuring imaging with minimized specimen motion blurring. The camera integrates for an exposure time t_{exp} over multiple cycles and multiple light pulses to retrieve a detectable scatter signal from the specimen surface. Typically an exposure time in the order of t_{exp}=100 ms is employed. Depending on the specimen surface quality it is slightly varied to avoid under- or overexposure.

An image series acquired in the course of a fatigue test typically comprises several thousand images, and a subset of it is showcased in figure 3.3. The figure shows the highly loaded region of the specimen (red region in figure 3.1b). In these intensity-inverted images, the dark spots present at the beginning of an experiment originate from light diffraction at particles, pores, or residuals on the specimen surface. As cyclic loading is applied, see figure 3.3b–i, multiple protrusions, and a few cracks initiate (black arrows) across the specimen. While some of the protrusions and cracks stagnate, others manage to grow and propagate (red arrows). The underlying hypothesis is that as protrusions grow, more light is scattered at these surface defects, and the local intensity associated with damage instances increases. This would enable tracking the local evolution of cracks and protrusions. The validity of this hypothesis is explored in the results section by employing correlative surface topographies obtained by atomic force microscopy.

Despite the high mechanical frequencies, it is anticipated that the mesoscale specimens' pronounced surface-to-volume ratio, the low plastic strain, and forced convection through specimen motion permit cycling at faster rates without notable heating on the global specimen scale. Moreover, the influence of the LEDs in terms of heating is assumed to be negligible. This can be ascribed to the high reflectance due to the polished specimen surface condition and the flat light incidence angle.

3.2 Automation of scanning electron microscope

Another fundamental requirement for the creation of statistically representative multimodal damage data sets is the automation of analytical techniques. Due to the nature of HCF and VHCF only a few damage locations emerge in distinct locations

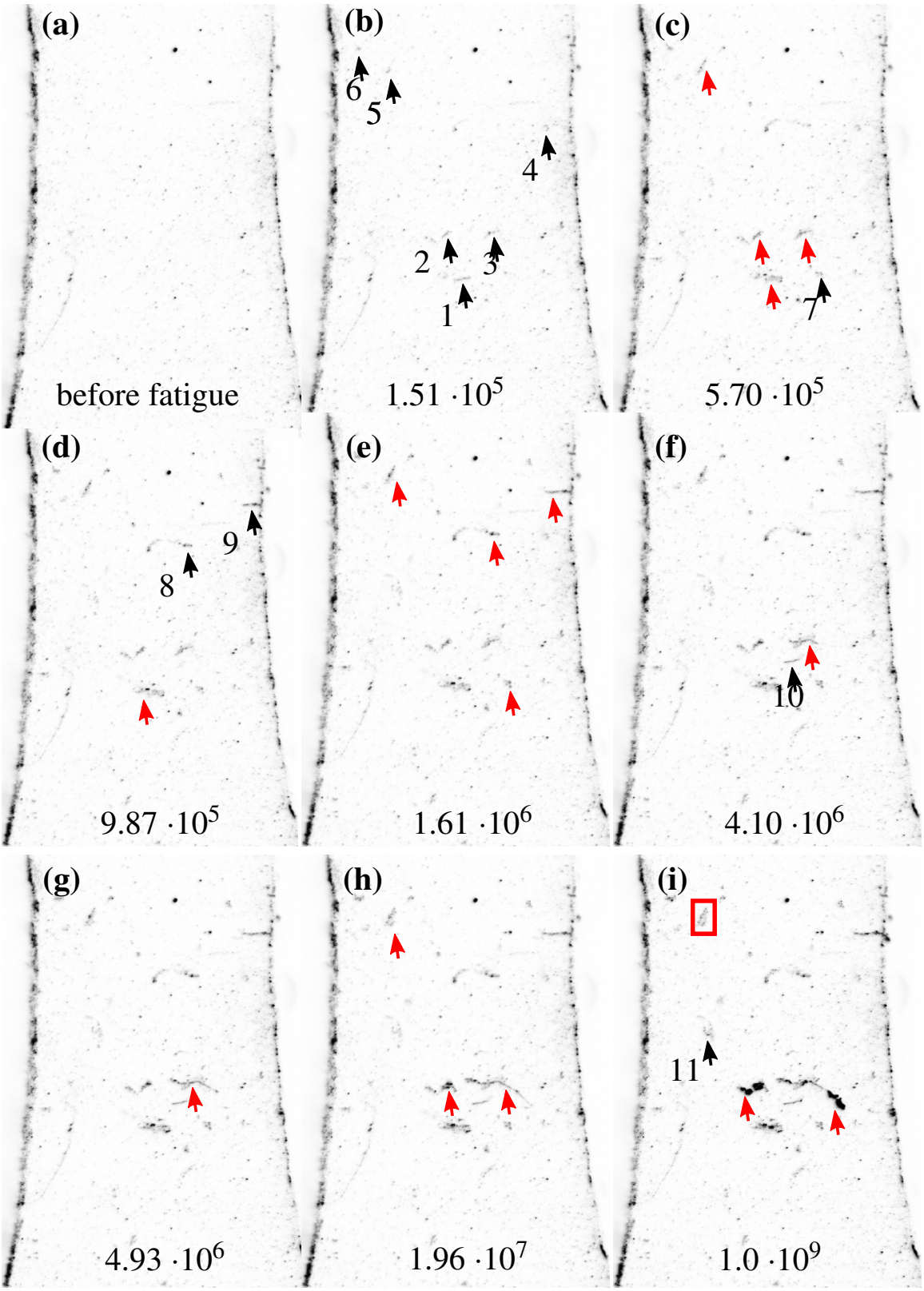

Figure 3.3: Inverted oblique illumination image series with annotations. Black and red arrows indicate the emergence and growth of a damage instance, respectively. The red box marks a protrusion that is treated in figure 3.10. Numbers at the bottom of individual images represent the approximate cycle numbers at which they were acquired. Adapted from [36].

where critical microstructural scenarios occur. Aside from being sparsely distributed across the specimen, the damage instances exhibit very fine features. Therefore, when imaging damaged specimens, typically large areas need to be scanned with high magnification to capture relevant damage features in the image texture for a multitude of damage instances. The scanning electron microscope was chosen as an automation target as it allows for resolving relevant damage features and exhibits a pronounced depth of focus. In order to address the problem of limited field of view at high optical magnifications, alternate methods other than standard single frame capture were required.

Therefore, in a collaboration with a co-worker, Mathis Bellmer, multiple C# control programs for the Zeiss Supra 40VP were written using the Zeiss API. While the SEM provides different detectors that collect various kinds of specimen information, for the characterization of damage morphology the surface topography-sensitive Everhart-Thornley secondary electron detector was utilized. Subsequently, these programs were extended to operate also under the specific EBSD configuration.

- **Tile image acquisition for stitching:** Subsequently to optimizing the acquisition parameters, this control program acquires a set of images (tiles) given a rectangular region of interest by corner points. Therefore, a tile overlap, resolution, and dwell time are prescribed. In the following, these tiles are combined into a single image (*stitching*).
- **Targeted coordinate image acquisition:** This approach rigidly transforms an image pixel coordinate list into SEM stage coordinates in order to capture high-resolution images at target locations. For the coordinate transformation, reference markers or invariant characteristic features in both imaging domains can be employed. The objective behind this approach was to derive an image pixel coordinate list corresponding to damage locations directly from frame differencing and morphological operations on the images in figure 3.3.

Since at a later stage, the objective is to register (spatially align) SEM data with data from other imaging modalities, the former imaging methodology of tile image acquisition for stitching was followed predominantly. This subsequent registration requires point correspondences of features which ideally should be taken from regions where substantial mechanical deformation and image distortions are absent. As opposed to the targeted coordinate image acquisition at damaged locations, this imaging methodology captures more undeformed regions which facilitates the straightforward assignment of point correspondences. As indicated in the introduction, this stitching image acquisition was performed twice, before and after fatigue testing with slightly different settings.

The acquisition parameters used to capture the tile images are described in section 4.3.1. Image resolutions of 2048×1536 to avoid information loss and typically 15×13 tiles with a tile overlap of 3–5% were selected. Consequently, the resulting stitched images had a resolution in the order of $30{,}000 \times 20{,}000$. This was required to differentiate between protrusions and cracks by resolving the distinct features. A frequent issue that occurs when acquiring a set of images automatically includes actuator backlash and slip at the motorized translation stages which results in a shift between the adjacent images [215]. In literature, Chen et al. [20] applied normalized cross-correlation to correct such shifts for stitching SEM images of magnesium surfaces decorated with gold nanoparticles. To identify the appropriate peak in the correlation map, background estimation (using a Savitzky-Golay smoothing filter) and background subtraction was performed in their work. In our case, for stitching, the implementation in [216] was utilized. The algorithm uses the inverse Fourier transform of the phase correlation in order to determine the translational shift between the neighboring tiles. This assumes that the tile overlap is chosen sufficiently high to compute the phase correlation even in presence of tile shifts. A stitched image is depicted in figure 3.4.

Furthermore, the brightness between tiles can differ substantially, which is referred to as *shading*. These differences arise due to the imaging configuration in the SEM. Amongst others, shading can be ascribed to a change in specimen surface orientation with respect to the detector and to a change in electron source luminescence between tiles. Other potential sources of brightness differences are electron beam-induced oxide deposition and other sources of charging. Shading can be to some extent accounted for by background subtraction or different image blending strategies in the tile overlap regions [216]. In our case, two different image types were considered for stitching – the as-acquired input images as well as segmentation masks derived from the former by manual annotation. The same image positions, i.e., translational shifts, as computed based on the raw input images, were also applied to the mask images. Linear blending was applied at the overlapping regions for the stitching of the input images. However, for the mask tiles, this was not appropriate since interpolation would lead to

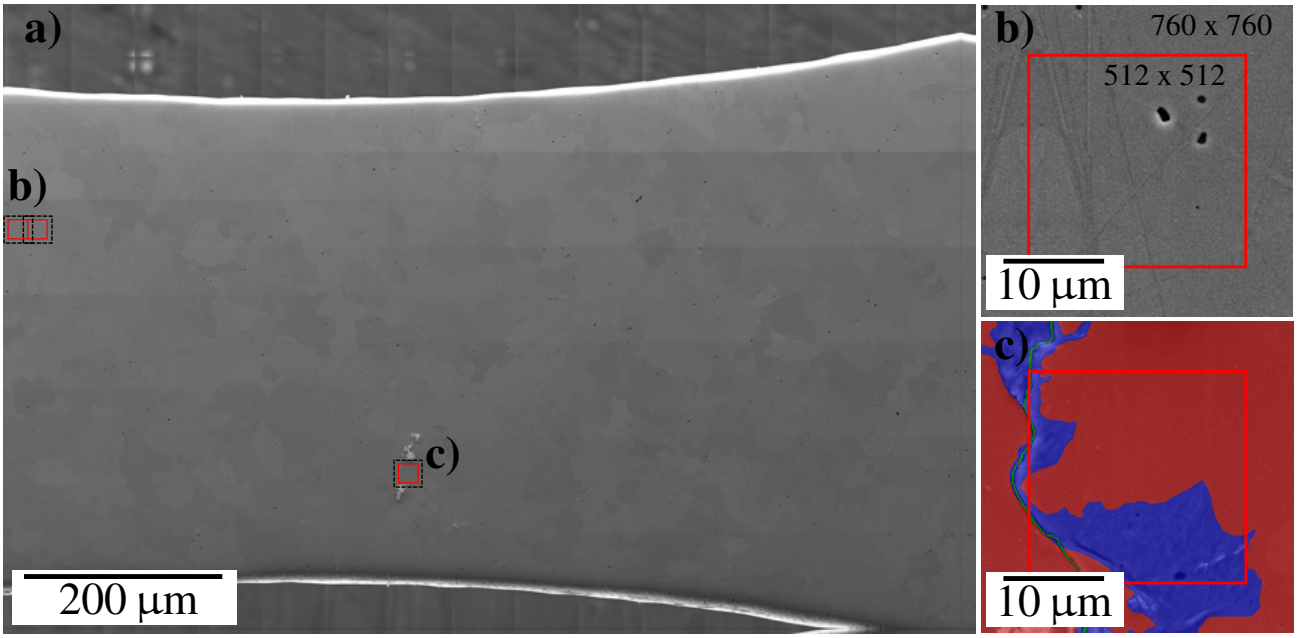

Figure 3.4: a) A stitched secondary electron scanning electron microscopy image of a mesoscale fatigue specimen and some detail views in **b)** and **c)**. **b)** A miscellaneous area on the specimen to indicate the tiling approach used for deep learning. **c)** A region containing a microstructurally short crack and slip markings with the manually labeled damage mask superimposed. The colors red, blue, and green represent the background, extrusion, and crack classes, respectively. Adopted from [37].

invalid class labels in regions of conflicting class labels. In order to resolve the latter, the maximum class label value was chosen in case of conflicts, while the class labels were chosen such that they correlate with the severity of the damage type (cracks >extrusion >background). These pixel class label conflicts can either occur due to (inadequately corrected) tile shifts or distortions present in images.

Depending on the magnification, different types of physical mechanisms determine the contained image distortions. For instance, at higher magnifications, as the electron beam is confined to the center region of the electron optics, the time-independent spatial distortions [217] are reduced. This can potentially allow for better relative tile alignments. Charging, temperature drift or stage drift can induce time-dependent and non-uniform drift distortions in the SEM images, which are particularly relevant at high magnifications [217]. However, appropriate grounding and short dwell times can provide a remedy. During image acquisition, SEM operators are responsible for setting imaging parameters such as brightness and contrast, for ensuring electrical conductivity, and for avoiding defocus, and astigmatism. These parameters collectively influence the signal-to-noise ratio. Amongst others, robust automatic astigmatism and defocus correction being the state of research and therefore unavailable in standard SEM devices [218], results in images containing notable variance. Moreover, the applicability of these correction approaches is often limited to specific surface morphologies or specimen geometries.

These distortions do not only affect the stitching and registration procedure. The local image texture can be influenced by distortions as well. Depending on the specimen geometry, such distortions can vary from tile to tile. Aside from SEM and operator subjectivity-induced variances, the appearance of slip traces and cracks in surface-sensitive SE2 SEM images can differ due to several inherent factors including the material, specimen fabrication, and loading condition during fatigue. For instance, the microstructure and crystal structure determine how materials accommodate applied loads through the dislocation slip processes and therefore affect the shape and size distributions of slip traces and cracks. The large variety of materials employed in engineering leads to a substantial versatility of damage-related image textures. In Nickel specimens, several parallel but distinct and spaced extrusions delimited by grain boundaries were observed [219]. The specimen fabrication, such as metallographic polishing steps, can alter the surface roughness and therefore the image background texture as well as the formation mechanisms of fatigue damage. A common issue is that specimen edges are often poorly preserved during polishing steps which impacts local image textures. Loading conditions, e.g. axiality, determine the slip characteristics [220] and thus the shape of the slip traces to a large extent. Furthermore, slip markings have shown to form a cyclic load-induced

oxide layer at their surface, which significantly exceeds native oxides in thickness [91]. This can influence the representation of slip markings in SE2 images. Nonetheless, comparatively stable imaging conditions in terms of a consistent viewpoint, minimal occlusion, and irrelevance of environmental influence factors are obtained for SEM imaging with respect to the vast majority of natural image data sets.

Since achieving a robust and automated workflow for damage data set generation irrespective of the material is one objective, the aforementioned stitching image acquisition routine was applied to create data sets for three materials. These materials and corresponding data sets are described in the subsequent section and were utilized to train data-driven models and characterize their damage detection performance. The selection of materials was based on obtaining a range of diverse microstructures and fatigue mechanisms. Despite the different globular and hierarchical microstructures as well as face- and body-centered cubic elementary cells prevalent for the chosen materials, they represent only a subset of the microstructural diversity found in engineering alloys. Nevertheless, it enables an investigation with respect to the generalization of damage detection models for distinct engineering materials.

3.3 Establishment of surface fatigue data sets

This section covers the pixel-wise manual annotation of the acquired SE2 images and the resulting data set. A systematic reference data set was established as a fundamental requirement for data-driven techniques. Due to the demand of accurate damage localization with respect to grain boundaries and other defects, segmentation (i.e. assigning a class label to each pixel in the input image) was selected as the task of choice. The decision to perform a pixel-wise manual annotation is justified by the superior performance of supervised learning segmentation models which require many input and segmentation mask image correspondences during the training stage. Since the objective is to distinguish between crack, extrusion, and background regions, the classes were chosen accordingly. In this section, unless explicitly specified otherwise, the concept of extrusions is used synonymously with arbitrary slip marking types.

The annotation process involved drawing the perceived border of damage locations manually, fine-tuning the regions, and assigning a semantic class label. However, the signal-to-noise ratio in the images impeded the accurate identification of the damage borders for some materials. Labeling was performed on the raw input images rather than the stitched image in order to allow for subsequent adjustments in the stitching procedure. Moreover, prospectively having manual annotations before stitching can assist in the stitching process if the overlap region between adjacent raw images contains damage instances. After the annotation and stitching process, the data was split into multiple tiles of 760×760 pixels as shown in figure 3.4. This was done in order to avoid GPU memory limitations during training, as proposed in [197]. Since image border padding was not performed in the encoder convolution layers of the applied U-Net architecture (see Section 4.5.2), the predicted mask tile size is reduced in comparison with the input tile. Therefore, prior to extracting the tiles, mirror padding was applied at the stitched image boundary. Furthermore, the tiles were extracted with an appropriate overlap to account for the reduced mask tile size and to take the entirety of the data into consideration for training and testing. This procedure was followed for each material. The set of tiles obtained was split in a training ($\approx 80\%$) and testing data set ($\approx 20\%$) with the exception of copper which owing to limited data availability was solely used for testing purposes. For domain generalization and adaptation studies, where the transferability of a model to another domain is investigated, it is a common practice to distinguish between so-called *source* and *target* domains. In this case, the ferritic steel material, posing the main research objective, is considered the source domain and the transfer to alternate target material domains is examined.

3.3.1 Source data set: Slip markings and cracks in ferritic steel

Initially, the damage observed in a ferritic steel EN 1.4003 with a body-centered cubic crystal structure was investigated. Training tiles originated from multiple specimens, which were fatigued with differing strain amplitudes. Cracks ranging from microstructurally short cracks to long ones were observed in the material.

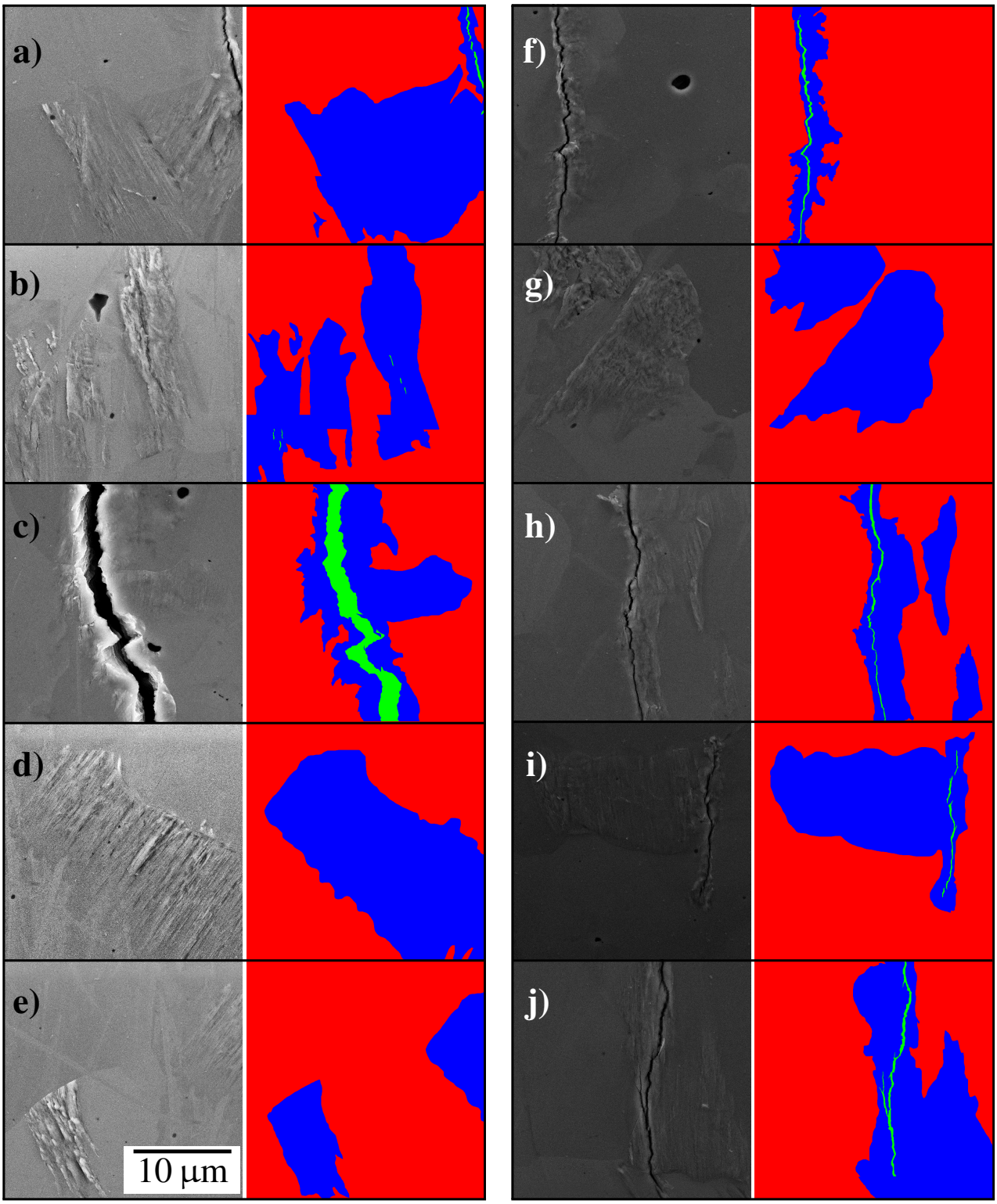

Figure 3.5: Pairs of SE2 input tiles and their manually labeled mask correspondences showing the diversity of the input data set. The figures show a subset of the ferritic steel data set (source domain). The scale bar applies to all subfigures. Adapted from [37].

As apparent in figure 3.5 typical cracks (green) present in the data set show distinct contrast compared to the background and frequently are encompassed by extruded volume (blue). This extruded volume is either present due to the accumulation of plasticity prior to crack initiation (figure 3.5h) or resulting from stress concentration in the plastic zone of the crack tip (e.g. figure 3.5f). While in micromechanics the term extrusion refers to the formation of a unique surface topography through the accumulation of dislocation steps and vacancies prior to crack nucleation, here both cases are merged into the extrusion class. Even though slip markings, according to [221, 222], can be classified based on their shape in ribbon-like, tongue-like, chord-like, and band-like extrusions, as well as macroscopic protrusions superimposed with the former types, the ferritic steel data set contains almost exclusively protrusions. This can be ascribed to the low defect density in the ferritic steel and the corresponding unimpeded movement of dislocations. The data was not annotated to distinguish between these extrusion sub-classes. Some protrusions are fissured and few even contain a multitude of tiny cracks. It is noteworthy that the labels do not exhibit pixel-level accuracy since the boundaries of the extruded areas in the ferritic steel are faded and thus not always evident.

In the appendix section A.1, sources of variance in all three damage data sets are assessed.

3.3.2 Target data set i: Slip markings in copper

Additionally, a fatigue damage data set of polycrystalline oxygen-free high conductivity (OHFC) copper exhibiting a face-centered cubic crystal structure was created. The specimen was fabricated and fatigued as outlined in a preceding work [211] and generously provided for further analysis. Due to its low yield strength, scarcity of defects, and the applied loading conditions, slip markings observed in this data set are extensive and spread across grain boundaries (not depicted). The extrusion class entails instances where plastic deformation bands extend nearly across the whole specimen width (for a tile of which, see figure 3.6a–c as well as moderately localized protrusions comparable in scale to the ones observed in ferritic steel (figure 3.6d). Moreover, multiple fine tongue-like extrusions can be observed, see arrow annotation at the right border of figure 3.6c. Former larger slip bands observed in copper do not show unique well-defined slip trace orientations but rather indications of wavy slip and multi-slip. The copper data set in the context of this work was utilized solely for testing purposes.

3.3.3 Target data set ii: Extrusions in martensitic steel

Concluding the damage data sets, a complementary slip marking data set comprising extrusions in a steel EN 1.7228 with a martensitic microstructure was prepared. The lattice distortions and the hierarchical microstructure of the material result in confined extrusions, as apparent in figure 3.6e–h. In contrast to extrusions observed in ferritic steel, these extrusions can be classified as tongue-like and are one order of magnitude smaller. These confined tongue-like extrusions are at a similar size-scale as present OP-S particle agglomerates, see arrow annotation in 3.6h. Another image feature owed to the hierarchical microstructure is the distinct background texture. It is comparatively spotted and non-uniform, see figure 3.6h, and shows sub-structures comparable in size to the emerging damage instances.

Since the martensitic steel microstructure exhibits much smaller effective grain sizes than ferritic steel, the microstructure variance in a given area is comparatively more pronounced. This leads to the fact that more martensite images contain extrusions despite the lower number of total extrusion pixels, see table 3.1.

Each of these material domains contains distinct slip markings and distinct processing-induced variance which results in unique image textures. As a result, the performance of a trained network to segment damage in an alternate unseen material (domain generalization) can be evaluated and conjunct training with diverse materials data can be tested.

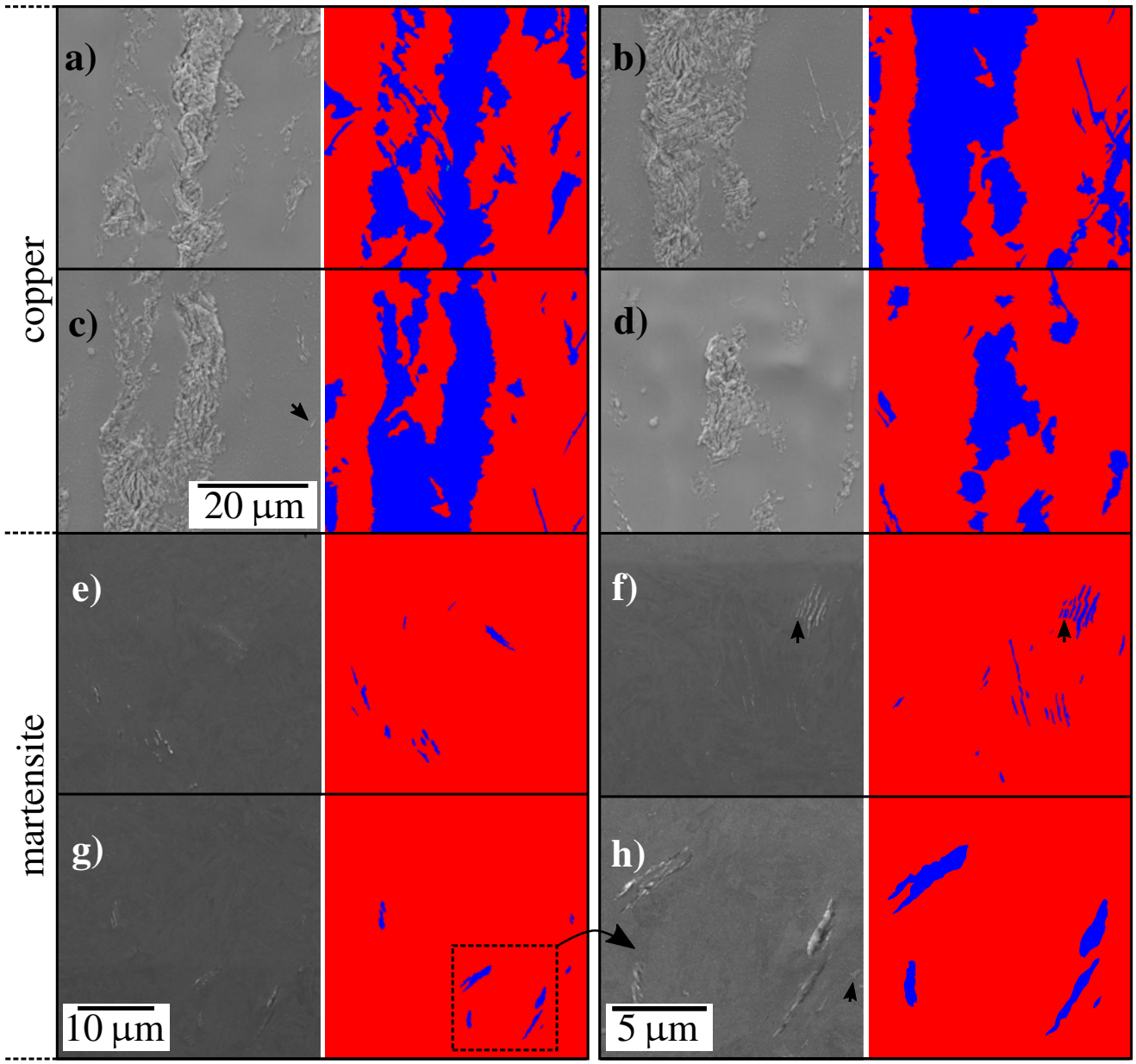

Figure 3.6: Pairs of SE2 input images and their manually labeled mask correspondences showing the diversity of the input data set. The figures show a subset of the copper (**a–d**) and martensitic steel (**e–h**) data sets representing target domains. The scale bar in subfigure **g)** applies to all subfigures except the detail view in subfigure **h)**. Adapted from [37].

3.3.4 Quantitative data set comparison

In the following, a quantitative comparison of the data sets is performed using a few characteristic data set metrics that describe the distribution of the background, extrusion, and crack class in each material data set. A general data set overview is provided in terms of the number of tile images N_t used for training and testing as well as the number of original stitched images N they originate from. Furthermore, image/pixel percentage for both damage classes p_x^y, and 80% percentiles of the pixel area ratio for both damage classes P_x^{p80} are described. Subscripts x denote the damage classes crack c and extrusion e, while the superscripts y denote whether it is an image i or pixel p percentage value. For instance, the value $P_e^{p80} = 10\%$ for the ferrite testing data indicates that 80% of the images have an extrusion area percentage of 10% or less. Whenever possible, the metrics are provided for the training and test set, denoted by subscripts *tr* and *te*, respectively.

Table 3.1: Overview of data sets. The annotations (s) and (t) indicate whether the data set is a source domain or a target domain. As an example to explain the notation of the data subsets, the f_{tr} refers to the **f**erritic steel **tr**aining data set. Adopted from [37].

	ferritic steel (s)		copper (t)	martensitic steel (t)	
metric	training (f_{tr})	testing (f_{te})	testing (c_{te})	training (m_{tr})	testing (m_{te})
N []	12		1	3	
N_t []	3940	860	357	763	168
p_c^i [%]	5.35	6.97	-	-	-
p_e^i [%]	33.70	34.30	89.92	36.56	33.92
p_c^p [%]	0.06	0.09	-	-	-
p_e^p [%]	2.34	2.21	12.13	0.18	0.17
P_c^{p80} [%]	1.15	2.00	-	-	-
P_e^{p80} [%]	10.75	10.00	20.45	0.62	0.63

From table 3.1 and the $p_{c/e}^p$ values, the substantial intrinsic class imbalance of the three data sets becomes obvious since extrusion and in particular crack pixels account for few percent, especially in the ferritic and martensitic steel data sets. Only a small portion of the pixels containing damage can be ascribed to the fatigue loading in the HCF/VHCF regime, which leads to the degradation of a few critical grain clusters. Moreover, a comparison between the different materials in table 3.1 indicates that there are more labeled ferritic steel instances which culminates in an additional imbalance between the data sets. When training on combined data sets is concerned, this needs to be accounted for.

3.4 Image series registration

Due to subtle specimen movement in the course of a fatigue experiment, an automated image registration routine was required to align individual images of the image series. Shifts amongst the in-situ acquired light optical images as illustrated in figure 3.7 occurred impeding damage tracking. The image features in magenta and green, representing the initial and last image, respectively, being shifted uniformly across the whole image indicates that the misalignment originates from a shift rather than a deformation. These shifts are typically in the order of 10 µm. Presumably, the slippage movement can be ascribed to the relatively large inertial forces acting at high frequent actuation, imperfect clamping, and to a lesser extent to gravitational forces. The exact specimen movement path is highly dependent on the actuation and clamping conditions.

Different feature and intensity-based registration algorithms were tested. Due to modest local fluctuations in image texture between subsequently acquired images resulting from stroboscope illumination timing and fatigue damage evolution, feature-based registration proved difficult for this task. The present noise levels caused by the stroboscope lighting and camera as well as damage-induced image changes complicate feature matching between subsequent images. An aggravating circumstance is the scarcity of surface features across the specimen due to the applied specimen surface finish (see inlay of figure 3.7). Particles and predominantly pores provided somewhat constant image features in these light optical image series. However, the

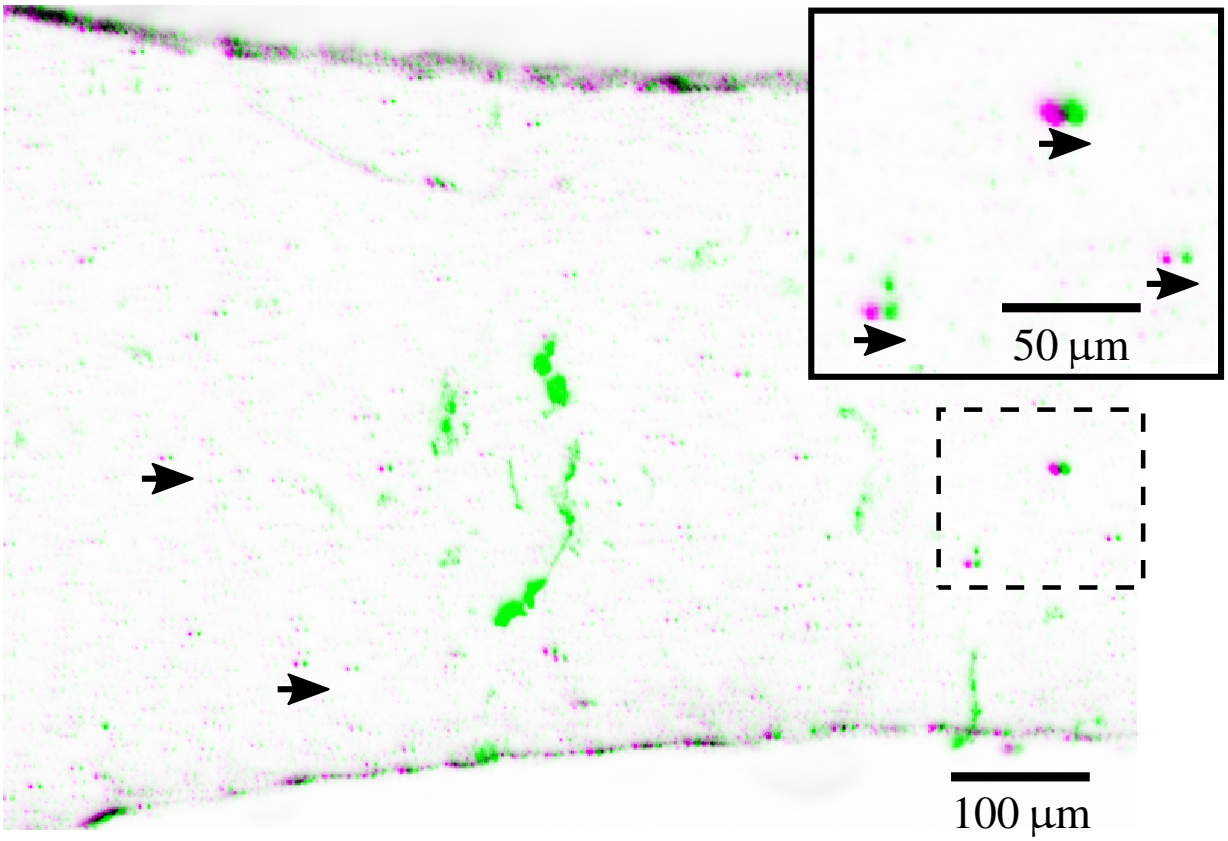

Figure 3.7: Pair of inverted oblique illumination images acquired in-situ at the beginning (magenta) and end (green) of a fatigue experiment in false color representation indicating specimen movement in the course of the experiment. Dark regions represent regions of overlap.

arising intensity distributions around different pores within one image are not distinctive (repetitive image texture). Computationally efficient utilization of unique long-range pore patterns would require image downscaling operations since pores are sparsely distributed across the image. However, downscaling does not preserve the small pore's features well. In conclusion, feature detection and feature matching algorithms proved unsuitable for the present surfaces. An alternative is altering the surface by applying dedicated speckle patterns typical in the field of digital image correlation (DIC). It was refrained from this to avoid affecting the very surface-sensitive fatigue test.

On the other hand, for the present task, an intensity-based method utilizing the so-called Mattes mutual information metric [223] proved appropriate yielding reproducible and stable registrations throughout the whole image series. Depending on the loading conditions and image acquisition rate, typical image numbers range from 100 to 12,000 images per fatigue experiment. The employed mutual information metric measures the similarity between pixel samples of two images in their intensity distribution and thus how well they can be mapped onto each other. In the course of the registration procedure of two images, the moving image is displaced such that the entropy between the two pixel distributions is minimized. The permitted degrees of freedom are displacements, translations, and rotations, i.e., rigid transformations. This euclidean distance preserving, linear transformation type is adequate if plasticity and crack growth do not cause major specimen deformation. Indeed, the macroscopic deformation is largely negligible due to the fully reversed bending load at very low stress amplitudes. The registration implementation employs pyramids, a concept used in image processing to capture features on multiple scales. In a nutshell, pyramids are generated by subsequent filter and downsampling operations [224] and increase the robustness of the algorithm.

On a higher level, the structure of the image series registration code is an iterative process where initially the first image $\mathbb{I}_0$ is used as a reference for registering subsequent ones $\mathbb{I}_k$ in ascending order until the last image $\mathbb{I}_K$. The computed transformation matrix from the prior image registration $\mathbb{M}_{k-1}$ is utilized as an initial guess for the consecutive one since the specimen shift seemed to be of continuous nature. Moreover, after K_p=10 registered images, the reference was replaced by the newly

registered image in order to avoid erroneous registration owing to large changes in damage structure. Finally, an image contrast estimate is computed by convolving the image with a Laplace kernel and averaging all pixel values. This metric allows for the removal of outliers affected by motion blurring from the further analysis. In the final stages of the fatigue test, such outliers can occur to due drastic changes in resonant frequency culminating in poorly timed stroboscope illumination.

Establishing an efficient routine to register the large image quantity alleviated the problem of image shifts and enabled the correct assignment of damage instances and therefore the assessment of damage kinetics.

3.5 Multimodal registration

Correlative microscopy holds the potential to give access to previously inaccessible information [225]. For instance, investigating the local interdependencies between localized cyclic damage evolution data and microstructure information requires spatial alignment of both. While many commercial correlative microscopy tools exist, they often rely on vendor-specific specimen holders and do not correct for non-linear distortions in the data sets [226]. In contrast, we applied image registrations to match the coordinate systems and correct distortions of the data sets. Registration methods can be subdivided into feature-based and intensity-based techniques. While automated feature-based techniques require methods of feature detection, extraction, matching, and outlier exclusion, intensity-based ones rely on the correlation of intensity textures in images or sub-images, see [227]. *Multimodal* data registration refers to the spatial alignment of the source to target data sets and is accompanied by challenges concerning data heterogeneity. This comprises aspects such as different imaging modalities, specimen states, physical pixel sizes, dimensionalities, viewpoints, and field of views [228].

With regard to this study, the image-based data sets depicted in figure 3.8 were registered, where (s) denotes source, (it) intermediate target, and (t) target data sets. The in-situ images were optimized to have minimal optical geometric distortions by considering solely the center of the frame where the highly loaded specimen region was positioned. Moreover, the relatively narrow spectrum of the LED spotlights circumvents chromatic aberrations. When it comes to SEM image data, it was claimed that distortion correction is required for accurate overlay [229]. However, in our case, appropriate imaging conditions for image tiles, their stitching (see section 3.2), and the usage of a custom specimen holder throughout the whole fatigue and analytical process chain minimized distortions in SEM images and specimen alignment-induced relative distortions between the data sets. Therefore, the EBSD data was the only data set for which distortion correction was necessary.

In this multi-stage registration procedure, the undeformed and comparatively undistorted stitched SE2 SEM image of the specimen surface before fatigue acts as an intermediate registration target. This intermediate state facilitates subsequent collective transformation of all (s) and (it) data sets to the target EBSD stage by the same transformation. Hence, it could be ensured that all individual fairly undistorted source data sets are aligned well amongst each other. Note that using direct transformations to the distortion containing EBSD map possibly results in an erroneous alignment amongst the transformed source images. In literature, automatic feature detection and matching for light optical and SEM data was reported [230] employing scale-invariant feature transform (SIFT) [231]. However, this proved inadequate for the data at hand since the data sets exhibit a wide range of physical pixel sizes and distinct grayscale textures. Therefore, a landmark approach using a selection of point correspondences is conducted to derive most affine transformations. Pores and static particles visible on the specimen surface in various modalities facilitate the user's selection and matching of these features. Initially, the stitched SE2 image after fatigue and the in-situ light optical image time series are affine transformed to match the SE2 image before fatigue, see arrow (1) and (2) in figure 3.8, respectively. Additionally, a deep learning model-inferred segmentation map indicating damage locations is derived from the SE2 image after fatigue testing (see section 3.3) and transformed accordingly. Thus, every experimental source data set is at the stage of the SE2 image before fatigue.

Carrying along the idealized geometry (without fabrication-induced variances except for actual specimen dimensions) and loading conditions, together with the embedded microstructure enables straightforward incorporation in micromechanical simulations. The absence of defect-related features in the geometry necessitates a different approach for its registration. An intensity-based algorithm relying on shape-based features is used to transform the idealized specimen geometry represented

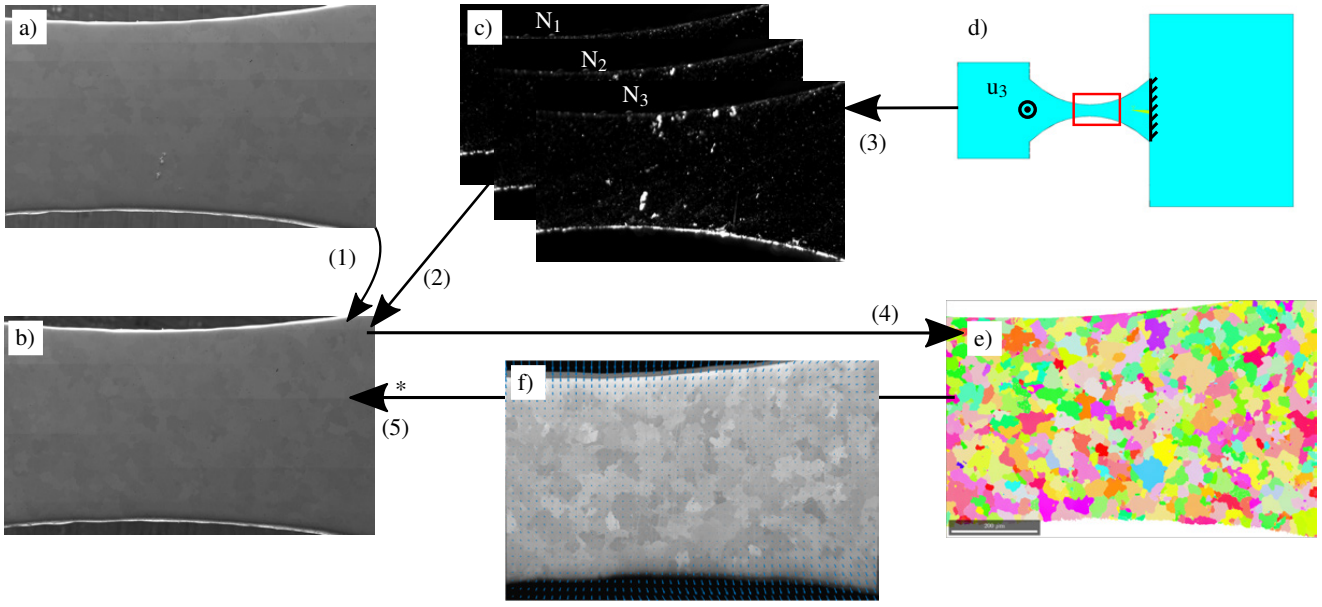

Figure 3.8: Overview of the multimodal registration process. The regular arrows and the one labeled with the asterisk describe affine transformations and distortion correction by an elastic transformation, respectively. Adopted from [36].

by a binary mask onto a binary mask derived from the light optical images as indicated by the arrow (3) in figure 3.8, followed by the transformation (2). Therefore, the Mattes mutual information metric [223], which measures the relation between two-pixel distributions, was employed. In the case of image registration, the measure contains information on how well pixel samples from both images are mapped onto each other. During registration, the alignment is optimized by reducing the entropy between the two-pixel distributions, see [232] for algorithmic details. In the employed Matlab implementation, so-called image pyramids, relying on subsequent filter and downsampling operations, are employed, e.g., [233]. These ensure that larger-scale features such as the specimen shape are taken into account as well.

Subsequently, using the landmark approach, a transformation is inferred to commonly transform every data set from the intermediate target stage (SE2 before fatigue) to the target EBSD data stage (4). To derive this affine transformation, common image features related to volume defects in the SE2 before fatigue and the SEM signal channel of the EBSD data are utilized. However, the superposition after this affine transformation is non-ideal, since the EBSD data contains spatial, non-linear distortions. The 70° specimen tilt during the EBSD measurement resulting in a position-dependent working distance, in conjunction with electron Larmor precession, causes such distortions [234]. These are amplified by specimen surface curvature that is prominent in the vicinity of specimen edges. The polishing introduces deviations up to 3° from an ideal planar specimen surface (see figure 3.9). Additionally to the spatial distortions, the orientation measurements are affected if large sample surfaces are scanned, see [235, 236]. Primarily, the spatial distortions in EBSD-inferred microstructures impede the correct assignment of damage derived from comparatively undistorted image sources to its underlying microstructural feature. Overcoming this challenge is an essential step towards performing reliable microstructure-property relationship analysis when dealing with large-area EBSD scans. This motivated the acquisition of the preliminary undistorted and undeformed reference since it allows for a spatial distortion correction of EBSD data as proposed by [236–238].

Finally, the correction of the EBSD data (5) takes place by computing its elastic transformation field relative to the *affine transformed* SE2 image before fatigue, illustrated in figure 3.8f. Therefore, a b-spline optimization incorporating landmark selection was performed, following [239]. The employed implementation in the software imageJ is called *bUnwarpJ*. This elastic transformation utilizes the minimization of an energy functional consisting of several weighted energy terms. Namely, the terms are represented by dissimilarity energy between the images, optional landmark constraints, regularization, and bidirectional consistency. For the purpose of registration, the dissimilarity term attempts to minimize the pixel intensity difference between the warped source image and the target image. The regularization term ensures gradual displacements without discontinuities in regions where landmark information is absent. Landmark information from selected point correspondences

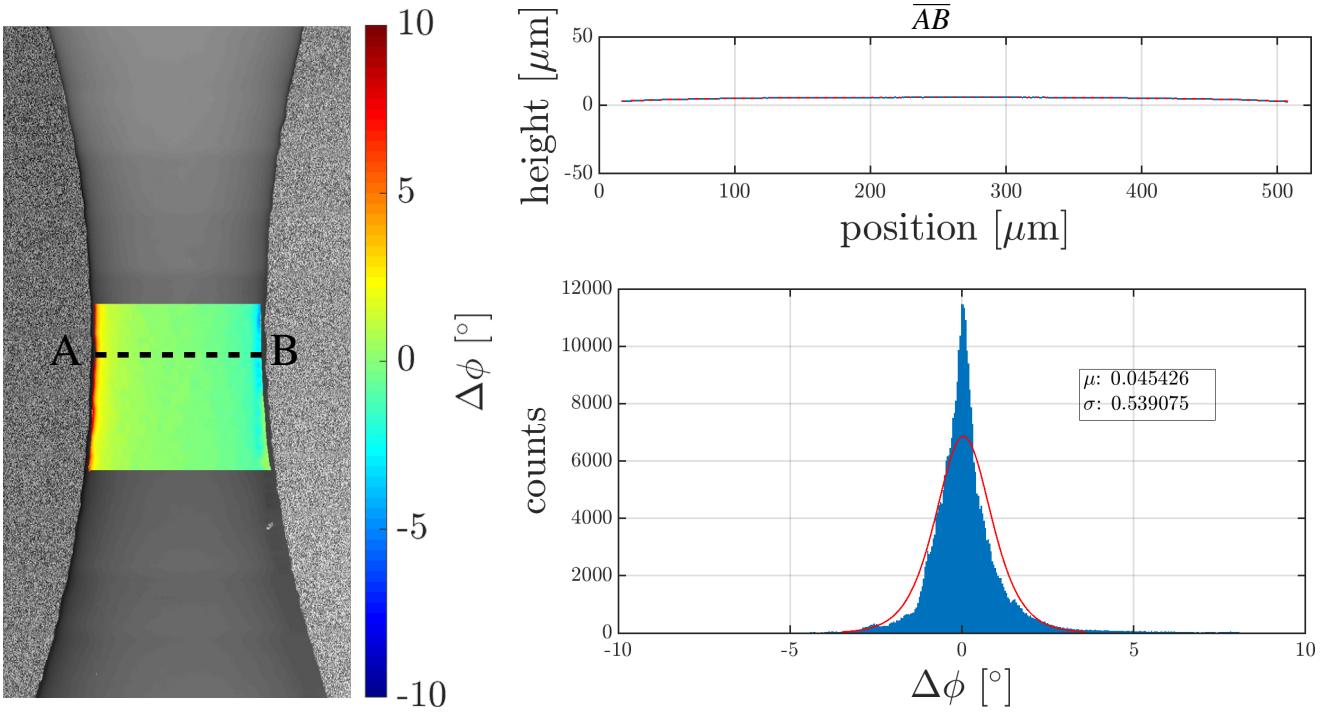

Figure 3.9: Overview of the specimen shape captured by laser confocal scanning microscopy. A heatmap superimposed onto the specimen, in the highly-loaded, tapered region shows angular deviations (partial derivatives in width direction) from an ideal plane surface. The path $\overline{AB}$ is plotted to scale in the top right subfigure. In a histogram, the distribution of the angular deviations for the whole heatmap is illustrated.

constrains the deformation. Depending on the weight applied to it, it can act as a soft or hard constraint. The bidirectional consistency energy refers to the inverse transformation and ensures the invertibility of the deformations. Images and displacements are constituted by cubic b-splines, as proposed by [240]. This information is translated to pixel-wise displacements to determine each sampled data point's corrected positions on the hexagonal EBSD grid. After computing the updated positions, the EBSD attributes (Euler angles, confidence index, image quality, etc.) are assigned to the closest grid point without altering the grid point positions. Subsequently, data cleaning and grain reconstruction as described in section 4.3.2 are applied.

3.6 Data fusion for segmentation of damage in light optical images

The objective of the in-situ image series is to derive the evolution kinetics of individual damage instances. Therefore, the damage locations need to be segmented correctly and the damage amount needs to be quantified in every image of the time series. In order to achieve this, an oblique illumination method is proposed for in-situ imaging since it achieves a better contrast and better directional resolution as compared to brightfield lighting [241, 242], which facilitates retrieving information on the surface defect height and enables segmentation. If uni- or bidirectional oblique illumination as described in section 3.1 is concerned, shadowing effects can occur due to the high aspect ratios present at slip markings and cracks.

The shadowing prevents a gray value-based thresholding segmentation, as illustrated in figure 3.10a and 3.10b. In this illustrative case, the segmentation threshold was selected automatically by Otsu's method [243], which attempts to minimize the intra-class variance in both the foreground and background classes. Even though the protrusion is continuous, see figure 3.10c, the segmentation leads to three separated regions. Presumably, this issue can to some degree be alleviated by increasing the azimuthal illumination directions in light optical imaging. Conditions of the illumination were chosen as indicated in figure 3.1 in order to promote the protrusion visibility regardless from their topography. However, due to non-symmetric illumination conditions, protrusions with a suitable topography appear relatively brighter than others. Therefore, a relative comparison of protrusion height between different protrusions remains difficult. This will be more thoroughly investigated in section 5.2.

(a) Oblique illumination microscopy **(b)** Threshold segmentation (Otsu's method) **(c)** Everhart-Thornley SEM

Figure 3.10: Segmentation of an extrusion based on a gray value threshold compared to the actual geometry from SEM.

The inverse issue, where two damage instances appear connected in the light optical images, arises as well, see blue arrow annotation in figure 3.11. This can be ascribed to the limited resolution or wear debris at the surface. In figure 3.11, spatially aligned and superimposed segmentation masks of light optical and SEM images are depicted for a highly loaded region, see figure 3.1b. The SEM damage map acts as a ground truth. It is not shown in a class-resolved state but rather unifies slip markings and cracks in the magenta color channel. The binarization threshold of the light optical modality (green) was selected to maximize the agreement in terms of the IoU metric (see equation 2.24) with the reference SEM damage mask. Regions of overlap between the segmentation masks of both modalities are represented in black.

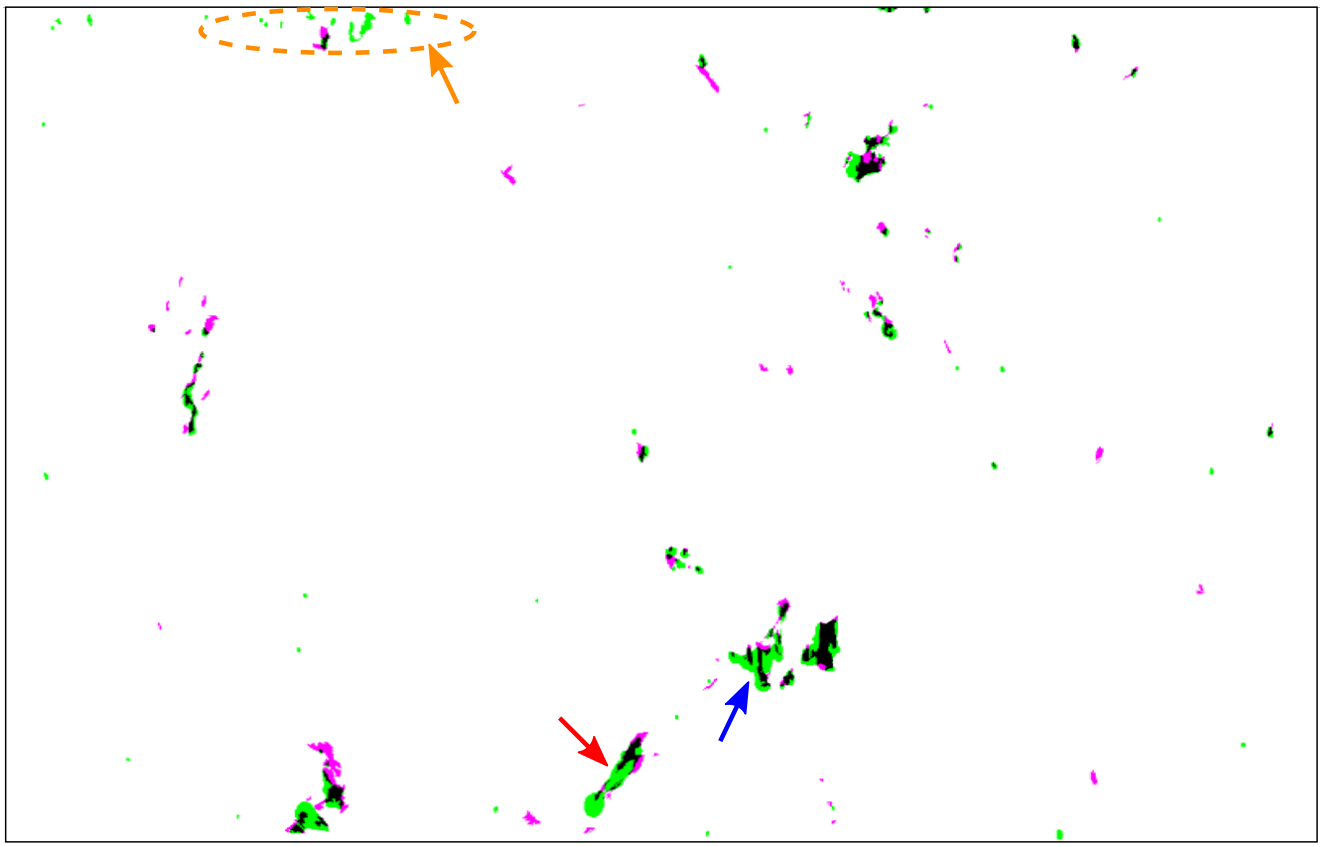

Figure 3.11: Comparison of the manually labeled SE2 segmentation mask (magenta) used as a reference with a binarized difference image (green) from the light optical image before and after fatigue. The binarization threshold of the latter was chosen such that the intersection over union metric with the reference image is maximized. Black regions indicate intersections between both modalities.

Both segmentation masks are generally in good agreement. This applies especially to the larger damage instances which are likely to dictate cyclic failure. There are few segmented areas in the reference that are not accompanied by instances in the light optical images. Mostly these cases correspond to spatially confined and shallow protrusions. Light optical damage instances that are not attended by the corresponding reference ones arise on one hand due to overexposure at the upper specimen edge (orange annotation in figure 3.11) in the automatically acquired SEM stitch maps. In these regions, no appropriate ground truth can be deduced. On the other hand, the light optically detected instances annotated in orange could be false positives. Specifically, particles settling at the specimen surface during the experiment are considered in the light optical segmentation as opposed to the SEM-based segmentation. Furthermore, shadowing of defect locations can occur due to the oblique illumination which does not pose an issue in SEM imaging and segmentation. Intuitively, the substantial directionality of the oblique illumination calls for a pixel-wise, asymmetric distribution of segmentation thresholds or sophisticated background subtraction techniques [244]. This was not implemented in the current version of the post-processing routine.

In contrast to the locations, the shape of the segmented instances frequently differs. This can be quantitatively assessed by computing instance-wise IoU values [245] and plotting them as a histogram. For instance, the crack instance marked with a red arrow in figure 3.11 appears broadened in the light optical modality. It can be ascribed to the debris originating from friction and fretting of the crack faces. In post-mortem SEM analyses, such debris was predominantly observed at microstructurally short and physically short cracks. The formation of this wear debris is closely linked to the fully reversed bending loading and to asperities of crack faces [246]. Previously, friction and roughness-based effects were demonstrated to attenuate the growth of slip-mediated crystallographic crack growth. During manual annotation of SEM damage instances, the debris was considered as background.

Despite their distinct topography, the distinction of protrusions and cracks in the light optical images poses a challenge. This applies especially to marginal cases such as the early stages of microstructurally short crack initiation and growth when the crack aspect ratio and opening is still small. Plastically deformed regions and extruded volumes accompanying such short cracks can also hamper the differentiation. Interfaces between highly plastically deformed regions and the largely unimpaired matrix often act as precursors to crack initiation [58]. In this case, locally recessed regions, so-called intrusions, are typically considered to be nucleation sites for cracks. These fine details in the slip marking topography suffer from shadowing effects of neighboring protrusions and are typically not resolvable in the images. On the other hand, plastic deformation often arises as a consequence of stress concentration at the crack tip [247]. This type of extruded volume, however, promotes the identification of cracks since it is typically confined to the immediate crack vicinity, especially in earlier stages of crack growth, and acts as a scattering site for the incident light.

As opposed to light optical images, SEM images allow for straightforward distinction of cracks and protrusions due to substantial differences in their pixel intensity profiles and image texture. In the SEM image, even with moderate magnifications of a few 1000×, subtle topographic information within slip markings can be resolved. Moreover, despite the increased depth of focus in SEM, unless the crack opening is pronounced, typically no secondary and backscatter electron signal can be retrieved in crack regions which culminates in characteristic dark regions.

In this work, it was attempted to alleviate these inadequacies in imaging through multimodal data registration. Specifically, the attempt included transferring the high spatial resolution and class-resolved information from post-mortem SEM segmentation maps (see section 3.3) to aid the segmentation of the light optical images. The implementation uses soft constraints by applying distinct thresholds for the different semantic categories transferred from SEM. The three threshold values for the semantic categories (background, extrusion and crack) are chosen automatically to maximize the overall intersection over union between both modalities. Since the registration and segmentation are integrated into a graphical user interface user intervention is possible. This was necessary in some cases to appropriately apply the class-sensitive threshold throughout the whole in-situ light optical image series.

Applying a semantic category-depended threshold matrix to the light optical images enables not only capturing more damage instances but also effectively suppresses false positives originating from particles. This is the case since particles are considered as background during the annotation of the SEM image and the background is assigned a higher threshold during optimization. Furthermore, the damage class labels are adopted from the SEM images. This is implemented such that, in

case of spatial overlap between a light optical intensity peak and an SEM crack instance, the segmented region of elevated intensity is considered cracked.

3.7 Analysis of damage evolution and crack initiation kinetics

In this section, the data processing methodology for evaluating the damage evolution based on the in-situ image series is introduced. Specifically, a rule-based computer vision pipeline is designed encompassing morphological and filtering operations along with the previously introduced segmentation to achieve tracking and quantification of damage instances.

As the surface topography of protrusion and crack defects is altered due to cyclic loading, their image representation evolves along with it. When protrusions are concerned, the change typically manifests in an increase of intensity which is monotonous in a first approximation. This is postulated to coincide with a height increase. The correlation between protrusion topography and its image texture in the applied oblique illumination light microscopy has been investigated and is described in section 5.2. In contrast, the primary mode of damage evolution for cracks is their propagation. This is reflected in the images by spatial propagation as well as alterations in intensity owing to plasticity and wear debris at cracks.

This section describes how the elevated intensity regions in the registered image series (see section 3.4) are evaluated. For the following stages, rather than using the full frame, as-acquired images, the images after multimodal registration are utilized which are confined to the region investigated by EBSD. In the following, the symbols as introduced in section 3.4 are used which here describe the registered images instead.

The final image of the series serves as a reference that contains information on damage instances and provides a starting point for the algorithm. Initially, as a pre-processing step, frame differencing was applied where the initial state was subtracted. Assuming an ideal registration, this difference image contains only information about all alterations that occurred during cyclic testing such as damage accumulation and particles adhering to the surface. Static image features related to pores, fabrication-induced surface contamination, and specimen edge are then effectively suppressed.

Subsequently, the difference image is segmented according to section 3.6. Every distinct elevated intensity instance in the resulting binarized image, see the green channel in figure 3.11, is indexed with consecutive numbers. Subsequently, connected foreground and embedded background regions smaller than five pixels are discarded. Such artifact regions can either arise due to impulse noise from defective pixels and memory errors [248] or due to subtracting frames with residual image shifts, i.e., non-optimal registration. Therefore a pixel connectivity of eight is specified, which means that adjoining horizontal, vertical, and diagonal pixels are considered connected. The result effectively represents an instance segmentation of damage locations.

The semantic damage category of each instance is determined by identifying its overlap with the segmented SEM images and applying a voting scheme. Since the light optical and SEM images have distinct physical pixel sizes, resizing with a bi-cubic interpolation scheme is applied to the SEM segmentation mask. Multiple damage descriptors were extracted for each instance including physical area, the aspect ratio of a fitted ellipse with the equivalent area, and accumulated as well as median pixel intensity within the segmented instance. In the case of cracks, a skeletonization step of the crack instances was performed to obtain an estimate of the crack length.

For the reference difference image ($\mathbb{I}_K - \mathbb{I}_0$), each set of pixels establishing an instance was stored in a list. In contrast, for the subsequently analyzed difference images, this pixel list was employed to assign contained damage instances to the ones in the reference image. Cracks observed in a final image sometimes result from bridging sub-cracks that were originally separated by one or multiple grain(s) posing a barrier. In such cases, the instances ascribed to the sub-cracks are assigned to the final connected crack instance in the reference image and their damage descriptors are aggregated. Analogous steps were performed for protrusion instances.

Images were analyzed in anti-chronological order, such that in the second iteration, $\mathbb{I}_{K-1} - \mathbb{I}_0$ was analyzed. The damage descriptors for each difference image and each instance are stored in a three-dimensional array. Assessing the degree of damage through such light optical difference images in a quantitative fashion is not straightforward. While the accumulated

pixel intensity damage metric is presumably associated with the protrusion volume, the median pixel intensity is assumed to contain information about its height. The asymmetric illumination conditions and the specimen surface curvature hamper the quantitative analysis as topographies and positions on the specimen play an important role in the total integrated scatter at damage locations [214, 249]. In particular, this complicates comparing different protrusion instances quantitatively. Nonetheless, an empirical study presented in the results section showed a decent correlation between specific topographic damage features with intensity over a set of protrusions. Moreover, evaluating the cyclic evolution of individual protrusions is feasible nevertheless, assuming that no severe changes in their topography cause shadowing.

When plotting the accumulated pixel intensity I for an individual protrusion as a function of the cycle number, typically trends as depicted in figure 3.12 are obtained. Alongside, the protrusion area A and area-averaged accumulated intensity H of a protrusion is plotted. All values are normalized with respect to their concluding values which is indicated by $\overline{(\cdot)}$. The asterisk indicates fitted curves.

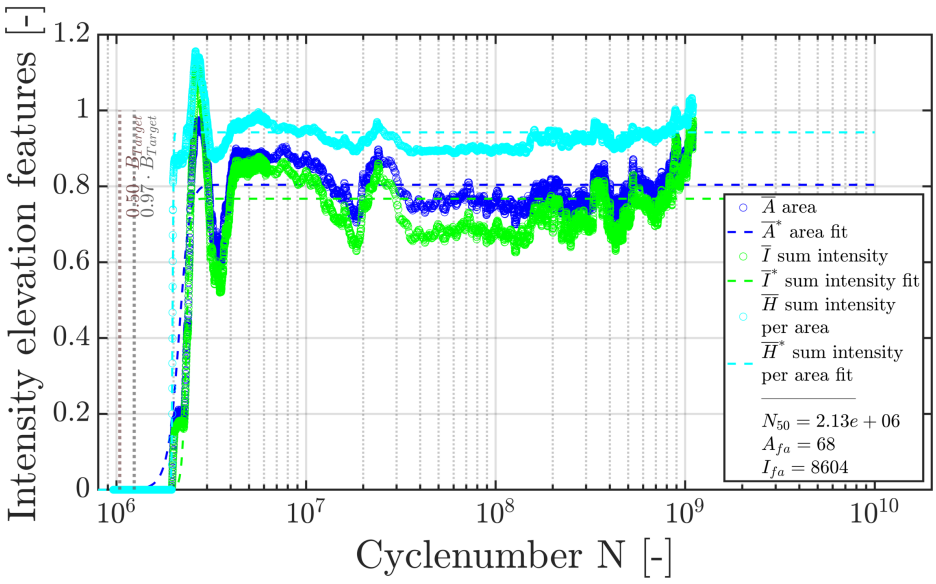

Figure 3.12: Damage features for an exemplary protrusion and fitted logistic functions. The two dashed vertical lines indicate the settling phase of the amplitude control.

From the evolution of accumulated intensity, specifically the early increase and saturation, it seems that the protrusion formation occurs fairly early with respect to the total fatigue life. This is in agreement with observations in the literature on extrusions in different materials [60, 222] where a static protrusion occurred early followed by subsequent roughening. Another observation that can be made is that fluctuations in the saturated portions mainly arise due to area fluctuations since the fluctuations are attenuated in $\overline{H}$. These area fluctuations can be ascribed to topography evolution, non-ideal illumination/blurring, and image noise in conjunction with threshold segmentation. The intensity curve progressions can be modeled by generalized logistic functions of the form

$$I = \frac{I_c}{1 + e^{-\kappa(N - N_{50})}} + I_{bias}. \tag{3.1}$$

The fitting parameters I, I_c, I_{bias} κ, N_{50} represent the intensity, the intensity increase, an intensity bias term, growth rate, and cycle number at half intensity. The performed fitting method uses a least-squares approach to optimize the fit by applying the trust-region-reflective algorithm [250]. Analogously, the curve fits for the area, and the area-averaged accumulated intensity are conducted.

A similar approach utilizing a limited exponential growth formulation was proposed in [251] to model the damaged surface portion in thin films. The underlying assumption of such a formulation with a single growth rate is that the growth of protrusions is governed by only one mechanism. The validity of such an assumption is assessed in the discussion section. Moreover, the relations of such curves to different stages of protrusion growth [221] and microstructural features are addressed there. Initial estimates and threshold values for the fitting procedure are summarized in table 3.2.

Table 3.2: Fitting parameters applied to approximate experimental intensity evolution data related to protrusion and crack formation and growth with equation 3.1.

	Protrusion			**Crack**		
Fitting parameter	Lower bound	Ini. guess	Upper bound	Lower bound	Ini. guess	Upper bound
I_c	0.75	1.00	2.00	0.75	1.00	2.00
κ	1E-6	1E-5	1E-4	1E-10	1-E-6	1E-4
N_{50}	N_0	$N(\Delta I_{max})$	N_{end}	N_0	$N(\Delta I_{max})$	N_{end}
I_{bias}	0.00	0.00	0.05	0.00	0.00	0.10

4 Experimental

4.1 Material and specimen specifications

Exemplarily, in this work, the focus lies on a ferritic stainless steel material EN 1.4003 (AISI 3Cr12). In table 4.1 the average values in weight % of the alloying elements obtained from spark emission spectroscopy are summarized. The high chromium alloying content renders this material corrosion-resistant and enables applications e.g. in hydrogen environments.

Table 4.1: Average chemical composition of ferritic stainless steel 1.4003 in weight %.

Material	C	Si	Mn	P	S	N	Cr	Ni	Mo
EN 1.4003	0.013	0.67	1.08	0.018	0.021	0.013	11.9	0.43	0.33

The material was supplied as a rod with a diameter of 21 mm, which previously underwent hot rolling, grinding, cold drawing, and annealing. Monotonous, macroscopic tensile tests in rod drawing direction revealed a macroscopic elastic-plastic transition which is dominated by the formation of Lüders' bands in the range 265–280 MPa (lower and upper yield strength) [252]. Furthermore, the material exhibits a Vickers hardness of 145 $\pm$ 3 HV1 in the region of the rods from which the specimens were extracted.

Planar, mesoscale fatigue specimens were produced from the as-received rod material. The primary objective of this work was to develop a methodology to investigate the microstructure dependence on the HCF response of such a bcc material. Therefore, some conditions set on the specimen preparation techniques comprise minimal induced residual stress, surface roughness, pitting, waviness, notching relief, and the absence of scratches. These surface defects can lead to stress concentrations and thus can mask stress concentration effects arising from the microstructure. Moreover, such defects can act as sites for dislocation generation and distort the material's response. Surface-sensitive electron diffraction techniques, such as the employed EBSD, require material-dependent surface finishes. Namely, the absence of oxide layers, contamination, lattice distortions, and minimal topography is recommended [119]. On the other hand, unvarying features at the specimen surface need to be present which are visible in the utilized microscopic characterization techniques. This requirement can be ascribed to correlative microscopy, which is relying on feature correspondences to align images from different modalities accurately as introduced in section 3.4.

Apart from the surface properties, specimen geometry poses an important factor. This is due to its influence on stress distribution and on resonant modes in the frequency domain. The planar mesoscale fatigue specimen geometry is illustrated in figure 4.1. Corresponding nominal geometry parameters for its clamp, beam and mass sections are summarized in table 4.2. The beam shape was adapted to avoid stress intensity at the beam transition to the constraint clamp area [211]. This was achieved by employing the so-called tensile triangle method introduced in [253]. It led to a relatively homogeneous stress distribution across the tapered part of the gauge section. At the same time, the geometry was optimized through modal FE analyses to solely induce out-of-plane first-order bending motion while suppressing in-plane bending or torsional modes. The following fabrication methods were chosen to comply with the above-mentioned surface-related and geometrical conditions.

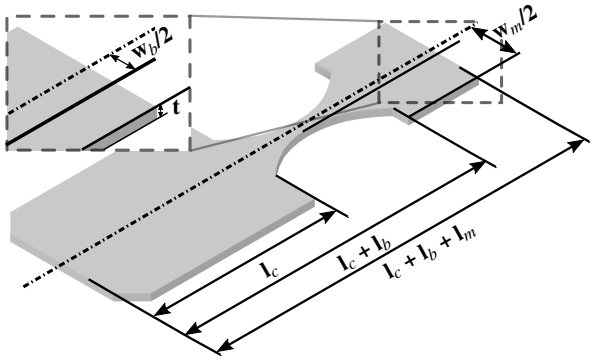

Figure 4.1: Isometric view of mesoscale specimen with relevant nominal parameters indicated.

Table 4.2: Target parameters for specimen preparation.

section	width w_x [mm]	length l_x [mm]	thickness t [mm]
clamping (c)	6.00	7.8	0.35
beam (b)	0.48	5.00	0.35
mass (m)	4.00	3.00	0.35

4.2 Specimen preparation

4.2.1 Electrical discharge machining and laser cutting

In this section, the experimental procedure is outlined to extract the aforementioned mesoscale fatigue specimen shape from the as-received rod material. The specimens' fabrication employed different processing steps, including electrical discharge machining to fabricate disks and laser cutting to define the specimen contours, shown in Figure 4.1. Wire electrical discharge machining was performed with a single pass of a brass wire of 0.02 mm and culminated in disks of 650 µm thickness. A single pass was considered sufficient since subsequent polishing was necessary to achieve surfaces apt for EBSD in any case. An ultrasonic cleaning step of approximately 1 min duration in 15% citric acid ensured a surface quality suitable for the subsequent laser cutting.

Two planar specimens were extracted from each disk such that the rod axis (i.e. rod drawing direction) was oriented orthogonally to the specimen plane. The positions at which the specimens were extracted were chosen such that a fixed distance from the disk center is kept to avoid regions with higher segregation emergence in the core of the rod. Aside from this, due to disk area constraints, the specimen was neither oriented radially nor tangentially within the disk. An allowance of 20 µm was used as opposed to the nominal geometry to compensate for the later electropolishing step which reduces the lateral dimensions as well. After pulsed laser cutting the heat-affected zone from laser cutting was removed from the specimen sidewalls. Initially, through metallographic cross-section investigation and Nital etching, the region that exhibited an altered microstructure was found to be approximately 10 µm wide. Such heat-affected surface layers, especially when mesoscale specimen, bending loading, and HCF are concerned, can impact materials properties significantly.

4.2.2 Metallographic polishing

For metallographic polishing, the specimens were then separated from the laser-cut disks. A thickness measurement with a micrometer screw was performed to sort the specimen into batches of five with uniform thickness. To increase the specimen throughput each specimen batch was bonded to aluminum cylinders for parallel processing. For this purpose, the aluminum cylinders were heated to 150°C on a hotplate and a thin film of epoxy resin adhesive was applied by melting it. The prerequisites for the adhesive are mainly negligible deformation during polishing and sufficient bond strength at both interfaces (specimens and cylinders) to avoid failure during the rotatory polishing where normal and shear loads act. Moreover, cavities

formation along the bond interfaces or in the bulk of the adhesive can be detrimental. Individual specimens were placed radially on the aluminum cylinders. Then the cylinders are transferred individually to a hydraulic compression stage where a force corresponding to 3 bar hydrostatic pressure is enacted onto the specimens via a piston. The hydraulic stage is then transitioned into a water bath for 30 s, where the adhesive undergoes initial curing. This process serves the purpose to obtain a uniform, parallel specimen embedding within a homogeneous adhesive film.

Subsequently, grinding with silicon carbide paper with a grit size of 4000 was performed on the first specimen side. Since the material in advance was confirmed to contain no metastable phases such as retained austenite, medium polishing forces of 30 N were applicable. Moreover, a continuous stream of water ensured lubrication and cooling throughout the rotational grinding process. The grinding process was prolonged for those aluminum cylinders which contained the batch with higher initial thickness. Successively, during the fine grinding and polishing stage, the grain size of the abrasive polycrystalline diamond particles in suspension is reduced from 9 µm to 3 µm. For both processes, distinct metallographic cloths, Struers® MD-Allegro and MD-Dac, were utilized, respectively. Both cloths were moistened slightly in advance and the corresponding diamond suspensions were supplied repeatedly. The forces were reduced to 20 N. The polishing duration for each process was adjusted to achieve the target removal which was monitored between each polishing step. When changing to a different grain size, the polishing machine and specimens were cleaned thoroughly to avoid carrying over abrasives from the former steps. This concluded the initial polishing of one specimen side.

To remove the specimens from the aluminum holders, the specimen holders were immersed in an acetone bath. Once the adhesive was completely dissolved, the individual specimens consecutively underwent brief acetone, water, and ethanol immersion baths before being hot air dried. This was performed to remove residuals adhering to the surface. The same process was performed for the other specimen side.

In the following, electropolishing with an electrolyte composed of perchloric acid, 2-Butoxyethanol, ethanol, and water (Struers® A2) was performed. The specimen acts as an anode and a curved electrode enclosing both planar specimen sides as a cathode. A DC voltage is applied to induce the chemical abrasion of the anode due to an oxidation reaction. Aside from removing the heat-affected zones from laser cutting, electropolishing, due to locally higher current densities [254], introduces a slight, beneficial specimen edge rounding circumventing crack initiation at otherwise sharp edges. Moreover, the surface roughness is reduced by removing smoothing asperities through the *anodic leveling mechanism*. These phenomenons collectively promote inference of microstructure-related influence factors. Mechanical stirring was performed by using a magnetic and heating stirrer. This was motivated in electrolyte exchange and to circumventing *pitting corrosion* [255, 256]. A small parameter study was performed to identify an appropriate potential difference, polishing duration, temperature, and electrolyte flow rate.

To ensure ideal surface conditions for electron diffraction techniques, a surface finish on both sides of the planar specimen was performed through 10 min rotational polishing with a colloidal silica polishing (OP-S) using a Struers® MD-Chem polishing cloth. Empirical studies showed that electropolished surfaces provide sufficient but slightly worse quality compared to OP-S-polished ones for surface-sensitive electron diffraction techniques. However, in this case, large specimen areas were scanned where short electron beam dwell times are crucial to achieving time efficiency.

4.3 Material & specimen assessment

In this section, the preliminary material assessment of the characterized steels in their heat treatment condition is described. In order to assess the microstructural properties of the steels and their differences, characterization techniques such as etching processes in conjunction with light microscopy and EBSD on specimen surfaces and cross-sections are employed. This has the purpose of deriving information on texture, phase distribution, grain boundary misorientation, and grain size distribution.

4.3.1 Scanning electron microscopy

To reduce spatial distortions in the reference SEM image and the one after fatigue image after fatigue, high magnifications, long dwell times, and low working distances were utilized here as suggested by [257]. Tile images acquired in such a manner were stitched subsequently to account for the whole highly loaded beam section despite the high magnifications. This procedure described in section 3.2 allows resolving damage features appropriately and specimen-scale imaging without low magnifications, typically associated with pronounced optical distortions [217]. At the same time, drift distortions or scan line shifts were not apparent. Sufficient electrical grounding to remove the charge was ensured by mounting the specimens in the bending resonant clamp with large contact areas. Acquisition of the stitched images was performed on an automatized Zeiss Supra 40VP SEM. During image acquisition with a surface-sensitive Everhart-Thornley secondary electron (SE2) detector, an acceleration voltage of 20 kV, working distances varying from 10 to 14 mm, magnifications ranging from 1500× to 2200× and an image resolution of 2048 × 1536 were applied. The higher magnifications were applied for the post-mortem SEM images to resolve the comparatively finer features of slip markings. Brightness and contrast settings were optimized to cover the whole intensity spectrum while minimizing saturation. A correction of aperture alignment and astigmatism was performed. Due to the scarcity of surface features automatic focus control was not available. Instead, the specimen center region was utilized to focus on residual silica particle agglomerates from polishing and the focus was maintained throughout the stitching image acquisition. These settings enable capturing relevant, distinctive features of the protrusions and microstructurally short cracks, which is a prerequisite for the distinction of damage classes by convolutional neural networks. Such topography-sensitive SEM stitch images were acquired in the highly-loaded regions of the specimen and on both sides of the planar specimen since the bending loading was fully reversed. Therefore, the method introduced in section 3.2 was used.

4.3.2 Electron backscatter diffraction

The EBSD measurements utilized the same Zeiss Supra 40VP SEM which is equipped with an EDAX TSL Digiview EBSD system. For EBSD data, the working distance (WD), scan step size, and aperture were chosen to be 18 mm, 0.6 µm and 60 mm, respectively. The EBSD camera image background was collected and subtracted from the Kikuchi patterns to avoid overexposure allowing the full utilization of crystallographic information. This was done after all parameters were set. Activating high current mode enabled high electron collection rates. Dynamic focusing ensured good pattern quality across the whole region of interest. For indexing individual pixels, conventional Hough-based indexing is performed. After an initial step, in which residual γ phases were confirmed to be absent, solely the body-centered cubic unit cell was used for indexing. Similar to SEM imaging, EBSD data was collected for both specimen sides and the highly-loaded region.

At this stage, the elastic transformation introduced in section 3.5 is applied to correct distortions in the raw EBSD map. Subsequently, as a part of the automated post-processing routine, EBSD pixels that exhibit a confidence index (CI) below 0.05 are discarded, and the grains and grain boundary segments are reconstructed. For different materials, CI=0.1 was reported to be a suitable CI threshold at which 90% of the indexing solutions are correct [119, 258]. In the present work, the primary microstructure containing only the α phase results in overall higher CI values as compared to multiphase materials. Since visual inspections indicated low CI regions to be ascribed to pores and GBs, the threshold of 0.05 was chosen. For grain segmentation, the Voronoi tessellation-based clustering algorithm following [259] implemented in the MTEX toolbox [260] is utilized. Therefore, MTEX version 5.1.1 was utilized. Pixel disorientations between adjacent EBSD pixels exceeding 5° were considered to separate grains. This parameter choice has previously been observed to affect the grain segmentation significantly [17]. Grains consisting of less than eight indexed pixels are discarded to account for noise and false indexed pixels at grain boundaries. This threshold of eight was selected due to the small physical step size with respect to typical grain sizes. A smoothing operation with a half quadratic filter [261] was performed with the objective of reducing noise. During this step, missing data points were recovered through interpolation. Such a filter was previously shown to preserve intragranular misorientation well and therefore facilitate investigation of dislocation structures [262].

4.3.3 Atomic force microscopy

While not part of the proposed workflow, a few atomic force microscopy (AFM) topographic maps were collected to evaluate the in-situ imaging during fatigue, described in the following section. Specifically, these had the objective of assessing whether a *quantitative* damage information can reliably be drawn from the in-situ imaging. Therefore, a Veeco® Dimension V AFM was utilized to conduct scans after fatigue in regions of protrusions at a specimen's surface. A scan size of $55 \times 55\,\mu m$ was appropriate for all target protrusions, and the data samples per scanned line were chosen to be 1024, resulting in a spatial resolution of $r_{x/y} \approx 53\,nm$. For scanning a proportional gain of 30, an integral gain of 15, and a scan rate of 0.5 Hz was employed. The scans were performed in contact mode with a Bruker® RTESPA-300 silicon tip with a nominal spring constant of 40 N/m, nominal tip radius of 8 nm. A front, back, and side angle of $15 \pm 2°$, $25 \pm 2°$, and $17.5 \pm 2°$ at the apex affects which features are resolvable. For instance, the topography of fine intrusions can not be resolved. This, however, is not of major concern here since these contribute to the in-situ image formation to a negligible extent. Following this AFM scanning protocol, eight protrusions were mapped for subsequent correlation with the in-situ light optical images. In terms of post-processing, the open-source software Gwyddion [263] was used. The acquired maps underwent background subtraction by subtracting a fitted plane. Subsequently, the removal of line artifacts was performed by first aligning individually scanned rows by the median difference method. This approach offsets rows such that the median of height differences between adjacent rows becomes zero and preserves large features such as protrusions better. Then, the post-processing is concluded by detecting and removing scar artifacts as well as data leveling by offsetting the height values to shift the minimum to zero [264, 265].

4.4 Fatigue crack initiation experiments

4.4.1 Calculation of deflections for target stress levels

Since the specimen geometries do vary, for each specimen, the controlled position sensitive device (PSD) amplitude corresponding to a certain target stress amplitude at peak deflection is computed. For this purpose, the beam angular rotation and beam deflection at the beam-mass-transition, along with trigonometric considerations outlined in [211] are involved.

An Ansys APDL finite element (FE) code was developed to generate and mesh the specimen geometry, perform a linear elastic simulation, and derive the controlled PSD amplitude at the target stress amplitude from the angular rotation and beam deflection. The specimen geometry was parameterized by the actual beam width and thickness in the tapered gauge section as measured by light optical microscopy. Apart from that, nominal dimensions were taken into account for the total beam length, exact hourglass shape, and mass dimensions. Specimen edge rounding or other fabrication-induced deviations from the rectangular cross-section are ignored in the simulation. In terms of meshing, a parameter study was conducted to find the out-of-plane element size at which the surface bending stress response converged. The element type was selected as higher order 3D 20-node solid elements *Solid186* with midside nodes for parabolic displacements. For the resulting mesh, see figure 4.2a, appropriate element aspect ratios and mesh integrity are obtained.

To replicate the experimentally observed elastic deformation accurately, a modal analysis was performed. All degrees of freedom of the clamp portion were constrained assuming the absence of specimen slip and clamp deformation. After obtaining the modal analysis solution, the modal deformation was iteratively scaled until the maximum von Mises stress matched the prescribed target stress. Therefore, a binary search algorithm [266] was employed. Once the necessary specimen deformation was determined, the bending line was post-processed to compute the target PSD amplitude from the aforementioned trigonometric considerations. The resulting stress distribution in the highly loaded gauge section is depicted in figure 4.2b. It indicates a rather uniform overall stress distribution with slightly elevated stresses in the vicinity of the edge at the tapered section. During the code execution, the specimen geometry is stored as an IGES file, which is then considered during multimodal data registration (see section 3.5) to spatially relate microstructure and specimen geometry. This represents input information for follow-up crystal plasticity simulations conducted elsewhere. Additionally, element-wise Cauchy stress tensors are exported into an ASCII file and later used for the computation of loading-related features.

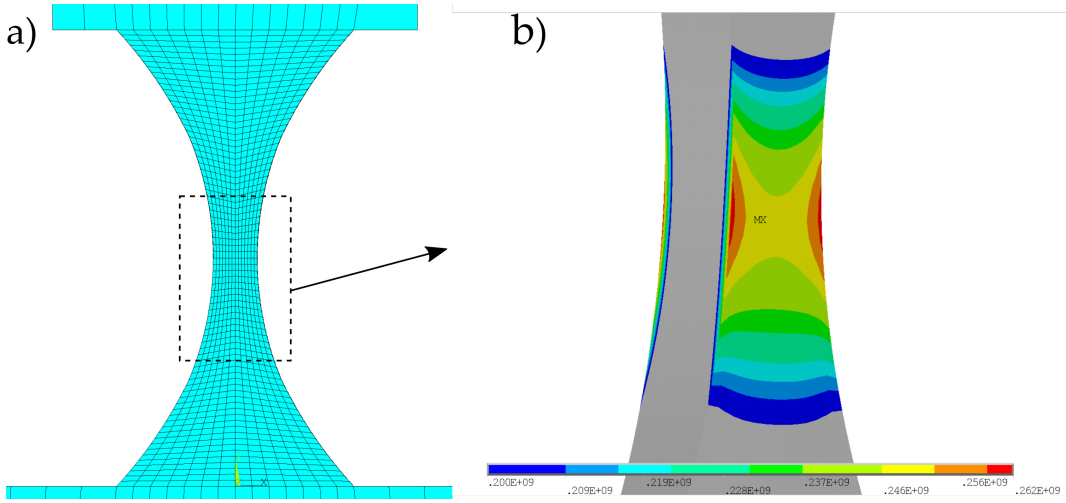

Figure 4.2: Simulation approach. **a)** The resulting meshed geometry showing varying element sizes along the beam. **b)** An oblique view on the highly loaded, tapered part of the bending beam which illustrates the von Mises stress distribution.

4.4.2 Conducting a bending resonant fatigue test

For all bending resonant tests conducted in this work, a rigorous protocol was pursued consisting of the following main steps.

1. **Specimen clamping and alignment**. In this step, the unit consisting of laser, beam splitter, and PSD is shifted using a linear stage equipped with a micrometer screw to appropriately focus the laser centrally at the specimens' beam-mass transition. This alignment was carefully performed after switching to the relevant specimen side (see step 2).

2. **Reference image acquisition**. Acquisition of reference light optical images on both specimen sides before fatigue using bidirectional oblique illumination and continuous illumination. Such stationary specimen images act as a reference to investigate the noise and motion blurring introduced through the cyclic deformation and stroboscope illumination.

3. **Initial frequency sweep**. Performing a frequency sweep to determine the initial amplitude and phase response at low piezoelectric actuator stroke.

4. **Fatigue cycling**. Running a fatigue test until at least the run-out cycle number is defined as $N_{ro} = 10^9$ whilst acquiring an in-situ image series indicating local damage information. Aside, from reaching the run-out cycle number, relative resonant frequency deviations from the initial one were monitored and used as a failure criterion when reaching $\Delta f_{res}/f_{res,0} = -0.1\%$. Neither of the failure criteria was used for actively stopping tests since crack growth did stagnate naturally and terminal failure was not observed at applied loads. In terms of load amplitudes, a range of bending amplitudes was applied on the different specimens which correspond to maximal surface von Mises stresses of 240–293 MPa. The interval between images (t_i) was adjusted in the course of the fatigue process. Initially, an image was acquired every second (i.e. roughly every 2,000 cycles). After settling of the local image textures (i.e., conclusion of the presumable static protrusion [221] growth) at approximately 10^7 cycles, t_i was gradually increased to t_i = 180 s. Such intervals are later empirically confirmed to temporally resolve image pattern evolution associated with subsequent HCF crack growth sufficiently well.

5. **Concluding frequency sweep**. Performing a frequency sweep after fatigue to determine the amplitude and phase response under the damaged specimen conditions at low piezoelectric actuator stroke. By comparison with the results obtained from step 3, the damage-induced total change in resonant frequency f_{res} and quality factor Q can be estimated. While f_{res} is in principle also traced during cyclic testing, the sweeps are performed at substantially lower loading amplitudes such that most phenomena contributing to damping [267] are less relevant.

6. **Post-mortem image acquisition**. Analogous to step 2, post-mortem images of both specimen sides are acquired at the same illumination conditions.

The fatigue setup is enclosed in an opaque casing which suppresses interfering light sources and to some extent, suspended particles settling on the specimens' surface.

4.5 Training of deep learning models for damage segmentation

This section describes the training approach of the DL segmentation models to semantically distinguish slip markings and cracks from the regular surface area in a pixel-wise fashion. It is largely based on a published research article [37] with substantial contributions from Akhil Thomas and the author.

4.5.1 Data variance and augmentation

In section 3.2, multiple sources of SEM SE2 image variance were introduced ranging from fabrication and fatigue-induced variance to measurement-intrinsic variances originating from imaging and operator subjectivity. Extending on this, section 3.3 presented the data sets originating from multiple materials resulting in further material-based variance. These variances affect the resulting image texture distinctly as can be observed in figure 3.5 and figure 3.6.

Typically, deep learning models are known to exhibit poor out-of-distribution generalization [268]. The pursued approach to cope with and generalize to the substantial variance was to train the network with an adapted data set. In this regard, the objective was to render the training data set a more comprehensive representation of the data that is to be evaluated. In this work, the variance was imitated by applying various intensity-based and spatial transformations which can be summarized as warping transformations [269] resulting in an artificially augmented data set. This process is typically referred to as *data augmentation*. During training, multiple altered image instances are created and supplied to the network. This further benefits the learning as it reduces the susceptibility to overfitting, similar to network-focused regularization methods such as dropout and batch normalization. The employed augmentation types comprise different intensity-based transformations, rigid transformations, and image deformations described in table 4.3. Depending on the image augmentation type, the transformations are applied to either only the input image (intensity-based augmentations) or to both the input image and segmentation mask. A comprehensive list of the applied augmentations can be found in table 4.3. The impact of augmentations was investigated in a systematic ablation study by applying the base augmentations (blue background) with individual other augmentations of this table. For performing the augmentations the python library Albumentations [270] was utilized. The different data augmentations were added to an image augmentation pipeline and applied with the specified parameters limits (x_{lim}) and probability (p) on the fly (i.e., online augmentation) prior to training. Rather than conducting a parameter search using libraries such as Tune [271], parameter limits for each transformation were selected based on visual inspection of augmented images. During the inspection, the criterion for appropriate threshold selection was the prevention of substantial visually perceivable information loss and of image distortions rendering the resulting image incomparable to the original data set. The probabilities were chosen based on the frequencies at which these variances are expected to occur in the data.

The objective behind introducing motion blurring was to mimic the directed distortion arising from astigmatism. The elastic transformation was incorporated to replicate present image distortions and morphological variance introduced by different materials (out-of-distribution generalization). Each augmentation threshold was chosen such that the characteristic features of the defects were still recognizable and the labels were preserved. Excluding non-physical augmentations and limiting their parameter space requires domain knowledge about the damage character in HCF and about SEM imaging.

4.5.2 Network architecture and training

The network architecture employed to solve this task is the U-Net, originally developed for segmentations in biomedical applications [197]. It consists of an encoder and decoder section where each level is linked by a concatenation (skip connection) as depicted in figure 4.3.

Table 4.3: Description of augmentation pipeline implemented from the Albumentations framework [270]. Adopted from [37].

Augmentation type	Description: aug. type / parameter(s)	x_{lim}	p
Affine transformation	Linear transformation / rotate, shift and scale limit	30, 0.1, 0.1	0.8
Rotation 90 °	- / -	-	0.25
Reflection	- / -	-	0.25
Elastic transformation	Local deformations / alpha affine, alpha, sigma, approx.	0, 40, 6, True	0.4
Optical distortion	Barrel or pincushion / distort limit, shift limit	0.1, 0.5	0.25
Gaussian blurring	Convolution Gaussian kernel / blur kernel size)	7	0.2
Motion blurring	Convolution motion-blur kernel / blur kernel size)	3	0.2
Gaussian noise	- / var limit	0.015	0.4
Contrast	- / limit	0.15	0.4
Brightness	- / limit	0.1	0.4

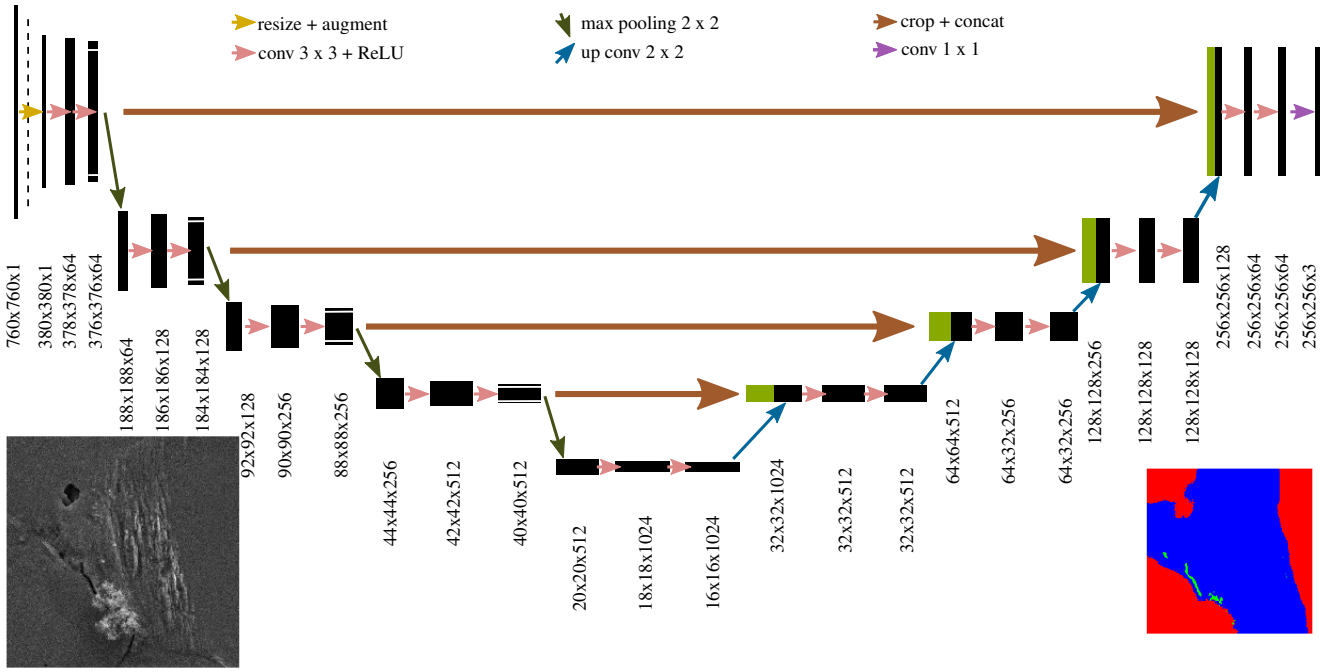

Figure 4.3: Adapted U-Net architecture utilizing convolutional, max pooling, and transposed convolutions in its encoder and decoder section. The cropped concatenation is performed at every stage of the architecture. An exemplary input grayscale SE2 image and segmentation mask are shown in the bottom left and right, respectively. Adopted from [37].

All operations applied in this architecture are introduced in section 2.3.4. The input image is passed into the network from the top left where the augmentation and an optional resizing step happen. Subsequently, the images undergoes multiple encoding stages, where each comprises two unpadded 3×3 convolution layers with **Re**ctified **L**inear **U**nits (ReLU) activation functions and a 2×2 max pooling layer. The alternating convolution and pooling operations facilitate computing multi-scale features which take short and long-range aspects into account. In every stage of the network, the number of feature channels is altered during the convolution steps since different numbers of convolution filters are applied. Each convolution filter (i.e., its weights and bias) is optimized in the course of the training process such that relevant features are extracted. Thus, as opposed to the max-pooling layers, the convolutional layers introduce many learnable parameters to the network. The same applies to the up convolutional layers ('up conv') in the decoding path, which are responsible to restore the high-resolution segmentation mask. The skip (concat) connections help achieve a better fidelity at instance borders since the uncompressed feature volumes from the encoder assist to retrieve the precise localization. Moreover, the skip connections provide a direct path during training (backpropagation of gradients) to optimize the encoder more efficiently.

In order to keep the training time maintainable, during model optimization, the images with a total pixel count of N in the training set were downscaled by a factor of four in area. During training, the loss functions weighted cross-entropy loss, dice loss [272], and focal loss (see equation 2.18) were tried out. Since empirical tests indicated the best segmentation results for the focal loss function, the presented models were trained accordingly. Batch normalization or other types of normalization techniques were not applied.

For training, the whole training image set was provided iteratively in batches of eight images for 300 epochs. This means that the entire training data was provided to the network 300 times in slightly altered versions (data augmentation, see table 4.3). The network was trained from random initialization. The training was conducted on an NVIDIA Turing RTX 2080 Ti graphics processing unit and monitored using Tensorboard by logging training and validation losses as well as intermediate network predictions on the test data set.

In terms of deep learning tasks, three distinct experiments were conducted using the data presented in section 3.3.

1. **Source domain**: In this case, the model trained by supervised learning on images of a specific material is applied to unseen data of the same material (in distribution/domain). $f_{tr} \rightarrow f_{te}$; $m_{tr} \rightarrow m_{te}$ (see Table 3.1)
2. **Material domain generalization analysis**: In this setting aforementioned model is applied to alternate materials domains but on the identical task. $f_{tr} \rightarrow c_{te}/m_{te}$
3. **Multi-domain training**: Rather than training with data from an individual materials domain, in this setting, it was tested whether materials generalization is improved when combining training data from different domains. $f_{tr} + m_{tr} \rightarrow f_{te}/c_{te}/m_{te}$

4.6 Data post-processing

Many methods outlined in chapter 3 are integrated into a graphical user interface (GUI) implemented in Matlab which was utilized to post-process the multimodal data of each specimen side individually. Instead of using scripts, it was opted for a GUI since at some stages throughout the post-processing pipeline, human interaction is required. The data post-processing steps were typically performed after all measurements were completed, as depicted in figure 1.1. In the following, the step-wise approach to data post-processing is outlined.

1. Initially, the specimen folder structure is searched for the presence of specific files to infer the measurement and post-processing stage, and an overview is provided. Amongst others, starting the subsequent steps requires that the *stitched* SEM images exist.
2. A second tab in the GUI is dedicated to all registration processes. This includes the registration of the in-situ time series which is performed first. This registration does not require manual interaction and can be reviewed by sweeping through

the registered images. Subsequently, multimodal registration is carried out. This is realized in the sequence depicted in figure 3.8. For each of the registrations, a few landmark correspondences are selected in Matlab's control point selection tool. For all linear transformations, a few far-spaced landmarks were sufficient. All transformation matrices and selected landmarks are stored. In the case of the in-situ time series, the transformed images are also stored to avoid running the transformation operations repeatedly on multiple thousand images. The elastic transformation to correct EBSD data poses an exception. Namely, the transformation is computed from within ImageJ which is called directly from the GUI. For the elastic transformation many landmark correspondences spread homogeneously across the whole field of view were selected. During bending resonance testing, images are taken of only one side of the planar specimen. Therefore, when the other specimen side is concerned, a slightly altered processing route was employed. In this case, light optical ex-situ images before and after fatigue were registered manually by selecting landmark correspondences.

3. Then inference with the deep learning model for segmentation of damage instances is performed. Afterward, whenever deemed necessary, manual corrections were conducted where subsets of pixels selected with a polygon selection tool were assigned a new damage state. The resulting corrected damage mask was transformed according to the stitched SEM after fatigue.

4. In another tab, the reconstruction, denoising, and smoothing of the microtexture data takes place as described in section 4.3.2. These operations were performed by using the MTEX toolbox. From there, different microstructure descriptors are calculated at the grain level, see table 5.4. Additionally, visualization for macroscopic and microscopic texture, grain boundary disorientations, and other distributions are created in an automated manner.

5. Grain-wise features of pores are computed by applying a threshold segmentation approach on the stitched SEM image before fatigue and aggregating the pore information on a grain level. An initial threshold is provided by Otsu's method [243] which can then be adapted manually.

6. The damage evolution analysis utilizes the corrected SEM-inferred damage mask as a presumed 'ground truth' and applies adaptive, damage class-dependent thresholds on the light optical images to maximize the accordance with the damage mask, see section 3.6. This is performed on the whole image series and culminates in segmented damage instances which are linked between images throughout the whole fatigue process. Subsequently, the curve fitting approach in section 3.7 is applied to each damage instance to assess the damage evolution.

7. Due to minor errors in registration and due to the fact that slip markings were found to localize to grain boundaries, obtaining a well-founded grain-wise damage label required manual validation. For this purpose, each damage instance was depicted in multiple modalities and the preliminary damage-to-grain assignment was corrected whenever necessary. The SEM after fatigue provides both, a grain contrast from polishing and information on surface damage. Whenever a protrusion extended into a grain, that grain was considered and labeled damaged.

8. All microstructure, loading, micromechanical, and defect features are exported along with the grain-wise label as a table which is further processed and utilized in machine learning, see section 5.6.

5 Results

5.1 Material & specimen assessment

The applied experimental techniques, see section 4.3, permit a thorough investigation of the primary microstructure and defect distribution. Structural features at the nanometer scale such as potential grain boundary segregation or precipitate distribution can not be resolved with the aforementioned workflow and characterization methods. Their comprehensive characterization typically requires destructive techniques such as atom probe tomography, nanoscale secondary ion mass spectrometry, or transmission electron microscopy, which can be applied subsequent to the process chain proposed in this work.

The high chromium content of EN 1.4003 amounting to 11.9 wt% renders the presence of precipitates or grain boundary segregation of chromium carbides very probable. For a large range of technically relevant steels, chromium in iron forms a substitutional solid solution. In literature, solubility limits of 9 wt% were reported previously [273, 274]. However, the solubility in technical alloys, is known to vary depending on the overall alloying composition. Indeed, the spotted appearance of light optical micrographs after Nital etching suggested the presence of finely distributed carbides across the whole micrograph in the longitudinal and cross-section of the rod, see figure 5.1a, b.

However, from such LOM micrographs, it is neither possible to indisputably validate the presence of precipitates nor to derive preferential precipitation sites or quantify precipitate size distribution to deduce implications for Orowan strengthening. Therefore, solution hardening, precipitation hardening, or a combination of both could be at play in this material. Moreover, larger elongated inclusions are annotated with arrows in both subfigures 5.1a, b and reach lengths of up to 50 µm in the rod extrusion direction. Additional energy-dispersive X-ray spectroscopy in longitudinal sections of the rod revealed these being manganese sulfide (MnS) inclusion lamellae (type II and type III) [275]. Owing to the mesoscale fatigue specimen fabrication, inclusions and are oriented along the specimen out-of-plane direction. The inclusions and resulting pores present at the surface exhibit equivalent diameters of 0.1–6.0 µm with a peak in the distribution between 0.1–1 µm.

In terms of the primary microstructure, microtexture EBSD images in figure 5.1c, d, representing the longitudinal and cross-section of the rod, reveal anisotropic grain shapes. A single-phase ferrite microstructure with rather large, irregularly-shaped grains is present. The irregular grain shape points towards GB pinning during recrystallization, potentially due to Zener-type pinning at precipitates. This in turn, this hints at larger volume fractions of finely dispersed particles since it favors Zener pinning [276]. The color code in these images represents the three standardized Bunge notation Euler angles as one of the red, green, and blue (RGB) color channels each. In the cross-section subfigure 5.1d, the average equivalent diameter amounts to $d_{eq} \approx 25.5\,\mu m$. The weighted average of intercept length corresponds to 20.3 µm and individual grains reach up to 70 µm in intercept length. The specimen alignment relative to the rod results in slightly elongated grains in the out-of-plane direction of the specimen (5.1c), where the grain intercept length exhibits a weighted average of 20.7 µm and for few instances reaches up to 113 µm (higher variance). In few grains, the material contains crystallographic deformation twins as annotated by the arrow in figure 5.1c. In bcc metals, deformation twinning was reported to occur prior to macroscopic yielding [277]. Moreover, deformation twins were found to be preferred sites for crack initiation in ferritic steels [278]. When relating typical grain intersection length with the specimen dimensions in the tapered section, it can be stated that on average 17 grains are present along the thickness direction and 23 grains in the width direction.

While aforementioned, microtexture maps were acquired from arbitrary metallographic sections of the rod, also larger scans were conducted on the whole highly loaded specimen surface. These scans contain approximately 1.2 million ferrite-indexed

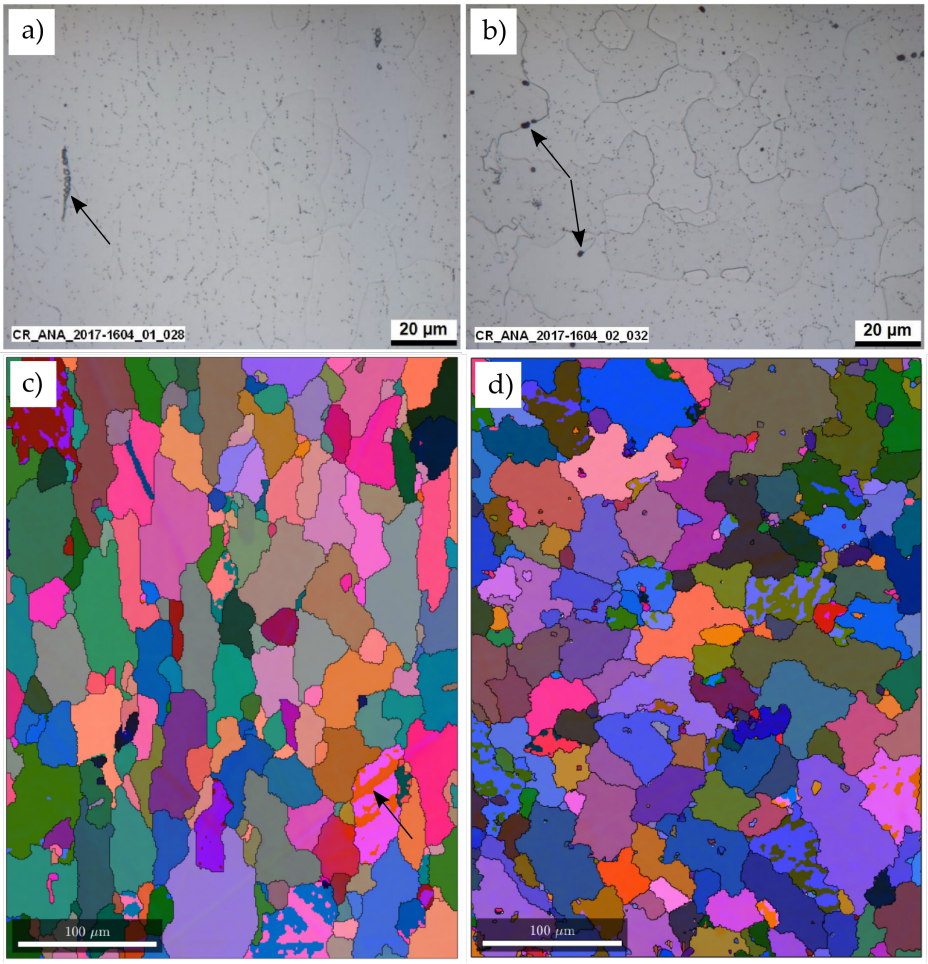

Figure 5.1: Overview of the microstructure. Subfigures **a)** and **b)** represent light optical micrographs of the Nital-etched 1.4003 surface as a longitudinal and cross-section of the rod, respectively. Manganese sulfide inclusions are indicated by the arrow annotations. The spotted appearance indicates the presence of precipitates at grain boundaries and grains. Figures **c)** and **d)** indicate Euler coloring EBSD maps in longitudinal and cross-section, respectively, adapted from [252]. In this color code, the three normalized Bunge notation Euler angles ϕ_1, Φ, ϕ_2 represent the red, green, and blue color channels, respectively. The anisotropic grain shape with elongated profiles in **c)** becomes apparent. Moreover, an instance of a crystallographic twin is annotated in **c)**.

points on a 0.6 µm step size grid and span an area of roughly $520 \times 980\,\mu m^2$. Pole figures plotted from such large scans for two arbitrary specimens' sides reveal a fiber texture in this material, see in figure 5.2.

Specifically, a $\langle 1\,1\,0 \rangle$ fiber texture where $\langle 1\,1\,0 \rangle$ equivalent directions are preferably aligned with the rod extrusion direction (the center of the pole figures) is observed, see the middle column of figure 5.2. This represents a typical texture for such extruded rod materials [279]. The multiple of random (MOR), i.e., how much more often an orientation occurs relative to a uniform orientation distribution is in the order of 2.8–3.6. This indicates a comparatively weak texture. From the comparison of both rows in figure 5.2, one can observe that while trends in the orientation distribution are very similar between the rows, the relative scale of specimens and microstructural domain sizes result in a non-representative macroscopic texture. This general trend was confirmed for both sides of four specimens, i.e. eight such surface EBSD scans.

On the other hand, the grain size distribution and particularly the grain boundary disorientation were found to be largely representative across specimens, see figure 5.3. Note that for the grain size distribution plots, truncated grains at the edge of the specimen or scan area were not removed. The actual misorientation distribution function (MDF) represented as a bar plot is shown to largely coincide with the untextured reference (Mackenzie distribution). Owing to the cubic symmetry, the highest observable misorientation is limited to 62.8° and in untextured materials, the most frequent observed misorientation corresponds to 45° [280]. As a consequence of the slight texture inferred from the pole figures (figure 5.2), the actually observed MDF is slightly shifted towards lower misorientations in comparison to the untextured reference.

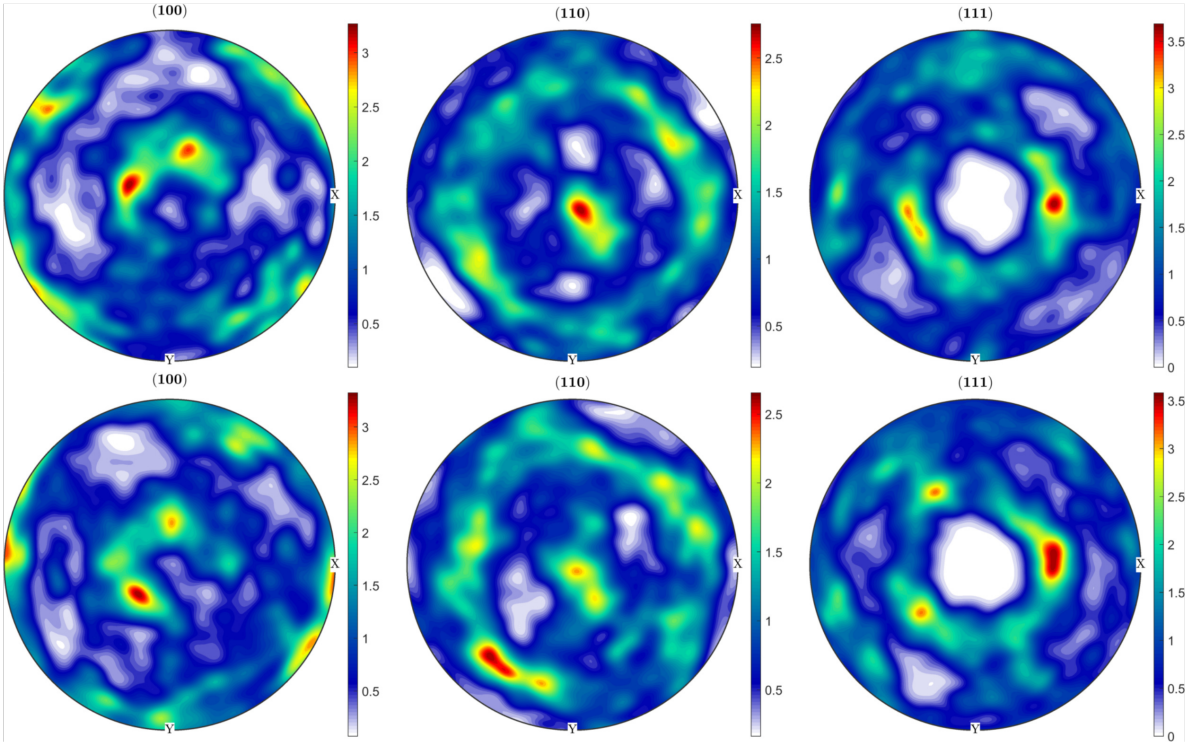

Figure 5.2: Orientation distributions illustrated as pole figures for $\langle 1\,0\,0 \rangle$, $\langle 1\,1\,0 \rangle$, and $\langle 1\,1\,1 \rangle$ crystal reference directions (columns) as well as two arbitrary specimens' sides (rows). The center of each pole figure represents the rod extrusion direction and the edge any normal direction. X and Y indicate the in-plane specimen axis and the specimen width direction.

Lastly, the variation in specimen shape is addressed. Tapered region width and thickness values measured from light microscopy exhibited a Gaussian distribution and corresponded to $468 \pm 5\,\mu m$ and $355 \pm 6\,\mu m$, respectively. Geometry variations among specimens, especially in specimen thickness, lead to differing stress gradients since the bending slope is used as a control parameter during fatigue testing. Such fabrication induced-scatter is inevitable and renders specimen dimension measurements necessary.

Aside from such macroscopic deviations, specimen edge rounding is introduced as shown previously, see figure 3.9. When considering the load distribution shown in figure 4.2, it is easy to assume that such a rounding can cause deviations in testing from the target stress state. Since different present constituents, α-iron and manganese sulfide inclusions, exhibit distinct chemical potentials as well as mechanical characteristics, local pore formation occurs. This occurs due to the electrolytic dissolution of MnS particles at low voltages [281] and load-induced detaching inclusions from the primary microstructure, respectively. The OP-S polishing causes a weak surface relief, i.e., marginal asperities at grain boundaries.

5.2 Damage topography characterization and assessment of the in-situ imaging

There are intricacies involved with in-situ imaging, owing to the high-speed specimen movement and the specific illumination conditions. To characterize the imaging properties, the dependence of its resulting local grayscale image texture on the slip marking topography (i.e., protrusions) is investigated. Therefore, atomic force microscopy (AFM) maps are acquired. Through this, the question is addressed, whether these light optical images can provide insight into not only qualitative but also quantitative damage assessment. Even the character of dislocation slip can be characterized with the AFM technique [69]. Moreover, AFM maps contain information on microscopic topography changes such as polishing-induced surface reliefs at grain boundaries. In Figure 5.4a and 5.4b, a comparison of both modalities is illustrated for two protrusion instances.

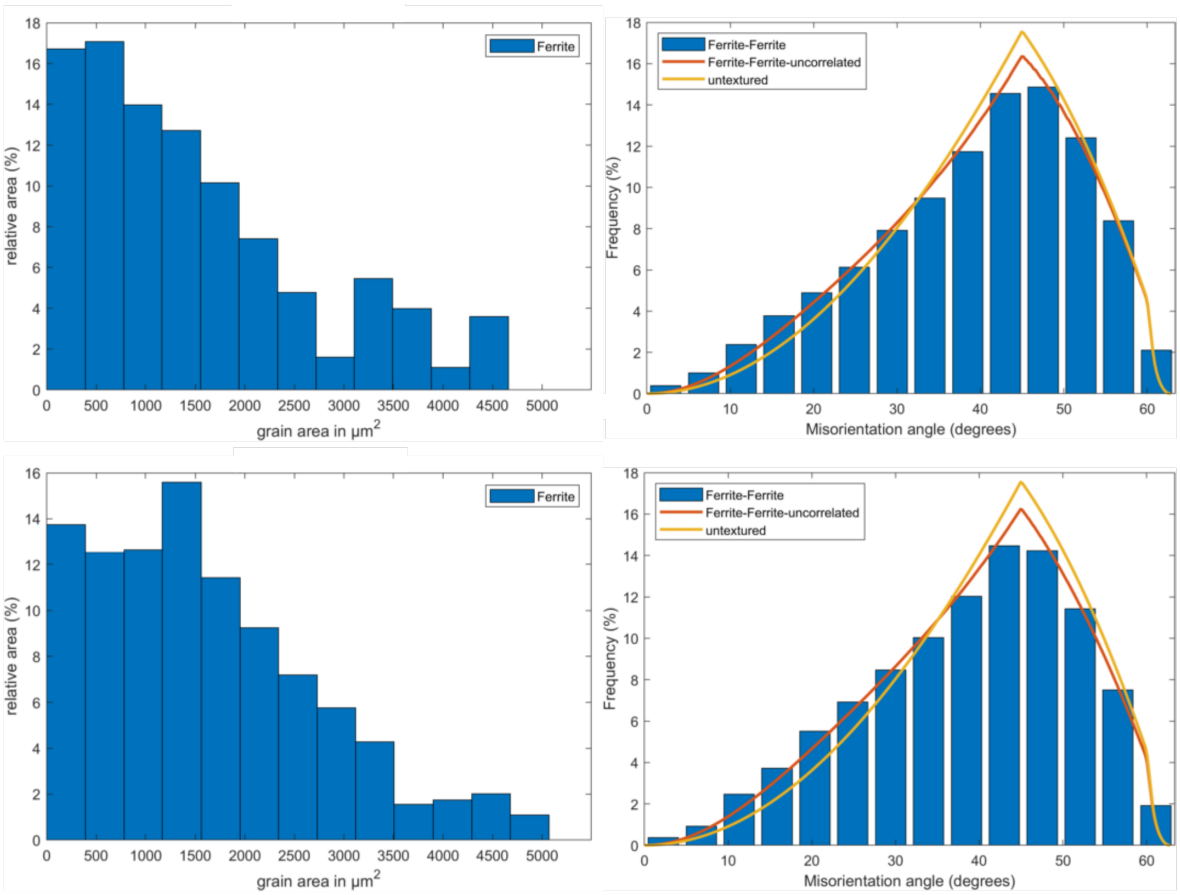

Figure 5.3: The left and right column of these bar plots represent the relative grain area distributions and the measured misorientation distribution function (with an untextured reference), respectively. Each row represents one of two arbitrary EBSD scans in the highly loaded specimen region.

For assessing which quantitative surface feature is captured in the local pixel intensities of the light optical images, multiple topographic features of slip markings were computed. Extracted features from the AFM information included various roughness metrics and geometrical quantities. Namely, the extruded maximum and average elevation, protrusion ground area, volume, and area originating from projecting the protrusion topography onto the axial LED's emitting plane A_{AP} (scattering cross-section) were considered. These features were computed relative to the host grain background elevation. Analogously, for each investigated protrusion, features were computed for the elevated intensity regions in the light optical images. The features comprised elevated intensity area and maximum, pixel-averaged, and pixel accumulated intensity. The data set used in this study comprised eight distinct protrusions. Subsequently, these features were evaluated with regard to their linear correlation using a feature correlation matrix. This indicated that A_{AP} shows the strongest linear correlation with the accumulated intensity (see figure 5.4c). For this pair of variables, the Pearson correlation coefficient was R = 0.88. It can be observed that especially for larger protrusions the data points deviate from the fitted line. The average of the maximum heights of all investigated protrusions amounts to 0.24 µm where the highest measures 0.39 µm. Moreover, from the AFM measurements height elevations of 9–45 nm at grain boundaries can be found considering ten distinct grain boundaries.

In the following, these methodological results are discussed. The image formation is assumed to rely on both specular reflection and diffusive reflection (scattering) at protrusions. This assumption is founded in protrusions superimposed by extrusions and intrusions typically exhibiting a range of spatial frequencies across the roughness spectrum (power spectral density). The scattering portion strongly depends on the ratio between RMS surface roughness (or scale of surface features) and the wavelength [214]. Some surface features resulting from accumulated slip are often comparable to the wavelength of the blue light used for illumination [198]. It is plausible that the correlation of A_{AP} to the accumulated intensity is more pronounced than that of the extruded volume. This can be justified in a first approximation by the accumulated intensity depending on the protrusion cross-section illuminated under the fairly flat light incidence angle. The observation that for larger projected

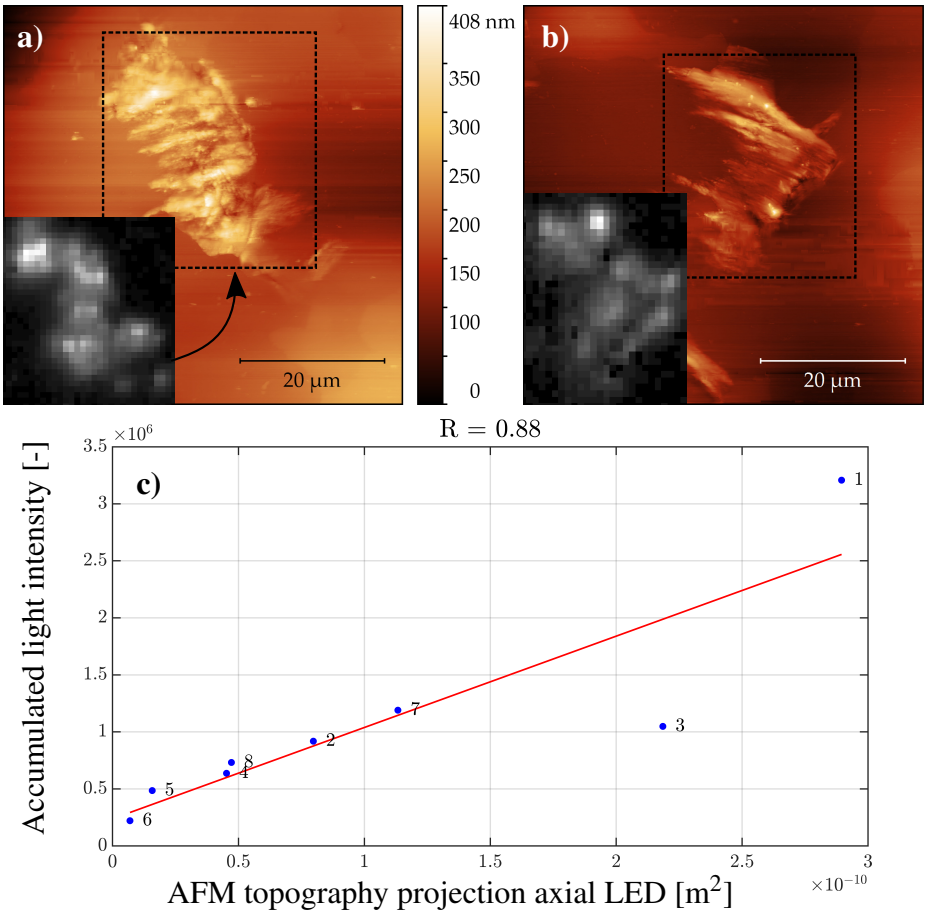

Figure 5.4: a), b) Two exemplary comparisons of protrusions in cleaned AFM topographic maps with corresponding light-optical images after fatigue (inlays). **c)** Correlation plot between accumulated intensity from the light-optical images and the projection of protrusion topography onto the emitting surface of the axial LED. Adopted from [36].

areas deviation from a linear behavior occurs (see figure 5.4c), can potentially be attributed to a more notable dependency of the accumulated intensity on the not considered protrusion surface orientation as protrusion size increases. Moreover, in this study, the position of the protrusion across the slightly bent surface is not taken into consideration.

The different image textures in both modalities, cf. figure 5.4a and b with their inlays, imply a distinct protrusion shape. Arising differences can be primarily ascribed to the bidirectional oblique lighting conditions. The substantial impact of the azimuthal illumination angle on the imaging of a static object was demonstrated in [213]. It was shown that occlusion prevents the realistic imaging of three-dimensional objects when few azimuthal illumination directions are applied. When temporal imaging of dynamically evolving surface topography such as protrusions under motion is concerned, this is further complicated. In large grain size and single-phase materials, these slip markings, as observed here, are often extensive protrusions. Especially, in this case, pronounced topography changes in the course of cyclic loading occur. Then occlusions can sensitively impact the imaging and potentially lead to an overestimation of topography change. On the other hand, the distinction between microstructurally short cracks and slip markings by inspecting local intensity distributions in these light optical images is not always straightforward. In conjunction with the fact that cracks in many materials originate at surface plasticity traces, this impairs the crack initiation life determination from these optical image series. Despite these non-ideal imaging characteristics of the optical setting, microstructural damage emergence can be detected reliably. Many of the aforementioned challenges can be alleviated through data fusion with highly resolved, feature-rich modalities such as SEM and deduced damage segmentation maps. Such damage maps reliably localize cracks, and surface slip markings pixel-wise and hence improve damage type distinction as well as counteract shading for improved damage connectivity inference. The procedure for image data fusion is outlined in section 3.6. In conclusion, this study indicates that utilizing the light optical time series to derive quantitative protrusion information is adequate in a first approximation. This is fostered through the data fusion approach and applies in

particular for smaller protrusions. In the future, improvements of the illumination conditions could facilitate a more reliable quantitative analysis of larger protrusion instances as well.

5.3 Validation of multimodal registration methodology and deformation mechanisms

To qualitatively assess the multimodal registration approach, a comparison between two inverse pole figure color-coded EBSD maps acquired before (figure 5.5a) and after fatigue (5.5b) is performed on the same specimen section. While the former contains the overlayed damage map (black) deduced from SEM imaging after fatigue and hence represents the proposed registration approach, the latter is superimposed with the intrinsic confidence index (CI) channel of the EBSD data. For straightforward comparison and illustration, both damage types, originally differentiated in SEM images by the DL damage segmentation model, were merged in figure 5.5a.

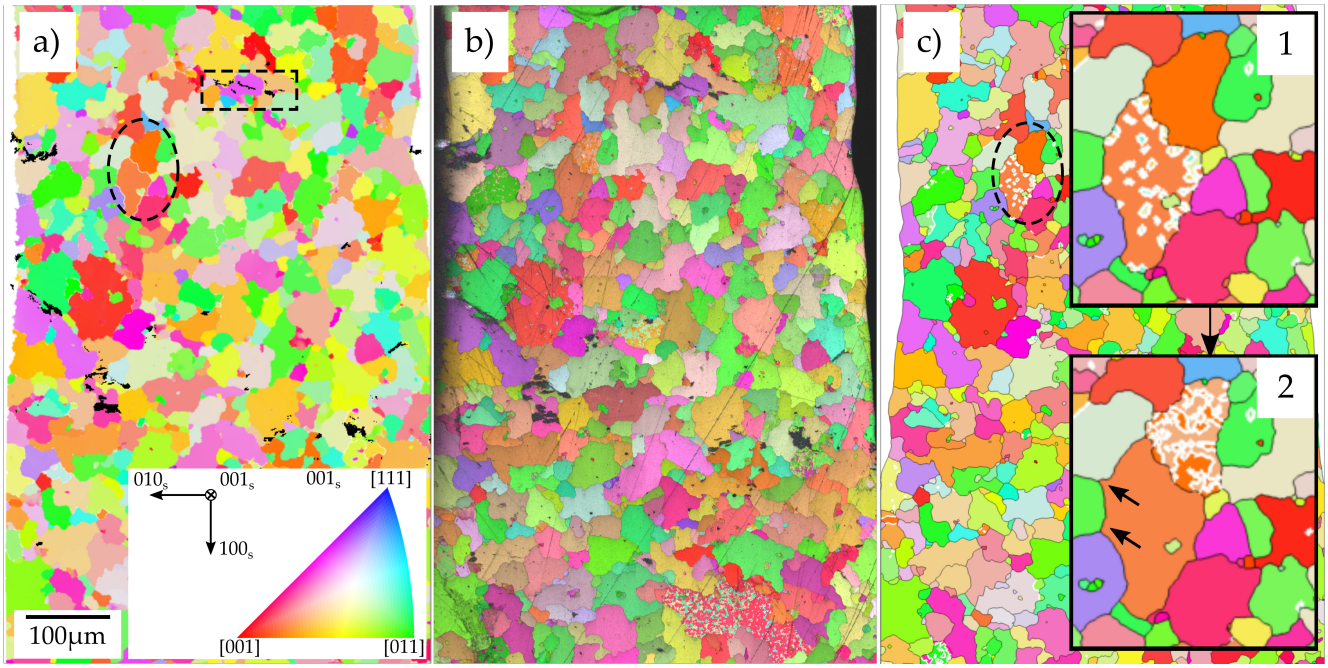

Figure 5.5: a) Inverse pole figure map before fatigue with the reference direction being the normal direction (ND, 001_S), i.e. pointing out of the specimen plane. Damage locations from DL semantic segmentation in black from the SEM image after fatigue are registered and superimposed onto the microstructure. The dashed box highlights a crack location investigated subsequently. **b)** Inverse pole figure map after fatigue in the same orientation setting. In this case, the intrinsic confidence index channel of the EBSD data is superimposed as a gray value distribution indicating damage locations and polishing artifacts. **c)** The same as **a)** but considering twin boundaries and without superimposed damage. The twin boundaries are colored in white. Two inlays show a magnified view of a region (ellipse annotation) with micro twins before **(1)** and after **(2)** fatigue. The EBSD maps were post-processed identically but show a distinct distribution of microscale deformation twins. Arrow annotations in inlay **(2)** show regions of grain deformation after fatigue. The scale bar, the color code reference triangle, and the specimen coordinate system apply to all images, except the magnified inlays. Adapted from [36].

The comparison of both images gives also an indication of the EBSD data processing. Figure 5.5a shows the inverse pole figure color-coded map after the EBSD cleaning and grain reconstruction routine outlined in section 4.3.2. Intragranular misorientations present in few grains are preserved during the cleaning routine. This allows computing meaningful intragranular misorientation features, such as grain orientation spread, for the grain-wise damage classifier. As mentioned in section 4.3, there are microtwins present in the material, see figure 5.5c (white borders).

With respect to a grain-wise ML approach, such microtwins would skew the data set towards small grains. Moreover, in ferritic steels, twin boundaries are known to behave differently from random GBs [282]. Therefore, for computing the machine

learning features, it was opted for an EBSD reconstruction which does not consider twin GB as grain delimiters. Such reconstruction results in a primary microstructure as depicted in 5.5a, where micro deformation twins are largely absent. Instead of considering twin boundaries as grain delimiters, the total twin boundary length contributing to a grain was considered as a numerical feature for each primary grain in the subsequent machine learning modeling. The shown twins are largely incoherent, i.e. their grain boundary plane does not contain the misorientation axis. As a criterion to detect these boundaries, the actual misorientation is permitted to deviate less than 5° (rather than complying with the $\approx 8.7°$ from Brandon's criterion [283]) from the orientation relationship $60°\langle 1\,1\,1\rangle$ (Σ_3 boundary). Aside from the microscopic deformation twins present from processing (see figure 5.1c+d or figure 5.5c inlay 1), deformation twins are present after fatigue as well (figure 5.5c inlay 2). Specifically, the twins, when exposed to high cycle fatigue loading, seem to change their position. This could be related to them dissociating into dislocations [284] which then propagate through grain boundaries. This alteration of twinning sites was found to be accompanied by grain deformation in the immediate neighborhood, cf. inlay and inlay 2 at arrow annotations at GB triple points. Nonetheless, crystallographic dislocation slip is the primary mode of plastic deformation. Thin residual scratches or plasticity traces shown in figure 5.5b were mirrored by traces of slight misorientation in the raw EBSD data prior to cleaning. Neither microtwins nor residual scratches/plasticity traces show spatial correlation with the emerging damage, on a first glance.

From figure 5.5a+b it is evident that multiple fatigue-induced damage locations (black regions) emerge across the specimen microstructure. Typical instances of microstructurally short cracks and protrusions developing the highly chromium-alloyed, ferritic EN 1.4003 steel are depicted in figure 3.5. Apart from some crack sections that propagate in a transcrystalline manner and few protrusions developing at processing-induced pores, the damage is localized to grain boundaries. This is assessed quantitatively in appendix figure A.1 by comparing relative Euclidean distances between real or randomly sampled damage pixels to their closest GB. The reference grain size to compute the relative Euclidean distance was the halved minimal Feret diameter of the corresponding host grain. Moreover, a comparison of figure 5.5a and 5.5b shows that the positions of damage locations with respect to the microstructure are in accordance. This applies to the whole highly loaded specimen region illustrated here, including the vicinity of specimen edges. Differences in shape and extent of individual damage locations arise between both images.

The discussion on the registration methodology is presented subsequently. Damage instances in both images (figure 5.5a and 5.5b) being spatially correlated within the microstructure implies that the applied affine and elastic transformations appropriately correct global alignment and trapezoidal distortion superimposed with further distortion effects. The proposed registration methodology, in contrast to grid calibration, in theory, provides the possibility to correct the specimen geometry-dependent part of the distortions as well. Indeed, even at specimen boundaries where such geometry-induced distortions arise due to edge rounding, the alignment of damage with microstructural features in both images is widely conformal. The multi-stage registration process was employed to exploit pair-wise similarities between the modalities, rendering the assignment of point correspondences easier. Similarly, it was observed that for registration of different SEM images from in-lens, backscatter and Everhart-Thornley type detectors, it is beneficial to use in-lens as an intermediate registration target as it combines information about atomic number and topography and therefore exhibits similarities with both other SEM image modalities. Furthermore, the multi-stage registration using a comparatively undistorted intermediate target (stitched SEM before fatigue) ensured that all modalities are well aligned. In contrast, registering directly to EBSD which contains notable distortions can cause misalignment between modalities. Depending on the degree of deformation and cracking, it can be beneficial to apply an elastic transformation to accommodate for the specimen deformation between the different specimen states captured in both SEM stitch images. The same applies to the registration of the in-situ time series. While for the in-situ registration outlined in section 3.4, the first image acted as the registration target, it can be beneficial to use the final in-situ image as the registration target as well. That way the multimodal registration path can be altered and the post-mortem SEM can be utilized as an additional intermediate target. This is beneficial in the case when there are barely any static features present on the specimen surface (e.g., pores, particles, residuals) since then damage image features can be used for the challenging multimodal registration process.

The good spatial correlation of damage instances between figure 5.5a and 5.5b indicates also that the DL damage segmentation is working well. Nonetheless, there are some discrepancies in the shape of damage locations. These can presumably be

attributed to the fact that the DL segmentation model was trained to detect specific local and contextual image textures in topography-sensitive SE2 images rather than small-scale plasticity in the vicinity of slip markings that do not culminate in a surface change. Moreover, in contrast to the confidence index metric, the deep learning approach can discern surface contamination from damage. Even though protrusions and cracks, as predicted by the DL approach, were merged for visualization, the possibility to reliably discern damage types distinguishes the proposed multimodal workflow from inferring damage from EBSD data after fatigue. Furthermore, comprehensive microtexture information is accessible due to the absence of plastic deformation in the collected EBSD data before fatigue. In the future, automated alternatives for the multimodal registration process are required to avoid the manual selection of point correspondences between both images. However, the appearance of the multimodal images strongly depends on the material and its processing route. In the ferritic steel at hand, after OP-S polishing, mostly pores pose surface features that can be utilized for multimodal registration. However, these features are not only sparsely distributed but also exhibit high similarity hampering their automatic matching between modalities. Having an automated registration concept would render the post-processing chain fully automated. Recently, methodologies that employ CNNs for multimodal image registration were introduced [285, 286] that can potentially alleviate this issue.

5.4 Fatigue crack initiation and propagation experiments

5.4.1 Damage growth kinetics

In the following analysis, the information content of the data sets with respect to cyclic damage evolution is evaluated. Initially, the overall fatigue life is assessed briefly. This solely serves the purpose of providing intuition for the fatigue loading regime and the final degradation state. Afterward, an actual objective of this work is tackled by breaking the total cyclic life down to its mechanistic stages such as protrusion formation, crack initiation, and short crack growth. The S-N plot relating the fully reversed stress amplitude to the number of cycles until failure for all tests is illustrated in figure 5.6.

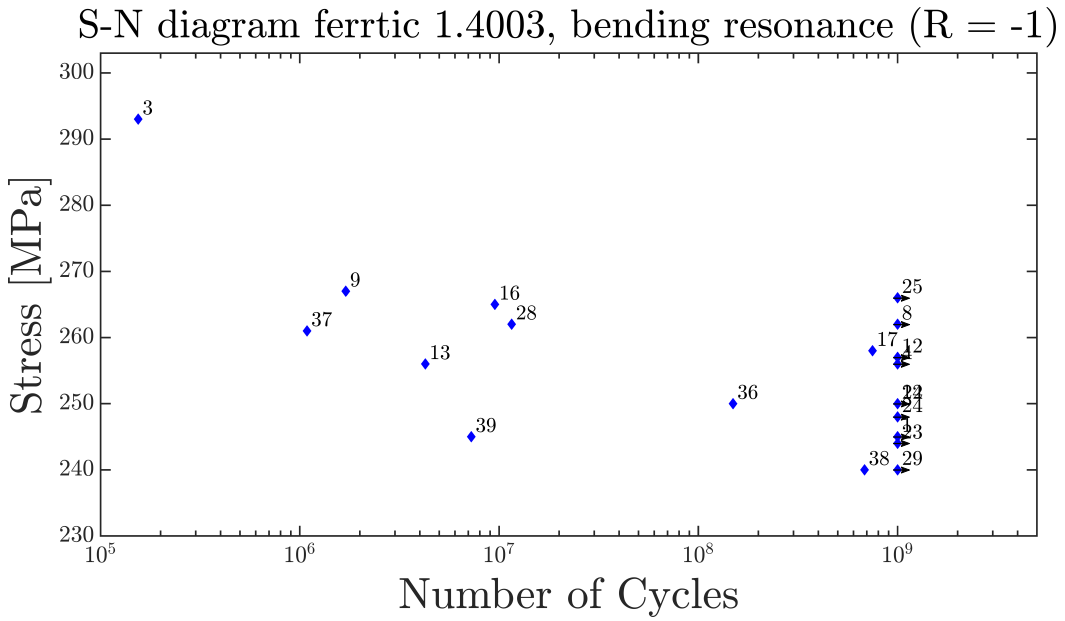

Figure 5.6: The cyclically applied von Mises stress amplitude plotted over the number of cycles until failure. Plotted stress amplitudes are extracted from the aforementioned simulation methodology. All tests were performed under high or very high cycle fatigue loading. Arrows indicate tests that were run-outs.

Most specimens investigated in the course of this work did not match the failure criteria (i.e., a relative frequency drop of 1E-3) before reaching the stopping criterion of 10^9 cycles and are therefore considered run-outs. Only cycles at loads above 97% of the target load were considered to contribute to the total cycle count. The applied loads culminated in tests located

around the fatigue limit where mostly microscopic plasticity in individual grains occurs and the irreversibility threshold is barely exceeded. As observed here, this fatigue regime is characterized by a substantial life scatter. In the bending resonant loading, fatigue failure is initiated always at the surface. The initial resonant and testing frequency ranged from 1910–2250 Hz for all tests owing to the scatter in specimen geometry. Within this frequency span, no frequency effect-induced change of dislocation glide mechanisms is anticipated. The specimen thickness variance mentioned in section 5.1 led to differences in bending stress gradients. During the adjustment phase of the experiment, a typical overshoot of less than one percent was observed.

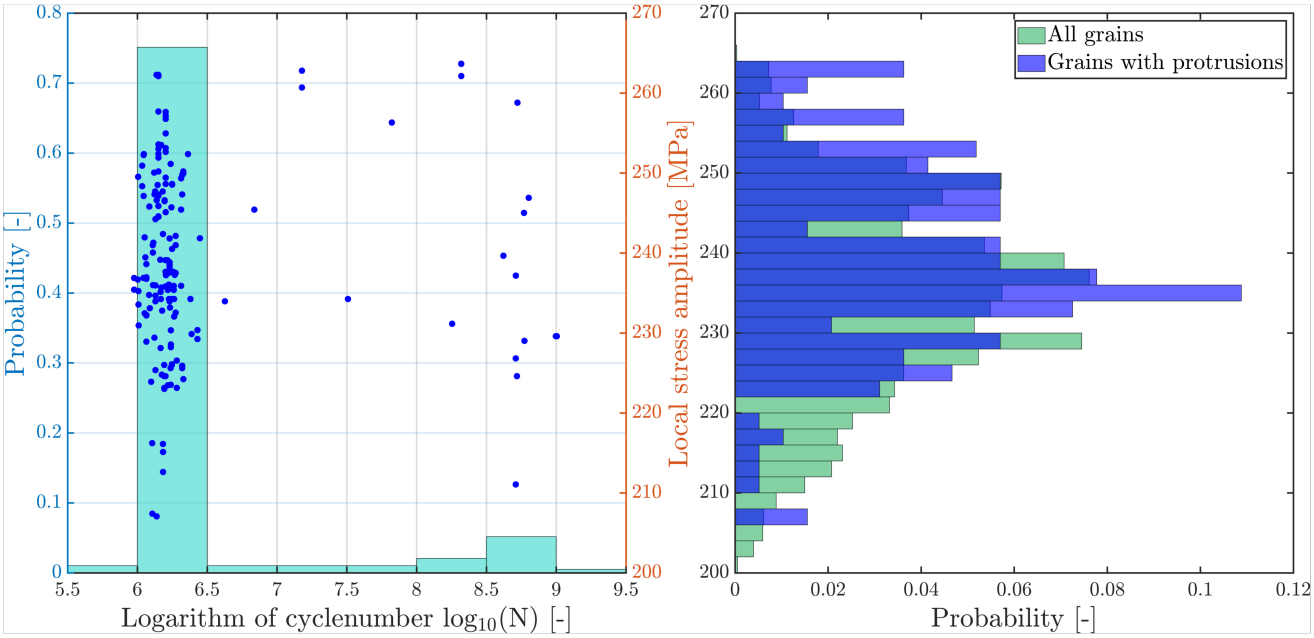

Figure 5.7: Protrusion formation life N_{50} as a histogram and plotted together with the local, isotropic-calculated von Mises stress from FE simulation. The probability distributions on the right compare which stresses arise in arbitrary grains and in the ones containing protrusions.

In figure 5.7, a damage initiation plot for individual protrusions is provided. These lifetimes originate from four specimen sides for which image series are acquired entailing 193 protrusions. They are inferred from the computer vision-based approach outlined in section 3.7 and confirmed by manual inspection. Specifically, the function in equation 3.1 is fitted to the accumulated intensity evolution of each protrusion instance. It can be seen that most primary protrusions growth occurs in the range $1{\cdot}10^6$ – $2.5{\cdot}10^6$ cycles. The stationary protrusion growth was found to be abrupt and stagnated shortly after the given cycles. Growth rates κ in the range of $1{\cdot}10^{-6}$ – $1{\cdot}10^{-4}$ were found to be typical. This is in line with observations made elsewhere [198, 222]. At later cycles, a few secondary protrusions form. Moreover, fluctuations in the final intensity are observed. These fluctuations are attributed to noise in the area as can be observed in figure 3.12 and shadowing due to topography alterations. Aside from cyclic damage formation, the figure also contains an assessment in terms of local stress amplitudes taken from the superimposed FE von Mises stress map. The right image shows probability distributions for the stress in arbitrary grains (turquoise) and damaged grains (blue). From this comparison, it is evident that the occurrence of damage instances is to some degree spatially correlated with grain von Mises stress. Note that this study does not capture the influence of elastic or plastic incompatibility on stress modulation.

Another analysis performed was evaluating the influence of Schmid factors on the emergence of protrusions. This analysis, rather than stress magnitude, captures the influence of crystallographic orientation with respect to the specimen axis. Multiple inverse pole figures (IPF) of specimen axis orientation distributions were computed, see figure 5.8, following [17]. The black dots in subfigures 5.8a and 5.8b describe the specimen axis orientation in the crystal coordinate systems of EBSD pixels for which damage was observed. Maximum Schmid factors for a range of grain orientations can be computed and illustrated as an isoline contour plot within the IPF (dashed lines), see figures 5.8a and 5.8b. During maximum Schmid factor computation in figure 5.8a, only $\{1\,1\,0\}\langle 1\,1\,1\rangle$ slip systems were accounted for, while isolines in Figure 5.8b included $\{1\,1\,2\}$ and $\{1\,2\,3\}$ bcc

slip systems as well. In this case, neither a distinction between slip systems of different slip plane families nor asymmetries within specific slip planes is made. In figure 5.8b, additionally, an orientation distribution function (ODF) was estimated from the loading axis distribution at damage containing pixels applying a de la Vallée Poussin kernel [287] with a halfwidth of 2.9°. In contrast, figure 5.8c shows the ODF (halfwidth = 2.0°) for all pixels contained in the EBSD scan as a reference. Different kernel halfwidths were chosen to account for the distinct amount of data points in both scenarios. The ODFs are measured in multiples of random.

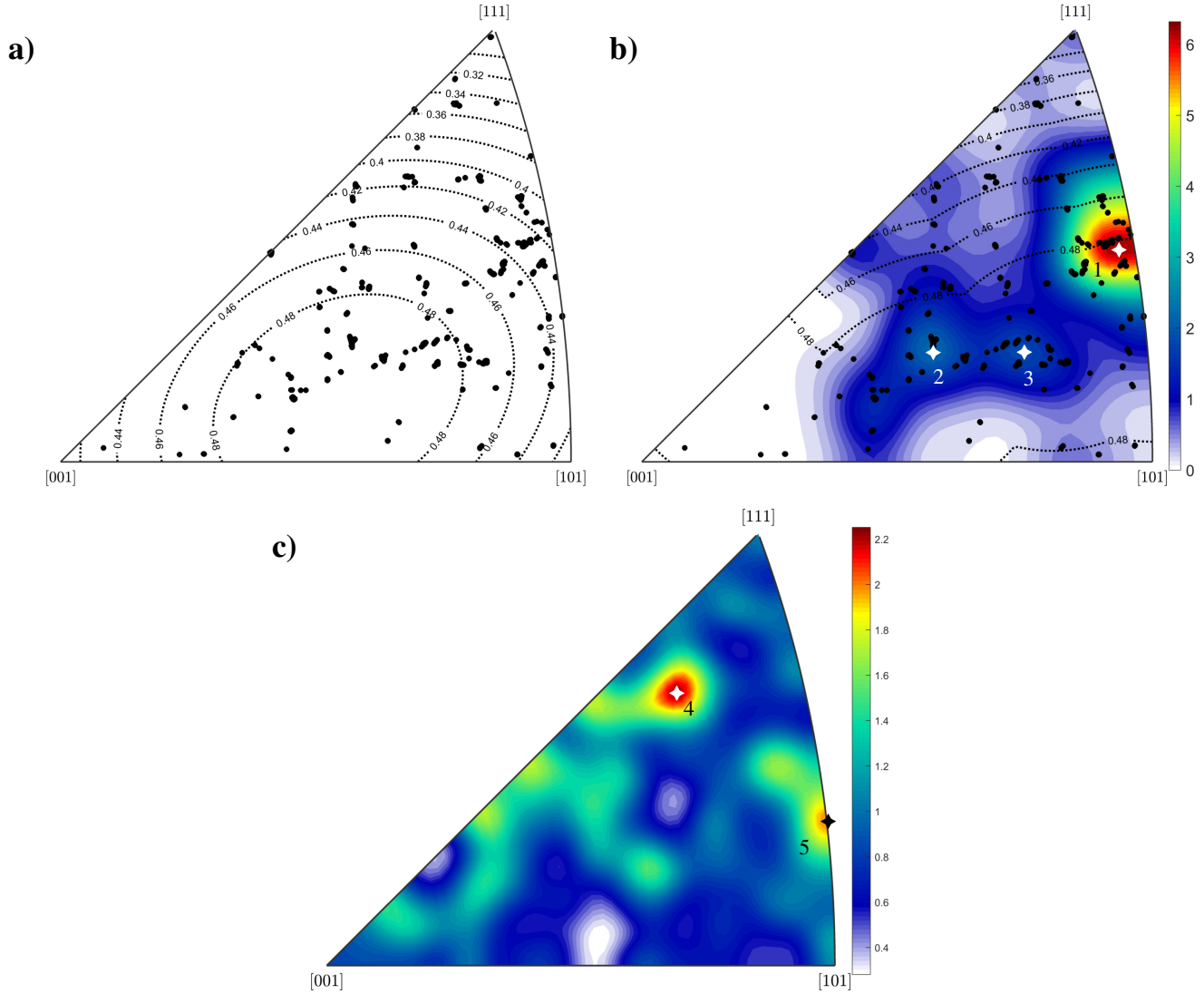

Figure 5.8: Inverse pole figures showing specimen axis $[1\,0\,0]$ distributions. **a)** loading axis orientations at EBSD pixels which contain damage and isolines indicating the maximum Schmid factor considering bcc $\{1\,1\,0\}\langle 1\,1\,1\rangle$ slip systems. **b)** shows the same as **a)** except that max. Schmid factor isolines were computed considering all three slip plane families. Additionally, an ODF is superimposed on **b)** for easier comparison with **c)**, which illustrates the ODF of the specimen axis for all EBSD pixels. Some considerable resulting peaks in the ODFs are annotated with star symbols. Values on the colorbar indicate multiples of random. Adapted from [36].

From figure 5.8a, it becomes apparent that indeed a large part of damage containing EBSD pixels exhibited orientations resulting in Schmid factors exceeding 0.48. This applies even more, when all three potential slip plane families in bcc are considered (figure 5.8b). The loading axis ODF considering only crystal orientations at damage locations in figure 5.8b shows three peaks, one pronounced peak (1) with MOR of approximately 6 and two minor peaks (2, 3) with MOR in the range 3–4. These peaks are situated in close proximity to the centers of the isolines of either slip system family associated with the highest Schmid factors. Similar to the correlation with FE stress, the Schmid factor can be considered a contributing factor but not a sufficient condition for protrusion emergence. Furthermore, protrusions were found to be clustered at high-angle

GBs. In a few cases, such slip markings were found to traverse through grain boundaries. Due to the apparent variety of independent variables and anticipated interactions between them, machine learning was applied to attempt to explore the different driving forces and their relative importance on protrusion formation which will be presented later.

In this material, protrusions are not the only precursors for high-cycle fatigue crack initiation. While some cracks originate at such slip markings, also preferentially-oriented grain boundaries and to a lesser extent pores and MnS inclusions act as sites for crack initiation. Due to the scarcity of crack instances across the observed specimens, a quantitative assessment and statistical analysis of the dominating crack initiation mechanism proves difficult. Hence, rather than aiming for a data-driven approach to infer driving forces for short crack growth, in the following, individual crack instances are considered analytically.

Therefore, the region annotated with a dashed box in figure 5.5, which contains a microstructurally short crack, is superimposed with the in-situ image series, see figure 5.9a–l. Moreover, the states before and after fatigue acquired with SEM in figure 5.9n and 5.9o, respectively, pose references. Figure 5.9m shows the DL-derived damage map downsampled to the resolution of EBSD and therefore appears coarser than figure 5.9o.

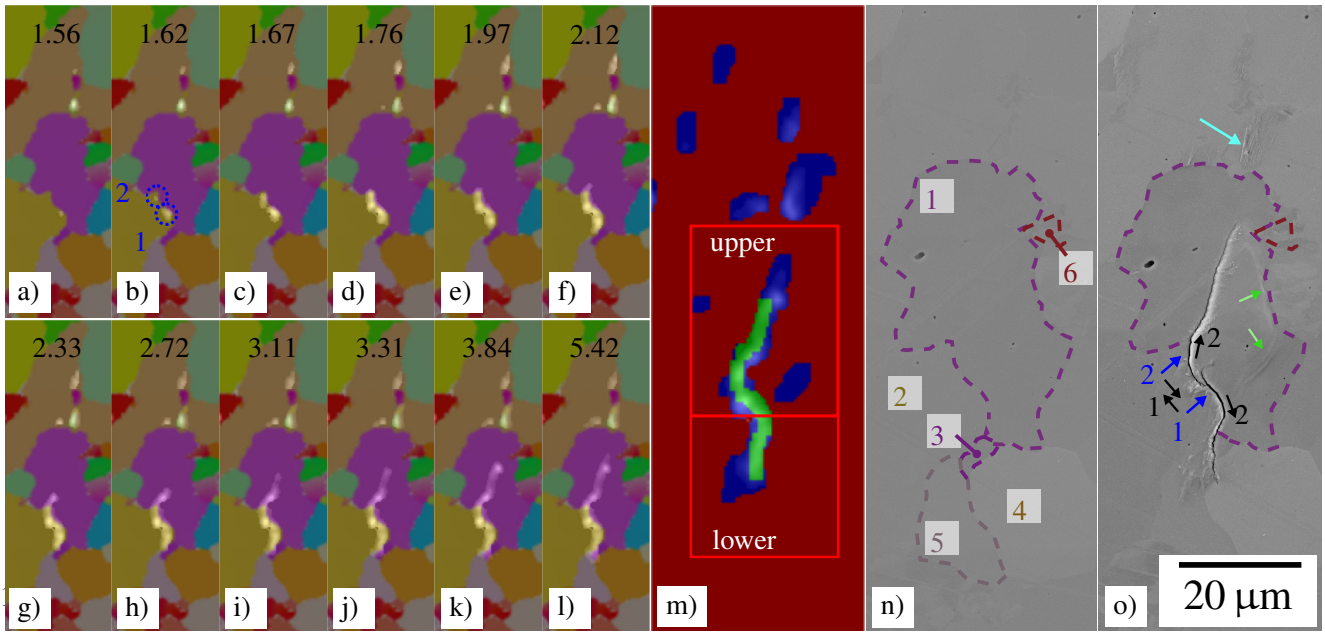

Figure 5.9: Subfigures 5.9a–l represent an image series of intensity evolution ascribed to topography changes from slip marking and crack formation and growth superimposed onto an inverse pole figure map. The numerical value at the top of each image represents the corresponding cycle number at which the image was captured in [$\cdot 10^6$]. In figure 5.9m, a multi-class segmentation map with cracks (green) and extrusions (blue) is shown, which was downsampled to the resolution of the EBSD. These are complemented with two SEM images captured prior to (figure 5.9n) and after fatigue (figure 5.9o). The SEM image prior to fatigue (subfigure 5.9n) contains grain numbers and the image after an illustration of the crack growth stages, where blue arrows indicate crack initiation locations and black arrows indicate crack growth. Moreover, green arrows point out a loop-shaped slip marking (better visible in the digital version due to faint contrast) The turquoise arrow in figure 5.9o annotates a surrounding protrusion. Adapted from [36].

From the post-mortem SEM in figure 5.9o, it is observable that two separate crack branches are present. This can not be resolved in the later stages of the in-situ image series (figures 5.9e–l). However, the two distinct elevated intensity regions emerging in figure 5.9b indicate that two distinct intergranular cracks originated in close vicinity (approximately 5-10 µm gap). From the image series, it can be deduced that upon crack formation, both crack instances exhibit bidirectional crack growth. Subsequently, the cracks converge and, based on the in-situ image series, seemingly merge. However, from figure 5.9o, it is uncertain whether the cracks merged or growth concludes with a stress-relaxed state that prevents the two cracks from fully merging. In contrast to this, the other ends of the cracks proceed to grow and exhibit arrest at microstructural defects. The upper crack part transitions from an intergranular crack to a transgranular crack and subsequently undergoes growth direction changes (see figure 5.9o). In subfigure 5.9o, damage which occured early during fatigue (see subfigure 5.9a) is annotated with an turquoise arrow.

In order to analyze the crack growth rates and barrier mechanisms, the crack branches were considered individually, as illustrated in figure 5.9m. The assumption is made that the elevated intensity solely originates from the crack growth, even though there are intensity contributions from the crack and the surrounding crack-induced surface plasticity. The crack growth of the upper and lower crack branch is illustrated in figure 5.10a and 5.10b, respectively.

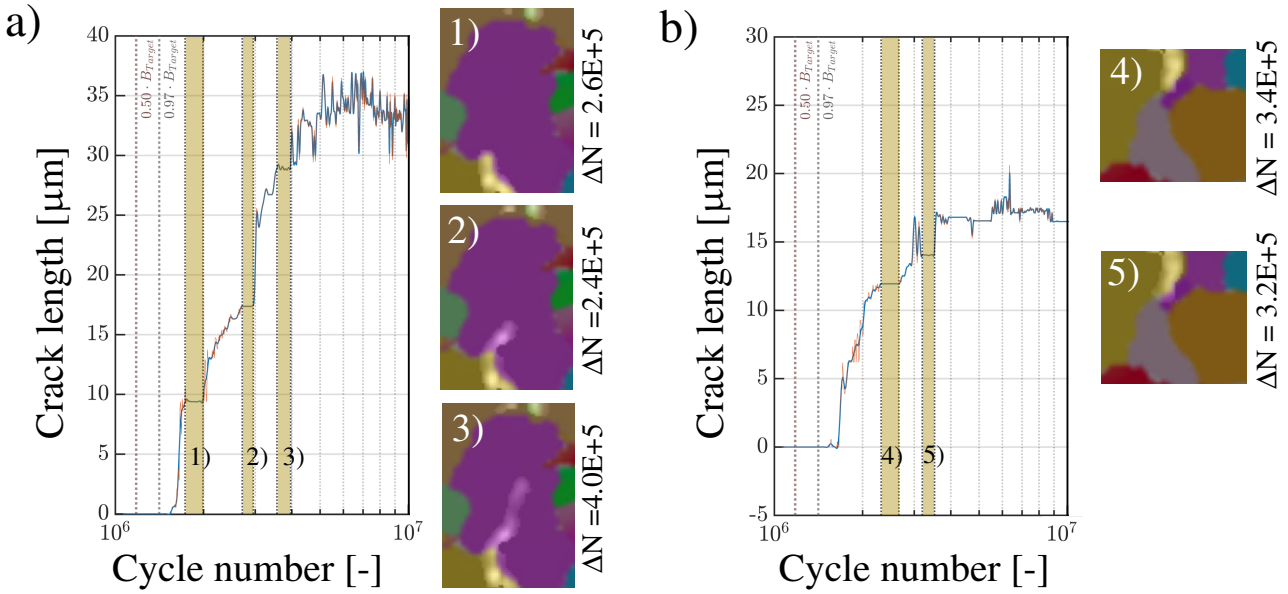

Figure 5.10: Cyclic crack length evolution demonstrating five instances of crack growth retardation. The insets 1–5 show the states at which the growth was retarded due to different phenomena. A pair of dotted vertical lines indicate the run-up of the fatigue amplitude. The vertically aligned ΔN values next to the insets quantify by how many cycles the crack growth was obstructed. Adopted from [36].

In the following, we describe the results by referring to specific grains and grain boundaries by means of the designations introduced in figure 5.9n. Crack initiation and growth occurred shortly after reaching the target amplitude (vertical dotted line in figures 5.10a and 5.10b). Both figures show distinct plateaus in the course of microstructural crack growth. At every plateau, the growth is obstructed for a few 10^5 cycles. While the plateau #1 in figure 5.10a occurs at a surface residing grain boundary when transitioning from intercrystalline to transcrystalline crack growth, the crack resting at the other plateaus does not show any indication of superficial defect interaction. However, in the plateau #2 in figure 5.10a, reorientation of the crack occurs, resulting in subsequent crack growth nearly orthogonal to the horizontally oriented specimen axis. In this section the growth rate reaches its upper limit of $da/dN = 1.5 \cdot 10^{-4}$ µm/cycle. A comparison of the upper crack path in the transgranular section (in G1) with potential slip planes indicated that some of the collinear crack segments with distinct directions approximately coincide with the potential $\{1\,1\,0\}[1\,1\,1]$ trace orientations. The slip plane traces of the $\{1\,1\,0\}$, $\{1\,1\,2\}$, and $\{1\,2\,3\}$ slip plane families are provided in the appendix. The crack observed in the lower crack part is retarded when the intercrystalline crack at the G1-G2 boundary meets the G1-G2-G3 triple point and transitions into a transcrystalline crack in G3 (figure 5.10b #4) and when the crack transitions trough the G3-G4 border (figure 5.10b #5). At triple points, typically a pronounced mismatch of mechanical properties is present which can, depending on the detailed configuration and presence of special GBs, promote or inhibit crack growth [288].

In another case, which posed the exception, crack initiation occurred at a pore situated at a low disorientation GB (25°). This damage location is depicted in figure 5.11. Upon early crack nucleation after $1.3 \cdot 10^6$ cycles. Crack growth occurs on slip planes in both adjacent grains consecutively, as indicated by the arrows in figure 5.11b. Initially, the upper grain boundary with high disorientation (54°) poses a barrier to crack propagation. This induces crack growth in the bottom grain where the crack exhibits a variable path alternating between crystallographic planes such that it overall grows approximately perpendicular to the specimen axis. Ultimately, the crack is stopped at another high disorientation grain boundary (46°). Whenever damage arose due to pore-induced stress concentration, it happened at very early stages of fatigue, sometimes even before reaching the macroscopic target stress, i.e. during the experimental run-in.

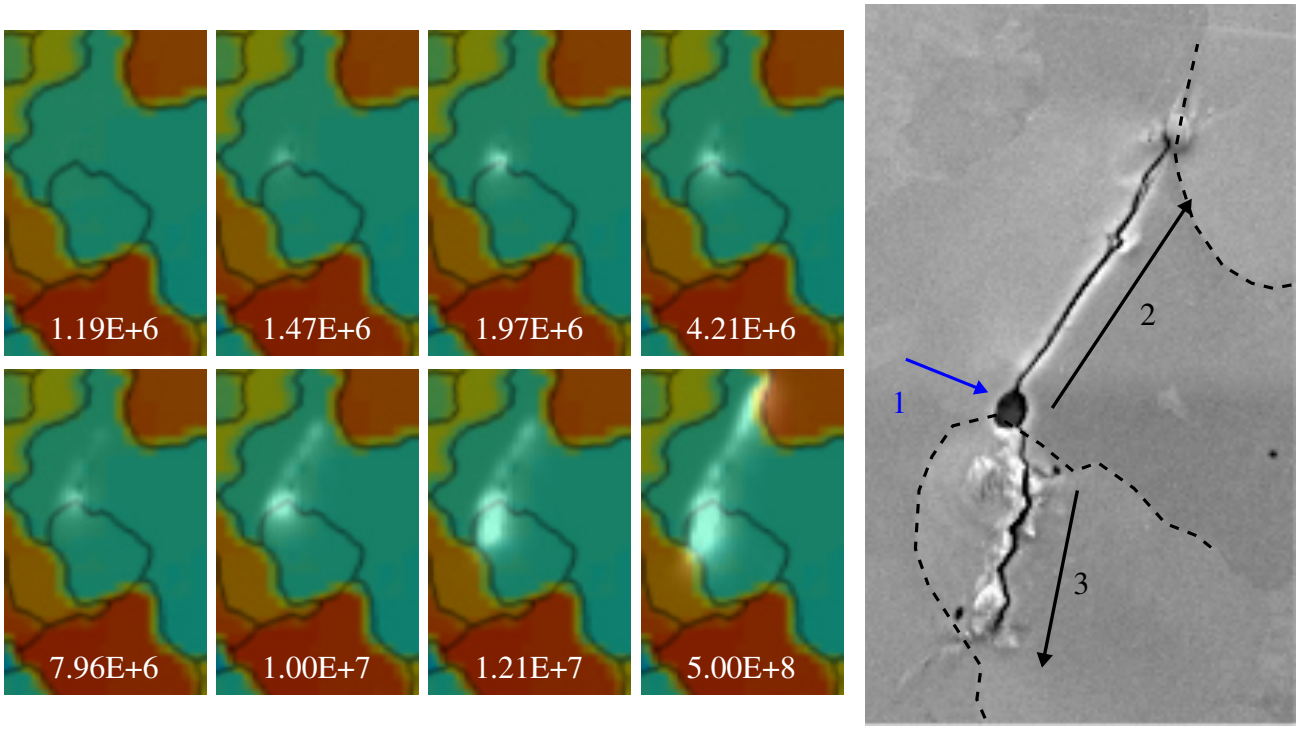

Figure 5.11: a) Light optical grayscale image series to track crack formation and evolution. The black lines indicate grain boundaries and the numbers the corresponding load cycle numbers. **b)** Secondary electron SEM image of the final crack state. The numbering and arrow annotations illustrate the crack initiation (blue) and growth process (black). Grain boundaries analogous to **a)** are faintly visible through discontinuities in the grayscale contrast (better visible in the digital version) and therefore highlighted with dashed lines. Adapted from [37].

5.5 Segmentation of fatigue damage locations

In this section, the results of the three experiments (1–3) outlined in the experimental section 4.5 are presented covering the semantic segmentation of fatigue-induced extrusions and cracks. Initially, U-Net models trained only with augmented ferrite data sets were evaluated on ferrite (source) to demonstrate the feasibility of DL-based damage detection (1) in section 5.5.1. Subsequently, the material domain generalization of the same models with respect to the target domains martensitic steel and copper was investigated (2) in section 5.5.2. The objective was to research whether the models learn a general representation of surface slip markings, or if more elaborate measures are required to ensure applicability to a multitude of materials. Extending on this, models were trained in a multi-domain setting on both steel data sets and tested on all data sets (3) in section 5.5.3. The large variance of the three data sets and their different sources are described in section 3.3. To account for the variance, all sections comprise the investigation of data augmentation's influence on segmentation performance. Thereby the importance of data augmentation for data sets that contain significant variance is assessed. Moreover, the question of whether standard image transformations can improve out-of-domain generalization to alternate materials by rendering the model invariant with respect to specific changes in input image texture is addressed by this. In order to account for the imbalance between the data sets, imbalance correction strategies were applied. The data augmentation hyperparameter studies were performed for downscaled images to keep the training time maintainable. The results presented here are to a large extent adopted from [37].

5.5.1 Source domain model evaluation

The source domain model serves the purpose of investigating the feasibility of DL-based semantic damage segmentation. Furthermore, the impact of individual augmentations on the segmentation performance can be considered a reference for

the following study of out-of-domain generalization. In accordance with the approach described in section 4.5, multiple U-Net models were trained from random initialization using the ferritic steel training set. Different individual augmentations and augmentation configurations described in table 5.1 were tested. The thresholds and probabilities characteristic for the individual augmentations can be found in table 4.3.

Table 5.1: Influence of augmentation pipeline on source domain performance. Note that in all tests, even in test #8, standard augmentations marked with blue background color in table 4.3 are applied. The subscripts 'c', 'e', and 'o', refer to crack, extrusion, and overall, respectively. Adopted from [37].

Test #	Resolution	Brightness	Contrast	Gaussian blurring	Motion blurring	Gaussian noise	Elastic transformation	Optical distortion	ferrite IoU_c	ferrite IoU_e	ferrite IoU_o
1		✓	✗	✗	✗	✗	✗	✗	0.75	0.49	0.74
2		✗	✓	✗	✗	✗	✗	✗	0.77	0.55	0.77
3		✗	✗	✓	✗	✗	✗	✗	0.74	0.52	0.75
4		✗	✗	✗	✓	✗	✗	✗	0.76	0.51	0.75
5	N · 0.25	✗	✗	✗	✗	✓	✗	✗	0.70	0.51	0.73
6		✗	✗	✗	✗	✗	✓	✗	0.76	0.52	0.76
7		✗	✗	✗	✗	✗	✗	✓	0.76	0.50	0.75
8		✗	✗	✗	✗	✗	✗	✗	0.75	0.52	0.75
9		✓	✓	✓	✓	✓	✓	✓	**0.82**	0.60	0.80
10		✗	✓	✓	✓	✗	✓	✓	0.81	**0.66**	**0.82**
11	N	✗	✓	✓	✓	✗	✓	✓	0.84	0.71	0.85

Results are evaluated utilizing the intersection over union metric, see equation 2.24, of the classes extrusion (IoU_e) and crack (IoU_c) in ferritic steel and later of extrusions in martensite and copper. Note that the background IoU is neglected as it is approximately unity in every case due to class imbalance. Nevertheless, the overall IoU (IoU_o) for ferritic steel is given. IoU proves to be a suitable metric since, under the loading conditions delineated in section 4.4, damage arises only at a few locations of the microstructure, which leads to a class distribution skewed towards the undamaged matrix (background).

Since the augmentation probabilities according to table 4.3 differ, it is difficult to draw conclusions between individual augmentation types. However, comparisons against test #8 (baseline) are feasible which is a baseline augmentation case containing only linear (affine) transformations such as shift, scale, rotations, and mirroring. The segmentation performance of cracks is higher than that of extrusions by 0.10–0.26 points depending on the model. The extrusion class profits more from maintaining the original resolution (cf. tests #10 and #11). It can be observed that brightness (#1) and Gaussian noise (#5) augmentations cause detrimental effects on the IoU_e and IoU_c, respectively when compared to the baseline experiment with minimal augmentations (#8). Taking into account only the models trained with down-sampled data sets, an increase of ΔIoU_c of 0.06 and ΔIoU_e of 0.14 for the evaluation on ferritic steel was achieved employing a custom set of augmentations (comparing #10 with #8). This custom set included only augmentations that optimized performance on the ferritic and martensitic steel domains.

In the following, the results of the ferritic steel trained model performing best on the ferritic steel test data set (table 5.1 #11) are examined in greater detail. Figure 5.12 illustrates a case study of good and poor segmentation instances in a–f and g–l, respectively.

Cases a–e in figure 5.12 show different types of protrusions that are segmented correctly by the network. Considerable changes in texture, brightness, and contrast can be observed in these protrusion images. Cases 5.12a,b show shallow extrusions while

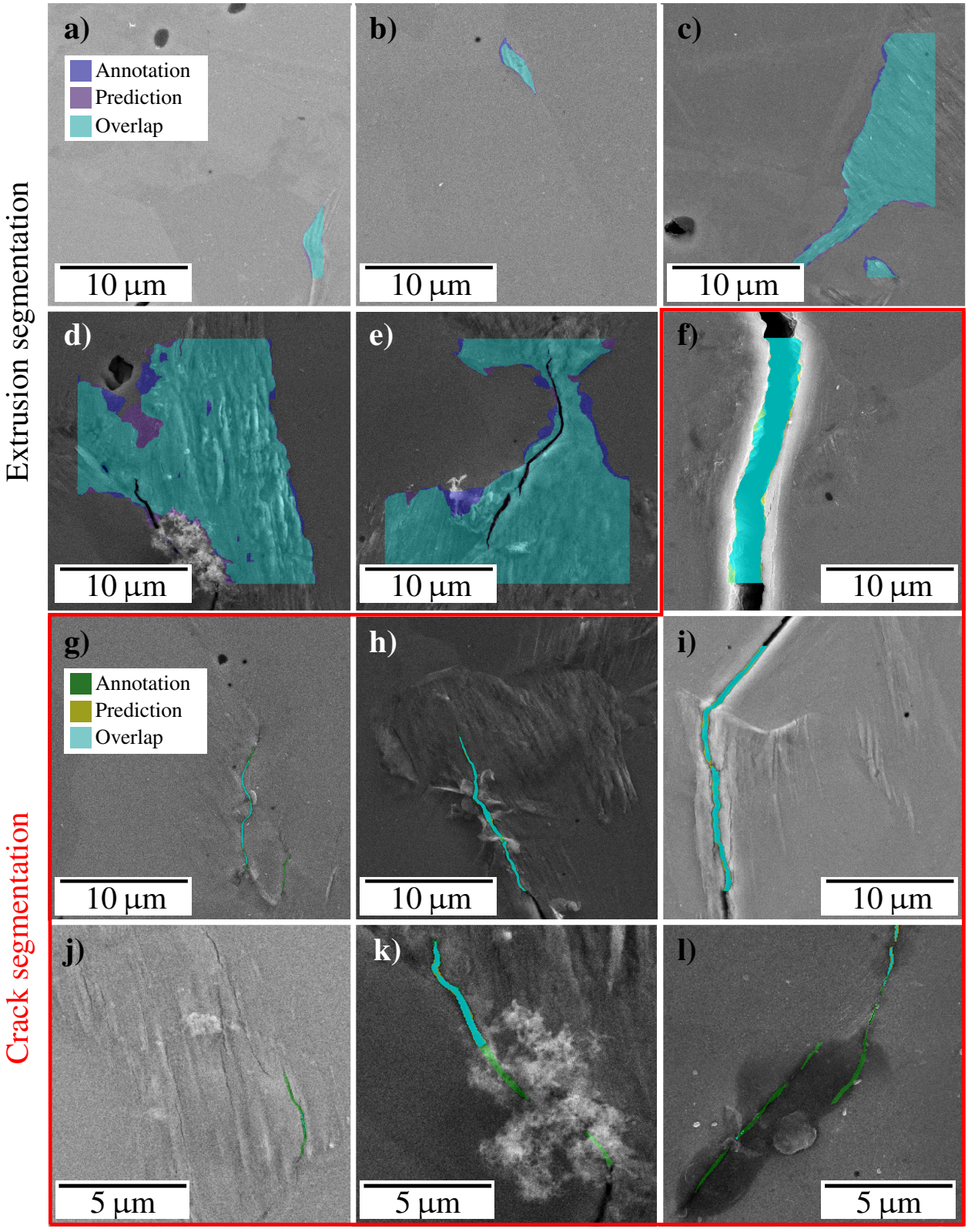

Figure 5.12: Examples for common cases of extrusion and crack segmentations. Subfigures **a–e** and **f–l** represent extrusion and crack segmentation cases, respectively. Note that the image border regions do neither show annotations nor predictions since the latter is not provided by the network in these regions. Images **j–l)** are an exception to this as they show a subset of pixels. Adopted from [37].

cases c–e show extensive ones. Example 5.12c shows that the presence of multiple slip trace orientations does not impede the segmentation of extrusions. In cases 5.12d and e, the network distinguishes the extrusion area from the cracks and crack debris very well.

Similarly, the images of cracks seen in figure 5.12f–i show that the network also learned to segment different types of cracks under various imaging conditions. Case 5.12f shows a large crack while case 5.12g on the other hand shows a microstructurally short crack that has been segmented accurately. In case 5.12h, a crack is detected correctly despite regions occluded by crack debris. A crack that is differentiated from the accompanying protrusion can be seen in case figure 5.12i. Cases j–l represent instances where crack segmentation was not accurate. In figure 5.12j, a fissured protruded area is depicted. Image 5.12j contains instances of arguably tiny microstructurally short cracks, which are on the verge of still being considered intrusions (and hence associated with the extrusion class). Fatigue crack tip friction-induced debris covering large parts of the crack is shown in figure 5.12k. Figure 5.12l shows a potential inclusion with a darker background concealing cracks and hence impeding its detection.

5.5.2 Domain generalization

This section addresses the capability of ferritic steel damage-trained models to generalize to other damage domains. From comparison of tables 5.1 and 5.2 it becomes apparent that the augmentations delineated in table 4.3 impact the performance of the network in distinct ways for the three test data sets. A substantially larger IoU scatter is observed in the target domains, where some augmentations prove particularly detrimental.

Table 5.2: Influence of augmentation pipeline on domain generalization performance. Note that in all tests, even in test #8, standard augmentations marked with blue background color in table 4.3 are applied. The subscript 'e' refers to extrusion. Adopted from [37].

Test #	Resolution	Brightness	Contrast	Gaussian blurring	Motion blurring	Gaussian noise	Elastic transformation	Optical distortion	copper IoU_e	martensite IoU_e
1		✓	✗	✗	✗	✗	✗	✗	0.55	0.05
2		✗	✓	✗	✗	✗	✗	✗	0.37	0.06
3		✗	✗	✓	✗	✗	✗	✗	**0.68**	0.07
4		✗	✗	✗	✓	✗	✗	✗	0.66	0.05
5	N · 0.25	✗	✗	✗	✗	✓	✗	✗	0.67	0.13
6		✗	✗	✗	✗	✗	✓	✗	0.67	**0.17**
7		✗	✗	✗	✗	✗	✗	✓	0.67	0.09
8		✗	✗	✗	✗	✗	✗	✗	0.51	0.06
9		✓	✓	✓	✓	✓	✓	✓	0.33	0.14
10		✗	✓	✓	✓	✗	✓	✓	0.26	0.14
11	N	✗	✓	✓	✓	✗	✓	✓	0.61	0.15

Domain generalization to copper: At first glance, it can be inferred that the models trained on ferritic steel can generalize seamlessly to the copper domain as IoU_e values for copper exceeds the source domain IoU_e for some models. However, upon detailed analysis, it becomes apparent that the tongue-like extrusions, which constitute a small damage area proportion in copper, are segmented erroneously (see figure 5.13a). The connectivity of multiple individual tongue-like extrusions in close vicinity is predicted as merged connected ones as highlighted by the arrows. Moreover, it is noteworthy that every model involving contrast augmentation showed comparatively poor performance for copper among the individual augmentations.

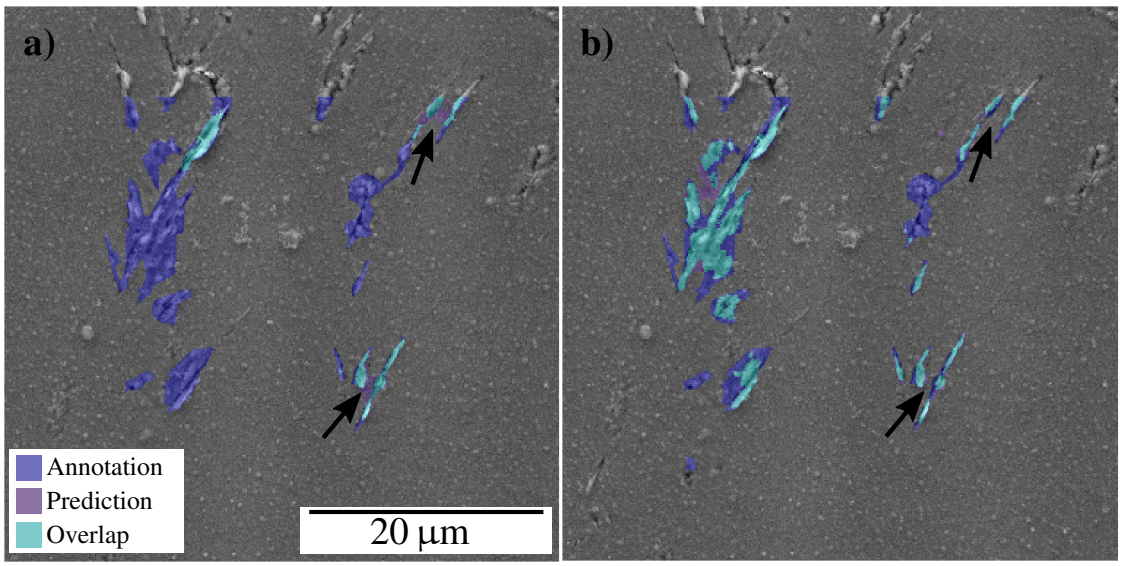

Figure 5.13: Tongue-like extrusions in fatigued copper segmented by a U-Net model trained with **a)** the ferritic steel data set and **b)** a conjunct data set of ferritic and martensitic steel.

Domain generalization to martensite: In contrast, domain generalization to martensite was unsatisfactory. Despite certain augmentations leading to an increase of 0.08 on martensite, the final absolute IoU_e of 0.14 is quite poor. Also, in contrast to the observations in the ferritic steel, the martensitic steel IoU_e increases when Gaussian noise is applied. In particular, the elastic transformation seems to improve segmentation performance.

General remarks: Apart from a few exceptions, augmentations improve the average performance across all considered domains relative to the minimal augmented experiments (#8). Most notably elastic distortion (#6) and to some extent optical distortion (#7) boost the overall IoU in every domain. On the contrary, brightness and contrast augmentation should be treated with care as these seem to potentially cause severe performance drops.

5.5.3 Multi-domain training

The results from the prior section indicated the demand for additional training in order to be able to generalize over multiple materials domains and microstructures with substantial domain gaps. At the same time, results showcased potential domain generalization of models for smaller domain gaps. Hence, training a model with few diverse materials (i.e. damage types) representative for damage types in various materials was pursued over alternative generalization techniques. In our specific case, indications were found suggesting that domain generalization of solely ferritic steel-trained models to martensite and copper is limited due to the ferritic steel material at hand exclusively containing protrusions with superimposed extrusions and intrusions. Therefore, further training with conjunct data sets adding martensite with tongue-like extrusion morphology, and evaluation on all test data sets were performed.

In table 5.3, the training settings and corresponding results are described employing the optimized augmentation setting from model #10 in table 5.1. Due to the imbalances between ferritic steel and martensitic steel data sets, see table 3.1, a few data set imbalance correction (DSIC) strategies were attempted in tests #3 to #5. Since many scientific problems and thus data sets are either intrinsically or extrinsically imbalanced, there is extensive literature on the imbalance correction, e.g. [163]. In this work initially, a sample weight correction (SWC) was tested in which correction factors dependent on the imbalance between training data sets are incorporated into loss calculation. These factors (weights) for ferritic and martensitic steel were computed based on the extrusion pixel ratio between the ferrite and martensite training data set. An alternate route was tested with the sampling correction (SC), where the number of images sampled from both training data sets was matched by over- and undersampling the martensite and ferrite data sets, respectively. These imbalance correction techniques were

complemented with a third hybrid approach which combined SWC with over-sampling of the martensite data set and in the context of this work is referred to as sample weight and sampling correction (SWSC).

Table 5.3: Results of different training and testing data sets including data set imbalance correction schemes. The letters 'm' and 'f' in the training set column denote martensitic and ferritic steel. The subscripts 'c', 'e', and 'o', refer to crack, extrusion, and overall, respectively. Adopted from [37].

Tests #	Resolution	Training set	DSIC type	ferrite IoU_c	ferrite IoU_e	ferrite IoU_o	martensite IoU_e	copper IoU_e
1	N · 0.25	m	-	-	-	-	**0.43**	-
2		m+f	-	**0.82**	**0.68**	**0.83**	0.31	0.38
3		m+f	SWC	**0.82**	0.62	0.81	**0.43**	0.32
4		m+f	SC	0.80	0.64	0.81	0.39	0.27
5		m+f	SWSC	0.79	0.62	0.80	**0.43**	0.31
6	N	m+f	SWSC	0.83	0.67	0.83	0.47	**0.58**

Performance on martensite: Initially, a target domain (martensite) trained experiment (#1) was evaluated on the martensite test data and achieved an IoU_e of 0.43. Typically, target domain-trained performances are considered to pose an upper bound in performance [29]. Indeed, this reference value exceeds the value observed from the domain generalization of the ferritic steel trained model (table 5.1 experiment #10) by a large margin. While the combined training (#2) is improving IoU_e on martensite compared to values achieved by domain generalization, it does not quite reach the reference value. Imbalance correction alleviates this discrepancy. Significant improvements of 0.12, 0.08 and 0.12 on the martensite IoU_e are observed when SWC (#3), SC (#4) and SWSC (#5) is applied, respectively. When trained on the original resolution, the SWSC yielded a martensite IoU_e of 0.47. Overall the martensite damage segmentation performance is worse compared to that of other materials.

Performance on ferritic steel: From the experiments, it can also be seen that including the martensitic data set for training does not cause significant performance reductions on the ferritic steel data set. In experiment #2, the slight decline in IoU_c when compared to table 5.1 experiment #10, is compensated by a small improvement in IoU_e. When additionally imbalance corrections in favor of martensite are applied marginal reductions in ferritic steel performance are noticeable.

Performance on copper: The model trained conjunctly on ferritic and martensite steel without data imbalance correction (#2), when tested on copper, showed an improvement over model #10 from table 5.1. It can be observed that this conjunct model is capable of assessing the connectivity of close proximity tongue-like extrusions significantly better than solely ferritic steel-trained models (see figure 5.13b). The issue of merging tongue-like extrusions is alleviated by adding strongly localized martensite extrusions to the training. However, the balancing methods decreased the copper performance slightly.

With respect to #11 in table 5.1, an overall IoU reduction of 0.02 can be observed for the ferritic steel data set, while the martensitic steel IoU_e boosts by 0.32. In general, it can be inferred that when additionally martensite images were provided to the network during training, a better generalization across extrusion types was achieved. After imbalance correction, the network was able to attain segmentation performances for extrusions in ferritic and martensite steel, which matched the performances of either network dedicated to the respective materials domain.

5.6 Random forest-based grain-wise binary damage classification

After successfully segmenting damage instances in an automated fashion, the damage information was mapped onto the microstructures. This facilitated training a machine learning classifier to predict whether grains will contain slip markings or

not from the initial state. Owing to minor residual misalignment after registration and the tendency of damage instances to be situated immediately at grain boundaries, the damage instances' assignment to individual grains was corrected by visually inspecting the SEM image after fatigue. Whenever a part of a slip marking extended into a grain, it was considered damaged. Following this approach, the degree of surface deformation differed significantly between damaged grains, as can be seen in figure 5.12a–e.

5.6.1 Feature engineering

In order to predict the emergence of protrusions within grains by a machine learning classifier, a set of descriptive features needs to be computed. Since environmental and surface topography influence factors were largely suppressed in this study, and only one material was investigated in the context of fatigue, the focus was placed on capturing microstructural descriptors and mechanical loading comprehensively. This entails microstructure morphology, crystallography (microtexture), and pore defect attributes as well as hybrid features that couple mechanical loading with the former. A list of features is provided in table 5.4. Justification for this choice of features was provided in section 2.1.3. For instance, the grain size is presumed an important feature as, in absence of precipitates, it affects the mean free path of dislocations and the intensity of dislocation pile-ups, and therefore the local strengthening behavior. Beyond table 5.4, further features are considered which are slightly altered permutations and variants of the ones in the table. In total, 120 grain-level descriptors were considered to capture various aspects of the microstructure, pore defect distribution, and loading. However, some of these engineered features are highly correlated. In particular, this applies to the features addressing grain morphology. By considering not only individual grains but also differences with respect to adjacent grains some contextual information is fed in during model construction. The features which supposedly infuse information of neighbor grains into the classifier are marked with a green check mark in table 5.4.

The set of features is extracted using an automated Matlab routine including different functionalities of the image processing and computer vision toolbox in combination with MTEX, a third-party toolbox providing a variety of crystallographic routines.

5.6.2 Evaluation of slip marking prediction classifier

This section presents the training and evaluation of a classifier trained to discern grains that will contain fatigue slip markings from grains resilient to surface damage accumulation. For this purpose, the initial state of the tested ferritic steel specimens and the applied loading captured in the registered data is transcribed into the aforementioned feature set. All features were imputed, i.e. missing values were filled by applying the mean value. Subsequently, all features were standardized to remove the mean and scale to a variance of unity.

The complete data set derived from the multimodal data contains a total number of 7633 grains when discarding individual microtwins as indicated in figure 5.5. This data set is generated from both sides of four specimens. The four specimens were exposed to a distinct cyclic load corresponding to von Mises stress amplitudes ranging from 240 to 262 MPa. Owing to this HCF loading, only 311 grains among these contained surface slip markings. All kinds of protrusions are taken into account, irrespective of their time of emergence. This includes also protrusions that emerged at pores. Microstructurally short cracks and surrounding crack-induced plasticity are not considered in this task. Hence, the data set comprises a rather pronounced inherent imbalance where only 4.25% of the grains exhibit fatigue slip markings. An underlying assumption is that in the HCF regime and the grain sizes of the material at hand, the localized damage instances, and both specimen sides are mechanically decoupled from each other. Furthermore, modeling grains as individual instances without detailed contextual information results in some degree of information loss. The data is split into train and test sets using stratified five-fold sampling to ensure the same distribution of damage instances in each fold despite their scarcity.

As the model of choice, a balanced random forest implementation from the *imbalanced learn* python library was selected [166]. A random search allowed identifying a presumably near-optimal set of hyperparameters. The number of individual

Table 5.4: Engineered features containing information on morphology, crystallography, and loading extracted from the multi-modal data. The adjacent grain (AG) column, and specifically the green check marks, indicate the features which contain information about neighbor grains.

Feature symbol	Description	AG
Morphological and topological features describing the shape and arrangement of grains and pores		
A_p	Grain area in pixel	✗
$A_{maxNeigh}$, $\overline{A_{neigh}}$	Max. and mean pixel area of adjoining grains	✓
N_{neigh}	Number of neighbor grains	✓
L_P	Perimeter of grain boundary (inner boundaries are neglected)	✗
AR	Equivalent area ellipse aspect ratio fitted to the grain	✗
L_{majAx}	Equivalent area ellipse major axis length	✗
L_{minAx}	Equivalent area ellipse minor axis length	✗
γ	Ellipse orientation angle	✗
$L_{maxSPtrace}$	Max. line intersection grain size in direction of maximally-loaded slip plane trace	✗
B_{sb}	Boolean indicating specimen boundary grains	✗
N_{pore}	Number of surface pores in grain	✗
$L_{poreSum/poreMax}$	Accumulated pore diameter and maximum pore diameter	✗
Crystallographic orientation, misorientation and quality-related features		
ϕ_1, Φ, ϕ_2	Mean Bunge Euler angles as crystal orientation representation	✗
$\omega_{min/max}$, $\overline{\omega}$	Min, max, and mean intergranular disorientation angle to any neighbor grain	✓
GOS	Intragranular angular disorientation spread from average orientation, see equation 2.10	✗
$\overline{KAM}$	Grain-averaged kernel average misorientation, see equation 2.9	✗
R_{tilt}	Proportion of tilt boundary candidates (< 15° angular deviation between GB segment trace and 2D GB crystallographic misorientation axis) with respect to overall grain boundary length	✓
R_{twin}	Proportion of Σ3 twin GB trace segments with respect to overall grain boundary length	✓
$\overline{CI}$	Grain-averaged EBSD confidence index, see [120]	✗
$\overline{IQ}$	Grain-averaged EBSD image quality, see [119]	✗
Loading-related and hybrid features		
C	Stiffness of a grain in the specimen axis direction	✗
$\Delta C_{min/max}$, $\overline{\Delta C}$	Min, max, and mean of specimen axis direction stiffness difference between adjacent grains	✓
σ_{vMmax}, $\overline{\sigma_{vM}}$	Grain mean and max continuum von Mises stress (FEM)	✗
m_{max}	Grain max. of pixel-wise computed Schmid factors assuming axial tensile load	✗
m'	Slip transmission factor considering alignment of slip plane normal and slip direction [289]	✓
M	Intergranular misorientation crack factor [290]	✓
f_{res}	Sample initial loading resonant frequency	✗

trees contributing to the model was set to 100, their maximum depth was restricted to 30. Each decision tree was trained with a bootstrap sample of seven randomly selected features. Leaf nodes with a minimum sample of one were permitted. Post-pruning (see section 2.3) was omitted.

Table 5.5: The confusion matrix for binary classification.

		Prediction	
		Negative	Positive
Actual	Negative	5326 ■	1996 ●
	Positive	67 ■	244 ●

Table 5.6: Performance of a balanced random forest classifier predicting the formation of protrusions.

Metric	Train	Test
Balanced accuracy	87.4 ± 0.3	75.3 ± 2.6
F1 score	25.2 ± 0.4	19.3 ± 1.2
ROC-AUC	98.4 ± 0.2	81.9 ± 2.2

The confusion matrix in table 5.5 shows that the model learns a decision boundary that culminates in a high number of false positives. This can be traced back to the application of class weights which weighs seldom damaged instances higher during training resulting in an over-prediction. In contrast, a classifier supplied with an unmodified data set and without class weights as a validation experiment was observed to put more emphasis on the negative class predictions resulting in many false negatives. Table 5.6 lists several performance metrics which are commonly used for imbalanced data. Note that while these metrics are commonly employed, the ROC-AUC can produce deceptive results when comparing different models on imbalanced data [210]. The individual performance metrics are provided as mean and standard deviations obtained from the five-fold cross-validation. The balanced accuracy metric poses the mean of the true positive rate (TPR) and the true negative rate (TNR). All other performance metrics are introduced in section 2.3. While all performance metrics listed in table 5.6 are defined in a range from 0 to 100%, the distinct values which they adopt for this specific task and data set can be traced back to the different confusion matrix elements which they take in consideration. It can be observed that there is a notable performance gap between the training and testing metrics.

For an individual specimen, the grain-wise classification performance is illustrated by plotting the confusion matrix element encoded as a symbol onto the corresponding grain centroid in figure 5.14.

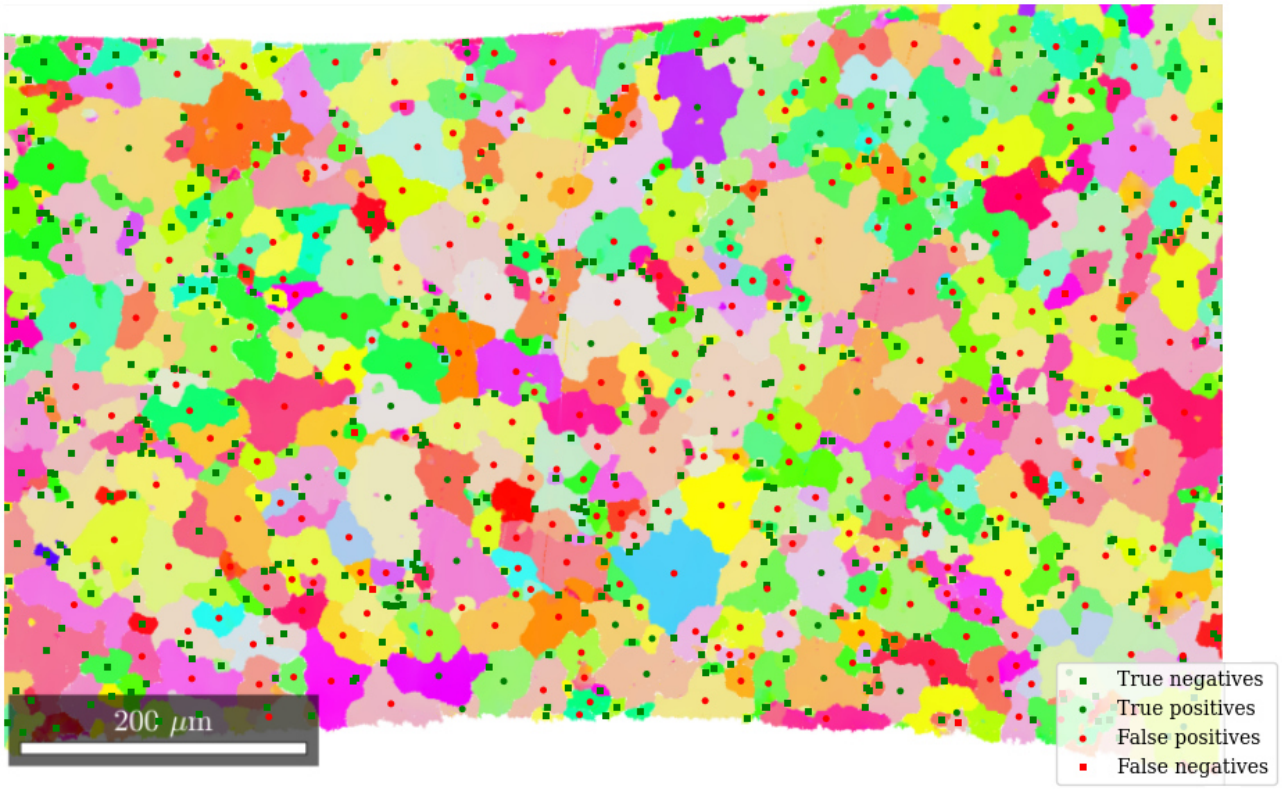

Figure 5.14: An inverse pole figure color-coded specimen microstructure map with the reference direction being the specimen normal (ND, 001_S). At the each grains' centroid a symbol is plotted that classifies its prediction in terms of confusion matrix elements (see table 5.5). A green circle and a square indicate true positive and negative predictions, respectively. The red circles and squares correspond to false positives and negatives.

It can be observed that many small grains are correctly classified as undamaged. Exclusively small grains were falsely predicted undamaged. This indicates that the feature or classifier does not capture the cause for some small grains developing protrusions. It can be observed that erroneous predictions are concentrated around the tapered region exposed to homogeneous and high load (cf. figure 4.2). Within this highly-loaded region, it appears that the classifier predicts all larger grains indiscriminately to contain damage. These observations hold true for the other seven specimen sides as well.

6 Discussion

Nuanced microstructural differences can alter the local materials response in the HCF or VHCF regime strongly. Here, in contrast to low cycle fatigue, local damage formation and evolution determines the fatigue life and its scatter. To transcribe the realistic local material behavior into a model with high fidelity, a comprehensive microstructure and defect representation as well as a solid mechanistic description at the microscopic scale is indispensable. Related modeling choices dictate whether the model represents the onset of plasticity, dislocation glide irreversibility, and crack nucleation realistically. This renders the task of predicting damage localization in these regimes very challenging. Aspects such as measurements and data post-processing need to be performed meticulously to capture the influence factors across the different scales and to preserve all relevant information. For instance, the disorientation threshold applied during grain reconstruction is known to affect the resulting microstructure substantially [17]. This in turn dictates the modulation of the applied stress as well as the effective impediments or nucleation sites for dislocations and cracks. The feature space is high dimensional. Even when considering solely microstructure, its comprehensive description requires a multitude of characterization methods and derived descriptors. However, non-destructive characterization methods which retrieve information on all microstructural flaws and other influence factors whilst encompassing all necessary scales are unavailable.

Therefore, for the HCF/VHCF regime, a probabilistic framework might be necessary, which considers and propagates aleatoric and epistemic uncertainties of measurements, data processing, and modeling. A similar framework was recently proposed to predict single crystal elastic stiffness parameters from spherical indentation of individual crystallographically-characterized grains in a polycrystal [291]. Nowadays, most predictive efforts revolve around computational models which attempt to describe the elasto-plastic polycrystalline response either phenomenologically or mechanistically in finite element frameworks. Moreover, machine learning approaches were proposed motivated by incomplete mechanistic knowledge [105, 106]. Probabilistic data-centered approaches to deduce relations in the data such as Bayesian neural networks or deep ensembles exist [292]. However, in presence of incomplete feature representations afflicted by uncertainty, even more grains are required to infer statistical correlations, causal relationships, and an understanding of variable interactions. This applies not only to the training of feature-based or representation learning approaches but also to physical, rule-based models during their appropriate model parameterization and validation. For both objectives, a well-curated, consistent, high-fidelity data set is highly beneficial. For data-driven approaches, such data can render model training more efficient, reduce data demand, and avoid model bias. An aggravating attendant circumstance that increases the data demand further is that HCF fatigue damage localization is limited to critical grain ensembles and therefore poses rare events. This also leads to an inherent imbalance of the data where only a small surface area proportion of the specimen contains damage. A large area needs to be sampled to obtain a reasonable representation of the extreme value distribution and to capture a sufficient amount of damage instances.

The necessary data quantity not only depends on the microstructural variance and exact loading conditions (i.e. the extent of the domain to be modeled) but also on the uncertainties introduced due to partial characterization and its non-ideal post-processing. Without rigorously assessing the uncertainties and the microstructural variance, the question of “How many grains suffice?” is very difficult to answer. Since a quantitative understanding of driving forces for early damage formation is unavailable, it is largely unexplored which combination of microstructural descriptors is essential to claim microstructural representativeness in terms of fatigue. While the grain boundary disorientation distribution was found to be largely consistent across multiple fatigue specimens (see figure 5.2), many more microstructure descriptors and their interplay affect fatigue damage initiation. Available data quantity also imposes constraints in terms of modeling choices. For instance, in the low data quantity regime, feature-based machine learning methods often outperform (pre-trained) deep learning variants for other tasks. However, it is arguable whether classical, feature-based machine learning methods have the necessary representation

power to model multidimensional, multi-scale problems such as fatigue, where variable interactions and spatial correlation patterns are crucial. Different approaches to encode contextual information about the spatial arrangement of phases or grain orientations have been proposed. These entail n-point statistics [293], pre-trained CNN encoders [294], graph-based approaches [295] or gray-level co-occurrence matrices [296]. These approaches are often applied in conjunction with linear or nonlinear dimensionality reduction techniques such as principal component analyses or t-distributed stochastic neighbor embedding (t-SNE) to extend engineered feature vectors. All aforementioned modeling techniques require large quantities of microstructure-informed damage evolution data.

The developed workflow enables the efficient creation of multimodal fatigue damage initiation and growth data sets. Resulting data sets combine microstructural information with surface morphology, and spatially resolved damage evolution information. The in-situ imaging captures the damage formation with unprecedented temporal resolution despite the high-frequency cycling necessary to conduct large-volume HCF testing. This facilitates the detailed investigation of individual crack instances and their growth retardation at specific crystallographic defects, see figure 5.10. The data provides the necessary spatial and temporal resolution to derive mechanistic knowledge through the analysis of individual damage instances. This way, the work at hand sheds some light on how grain boundaries quantitatively alter microstructurally short crack growth. Specifically, the transition from intergranular to transgranular crack growth and the deflection of the crack was found to inhibit the crack growth.

Aside from enabling such detailed analyzes, large data sets can be generated comparatively efficiently through the high degree of automation of the post-processing routine. These data sets can then enable data-driven investigations as demonstrated here. The data sets are attained by linking custom fatigue characterization with data post-processing methods, including deep learning semantic segmentation of damage locations, registration of heterogeneous image data, and data fusion. Provided thorough investigation of imaging properties, utilizing the presented in-situ light optical imaging can severely accelerate the acquisition of quantitative damage information. This is essential to establish an appropriate data foundation that permits deducing a mechanistic understanding in presence of pronounced uncertainties. Significant shortcomings in the imaging characteristics can be compensated for by combination and data fusion with the damage segmentation mask derived from the relatively high-resolution SEM damage images. Specifically, transcribing this high-fidelity information facilitates suppressing damage-unrelated light spots (false positives), accounting for shadowing, correcting the damage instance's shape, and discerning between semantic damage categories (cracks and protrusions) in the in-situ light optical images more effectively. Prospectively, the imaging can be further improved by circular oblique illumination [241], i.e., illuminating the specimen from multiple azimuthal directions. This concept is employed in rotating coherent scattering (ROCS) microscopy [213], which was demonstrated to improve imaging of slip markings profoundly in a master thesis scientifically supervised by the author [297]. The imaging demonstrated there can provide deeper mechanistic insights into damage formation while the larger field of view optics employed here is suitable to capture a variety of damage instances and hence statistical representations. Therefore, while the cycle-resolved light optical data in this work was solely used for analyzing individual damage instances, in the future it can also fuel the prediction of damage *evolution* by data-driven means.

Alongside the modalities that were utilized for correlative microscopy, further information on roughness, 3D microtexture, dislocation densities, residual stresses, chemical segregation, and internal inclusion defect distribution can be collected to round out the fatigue feature space. This would increase the data quality and reduce uncertainties. Complementary modalities can then be treated in a similar fashion by extending the multistage registration pipeline. In the following, a few potential future extensions to the outlined workflow are discussed. High-resolution EBSD can provide an estimate of the microscopic residual stresses and the geometrically-necessary dislocation density required to accommodate measured crystal rotations [298, 299]. The mean GND density estimate can also be obtained by conventional EBSD [300] and was found to depend on the step size applied during EBSD measurement [301, 302]. Appropriate step sizes are prescribed by the dislocation character and distribution within the material and for interstitial free steels were proposed to be around 0.5 µm [299, 303]. Therefore, the current EBSD data with an applied step size of 0.6 µm would facilitate a decent estimation presumably. Ideally, HR-EBSD is complemented with controlled electron channeling contrast imaging (ECCI) to additionally obtain detailed local information on statistically-stored, immobile dislocations and dislocation debris posing obstacles to crystallographic slip. This technique, as opposed to transmission electron microscopy, is non-destructive but restricted to surface-near interactions

and relies on backscattered electrons at lattice defects such as stacking faults or dislocations. When the Bragg condition for a specific reflector (i.e. lattice plane family) is approximately fulfilled, incident electrons tend to channel, unless crystallographic defects are present. In the case of grain boundary segregation, destructive techniques such as nanoscale secondary ion mass spectrometry (NanoSIMS) or atom probe tomography (APT) need to be consulted to retrieve information on their chemical nature and morphology. Such data can currently not be acquired in a large volume for many grain boundaries but could potentially be supplemented in a multi-fidelity approach where some portion of the data is enriched with such information. Especially when inner defects at the low nanometer scale are concerned, non-destructive measurement techniques are not available. While conventional X-ray computer tomography can achieve resolutions of approximately 50 nm in absorption or phase contrast for small specimens, corresponding lab-scale diffraction contrast tomography (DCT) can reconstruct reasonable 3D microtexture information only for grains larger than 20 µm providing a sufficient diffraction signal. For the ferritic steel material at hand, the smaller grain portion presumably can not be resolved with this technique and irregular grain shapes might pose a problem that would culminate in an unsatisfactory grain reconstruction. In this case, in spite of its suboptimal availability [304], synchrotron radiation facilities, owing to their beneficial beam characteristics, can to some degree provide a remedy where DCT with resolutions of approximately 1 µm is feasible whilst covering multiple $100 \times 100\,\mu m^2$ in a manageable measurement time. Another influencing factor not captured in the data is the variable magnitude of surface asperities at grain boundaries caused by the OP-S polishing relief. Their range was observed in a small sample size AFM experiment and outlined in section 5.2. Such asperities can act as dislocation sources or sinks and facilitate damage formation at grain boundaries. Extensive information on this can be gathered by using for instance a white-light interferometer. Furthermore, when multi-phase materials are concerned, the workflow can be extended to incorporate the phase segmentation efforts presented in [305] with contributions by the author.

In the following, the results of damage evolution, and deep learning models are discussed in detail.

6.1 Damage growth kinetics

The outlined approach facilitates the characterization of protrusion formation and microstructurally short crack growth. Formation of surface plasticity and crack growth are fundamentally different at the mechanistic level, and hence their corresponding elevated intensities regions in the in-situ images evolve differently. While surface plasticity such as extrusions and, to some extent, protrusions are comparatively localized and can be characterized through their topographic evolution, cracks exhibit directional growth. This has the consequence that different metrics are required to measure both phenomena. In the case of surface plasticity, the pixel accumulated intensity metric (see section 5.2) is appropriate in most cases. In contrast, cracks show pronounced variance in their growth behavior and directional propagation. Therefore, the accumulated intensity metric is inadequate in crack propagation analysis since growth direction information is omitted and temporary crack arrests are concealed. For the purpose of a crack analysis, morphological image operations such as skeletonization are required after segmentation to estimate the crack length. The deep learning predicted segmentation mask derived from SEM after fatigue (e.g., figure 5.9m) assists not only in the distinction of damage types and hence the decision of which damage metric to apply but also improves segmentation quality and connectivity assessment automatically.

As an example, the cyclic emergence of slip markings can be investigated, as was shown in figure 5.7. From evaluating the cyclic increase of accumulated intensity (see equation 3.1), it can be inferred that comparatively localized protrusions exhibit an abrupt growth, which terminates within $\Delta N \approx 2.5 \cdot 10^6$ cycles after reaching the target angular rotation amplitude under these loading conditions. This is in line with observations in literature where static extrusions were found to form quickly after reaching a steady state in the slip band [221, 222]. The height of such a static protrusion in a polycrystal according to the EGM [60] or Polák [59] model is then proportional to the wall volume proportion in the slip band, the wall vacancy density and the host grain size. Whether the latter proportionality holds true for the irregular grain shapes here or an effective local grain size needs to be taken into account is currently unknown. Since no TEM analyzes were performed information on dislocation distribution in slip bands is unavailable. Rarely, slip markings were found to traverse through grain boundaries. This was found in cases where grain boundary disorientation was small which is in accordance with previous observations

[306]. Similar cases of slip markings traversing through GBs were observed in literature for similar alloys [64]. In the course of further cycling, roughening of the initial static extrusion/protrusion occurs, including the formation of intrusions. This process was proposed to be related to random irreversible glide processes [307]. Therefore, there are at least two distinct physical mechanisms involved in the formation of static protrusions and their subsequent roughening. While the proposed growth model in equation 3.1 could fit the experimental intensity evolution data well, intuitively multiple mechanisms being involved would necessitate a more complicated formulation with multiple growth rates. The majority of the local intensity increase is associated with the growth of static protrusions. Afterward further roughening modulates the intensity profile to a smaller degree, cf. figure 3.12, where initially the intensity rises rapidly, and then fluctuations in the accumulated intensity are observed. Intrusions formed after roughening sometimes posed stage I shear crack initiation sites. Predominantly, however, cracks are initiated as intercrystalline cracks in regions close to evident micro plasticity. Intercrystalline crack initiation was reported to be the dominant mechanism in large grain bcc materials [70] since the slip plane asymmetry results in shape changes of the individual crystals and therefore decohesion at grain boundaries [68]. This represents a plausible line of reasoning for the material at hand as well.

The growth of microstructurally short fatigue cracks has been studied in the literature extensively. Typically, microstructurally short crack growth is described as a slip-assisted process occurring in specific crystallographic directions (slip systems) [308–310]. This could be confirmed for the considered crack through a slip plane trace study, see figure A.2. In this study, ripples in the transgranular crack path indicate that it alternates between the traces of two crystallographic slip planes. There are multiple reported crack retardation mechanisms, including crack branching [311] and impingement of the crack tip at grain or phase boundaries. Latter is sometimes associated with crack redirection or crack tip blunting through the dissemination of dislocations [312]. This blunting can be hampered by back stresses originating from dislocation pile-ups ahead of the crack tip [312]. The transition from the intergranular to the transgranular crack in the upper section (figure 5.10a #1) is potentially ascribed to the sharp GB trace direction change from approximately orthogonal to approximately parallel to the specimen axis. This transition is linked with a crack arrest which might occur due to crack reorientation to a specific slip plane when entering the transgranular mode. While propagating through G1, intermittent crack propagation (figure 5.10a #2, #3) occurs despite the absence of surface defects. However, retardation instance #2 is accompanied by a crack direction change towards the normal of the specimen axis. Such a behavior is often linked with the transition to macroscopic crack growth (stage II) [313]. Indeed, this transition occurs at shorter absolute crack length in miniaturized specimen [314]. Another, and in this case arguably more credible explanation could be the interaction of the crack with interfaces or redirection towards stress concentrations that reside subsurface. Similarly, surrounding damage fields such as the protrusions which formed above G1 (see figure 5.9) could act as stress concentrations and promote the crack growth towards them to release the associated strain energy. Further, it is noteworthy that presumably during crack growth, a plasticity trace loop develops annotated by the green arrows in figure 5.9o. Since there was no in-situ intensity elevation observed corresponding to this shallow surface plasticity, its cyclic emergence can not be traced back. However, it can be tentatively assumed that dislocations emitted from the crack's plastic zone interacted with and piled up at the G1-G6 grain boundary amongst others, and thereby affected the crack growth. From this study, it can be deduced that these data sets facilitate the investigation of how specific microstructural features affect stage I crack growth.

6.2 Assessment of fatigue damage segmentation

6.2.1 Source domain model evaluation

For the ferritic steel (source domain), the trained U-Nets in most cases achieved performances that resemble manual expert annotation. Especially cracks were annotated by the model reliably unless attended by artifacts. In the case of protrusions, the segmentation was satisfying but contained some inconsistencies with human annotation at the protrusion borders, see figure 5.12d. The borders are challenging to segment since they rather fade out than exhibiting a clear delimitation. As a matter of fact, human labels are afflicted with uncertainty in these border regions as well.

The scatter plots in figure 6.1, show the tile image-wise *IoU* plotted over the corresponding contained amount of crack/extrusion pixel for the ferrite trained model performing best on the ferritic steel test set (table 5.1 #11).

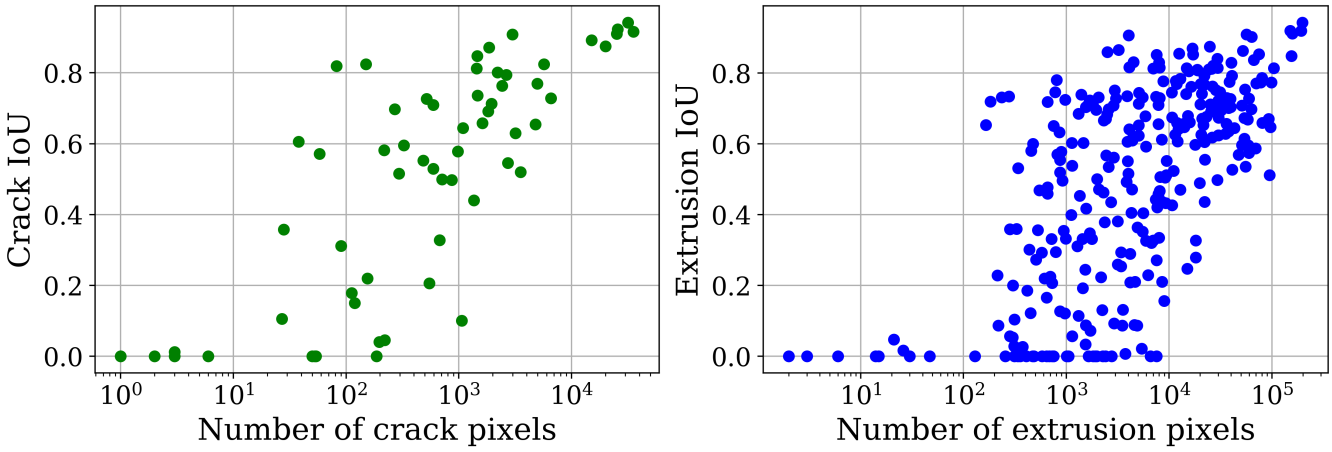

Figure 6.1: Tile image-wise intersection over union metric shown as a function of contained crack and extrusion pixels. Adopted from [37].

The characteristic shape of the distribution in figure 6.1 can be explained by the fact that extrusions composed of few pixels are usually rather shallow. Therefore, by virtue of the SEM image formation, low local contrasts arise and the detection is hampered. While these shallow regions can cause problems for the network, they are typically not critical defects determining fatigue life. In a notable amount of cases, image edge effects and in particular faded-out boundary sections of extrusions barely extending into a tile causes poor detectability. When the model is deployed it would rather operate on complete (i.e., untiled) images. Then incomplete information at tile borders is not of concern. Note that all DL model evaluations in this work were performed on tiles and therefore presumably underestimate the performance on complete images by a few percent. Another contribution to the trends in figure 6.1 is that the IoU metric inherently penalizes pixel discrepancies in smaller foreground instances more.

Generally, damage pixels for which foreground class affiliation was not obvious were manually labeled as extrusions. However, as the transition of intrusions to micro cracks is gradual, the discriminability is poor causing potential labeling inconsistencies. The same applies to cases of occlusion caused by debris where connectivity assessment of cracks is impeded. While in some instances only visible parts of the crack were labeled as such, in others, annotations were based on the assumption that the crack is continuous beneath the debris. There are different techniques reported in the literature to deal with occlusion. For instance, random erasing of pixel regions within the image was suggested in [315] to force the network to take into account the context in the entire image rather than only a small portion. In addition, specimen boundaries which suffer from over- or underexposure are often detrimental to extrusions and cracks segmentation.

This is presumably also linked to the fact that brightness augmentation was adverse for the extrusion class. Extrusions/intrusions or cracks often exhibit extreme intensity values. Their sharp edges cause substantially higher or lower secondary electron emission rates. Additionally, increasing or decreasing the brightness through augmentation results in an information loss in the high or low-intensity regions, concealing the features of the damage locations. This applies also to the specimen border where the specimen surface orientation relative to the SE2 detector causes over- or underexposure and concealing of damage. Due to the continuum strain distribution shown in figure 4.2, a slightly elevated portion of damage spots is located near specimen boundaries. The Gaussian noise especially reducing the ferrite IoU_c can potentially be attributed to features of narrow cracks not being preserved when the noise is applied. This can be worsened by downsampling. Downsampling turned out to be detrimental at all times due to information loss but did affect the extrusion class more (ΔIoU_c=0.03 and ΔIoU_e=0.05 between Test #10 and #11 in table 5.1) since it generally contained finer features which are more likely to get lost.

Moreover, the features in this damage segmentation task are comparatively localized (i.e., the spacing between intrusions and extrusions is small). In contrast, the lath-shaped bainite segmentation task addressed by the author in [305] relies on long-range image features. In such cases where downsampling does not cause notable information loss but compresses the relevant image features into a smaller pixel region, i.e. the model's receptive field, CNN models can benefit. Another aspect, which can be learned from comparison with the bainite segmentation task is the response of CNN models to image augmentation in presence of distinct input image variance. In the work at hand, pronounced variance in the data ascribed to specimen fabrication and imaging in particular made data augmentation indispensable. For microscopy data sets acquired in a very repeatable manner, image augmentation leads only to minimal improvements [305].

6.2.2 Domain generalization

In general, the domain generalization of a model depends largely on the domain gap, i.e. how similar the discriminative features are in images of both domains. Ferrite damage utilized for training consisted solely of extensive protrusions superimposed with extrusions and intrusions. This facilitated the domain transfer to copper in which most of the damage had a similar (but wavier) morphology. In contrast, the domain transfer to martensite solely consisting of localized tongue-like extrusions is impeded. Since no cracks were observed in the martensite and copper data set, this study is confined to extrusions.

Gaussian noise improves the performance of the ferrite-trained model on the martensite data as the latter exhibits an inferior signal-to-noise ratio. One aspect that is detrimental to the martensite IoUs is residual OP-S particle agglomerates at the specimen surface. These are surface polishing-induced artifacts at the same scale as martensite extrusions and can have a similar shape, as shown by the annotation in figure 3.5j. Apart from this, the overall poor model performances on the martensite domain are probably owed to a more pronounced domain gap between ferrite and martensite extrusions. In particular, this is attributed to martensite containing almost exclusively tongue-like extrusions. Furthermore, the damage sites in martensitic steel are much smaller in physical and pixel size relative to those in ferritic steel.

The performance observed in copper for some models exceeding that of the ferrite source domain can be explained by copper extrusions exhibiting more distinct boundaries to undamaged areas. In contrast to ferrite, plasticity in copper was macroscopic i.e. not confined to individual grains and substantially easier to detect by eye. Meanwhile, the pronounced performance drop in contrast augmented models implies that these models become invariant to contrast changes thus failing to make use of the distinct boundaries in copper extrusions. This feature is particularly important due to the absence of directional features in the copper extrusions i.e., no clear slip trace orientations are present as opposed to ferrite extrusions. Further, there are some limiting factors for the domain generalization of ferrite-trained models to copper. Since in contrast to the copper data set, there were virtually no tongue-like extrusions present in the ferrite data set, the accuracy for such extrusions was moderate. The merging of close extrusions and in particular tongue-like extrusions in copper (see figure 5.13a) can be ascribed to two characteristics of the damage instances in the ferritic steel training data. Namely, the absence of distinct extrusion boundaries (see figure 5.12d) and the predominance of extensive protrusions. These characteristics of ferritic steel are learned during training and hamper the detection of small-sized tongue-like extrusions in copper represented by fewer pixels.

The elastic distortions represent an essential addition to the augmentation pipeline, irrespective of the material domain. It can possibly be justified by attained (extrusion) shape invariance. For instance, warping the images of ferritic material with its straight slip traces can produce similar image textures resembling those of the wavy slip structures observed in copper. According to [197], in case of small training sets, a central role can be ascribed to elastic transformations. Microscope operators tend to have different preferred brightness and contrast settings. Hence, in terms of image intensity-affecting augmentations, it is essential to tune the augmentation hyperparameters for a large range of realistic image gray value distributions whilst minimizing information loss. Furthermore, the performance decrease (cf. #10 and #11 of table 5.1) related to downsampling-induced information loss in images is prevalent throughout every material domain and affirms similar observations in literature [316].

The less pronounced IoU variance in the source domain indicates that extrusions and cracks in its training data set are representative of the ones in its test data set. On the other hand, extrusions from the source domain are not really representative

of the ones in the target domains. Hence, the models trained on the source domain data set rely more on the careful selection of augmentation types so that they could be generalized well to the extrusions in the target domain. On the one hand, this can be achieved if the augmentations on the source domain extrusions render them representative of the ones in the target domain. On the other hand, augmentations can force the network to not learn features that are too specific to the source domain, thus learning features that apply to both domains.

Additional experimentation might be required to find clues on which network layer benefits in terms of generalization from the augmentations specifically. In [317], it has been demonstrated that out-of-distribution generalization is layer dependent. Typically, initial layers of any two trained networks share common characteristics and extract low-level features such as edges, corners, and blobs. In contrast, the following layers become increasingly specific and difficult to transfer to alternate domains, unless fine tuning (transfer learning) is applied.

The adaptation of the augmentation pipeline holds pronounced potential to improve the foreground class IoUs of each domain. It can be inferred from the source domain performance that the trained networks, with the right set of augmentations, can cope rather well with loading amplitude and subjective SEM setting-induced variance (different working distances, magnifications, and brightness/contrast settings). In this work, exclusively SE2 SEM images with a stable perspective were utilized, which posed a simplification to some extent. Therefore, conclusions on the domain generalization to alternate common SEM detector types and different imaging modalities are difficult to draw. Evidently, however, generalizability across material domains can potentially pose a difficult task. A difficulty lies in finding a mutually beneficial augmentation setting for every material domain. Further, even obtaining such an augmentation setting cannot ensure satisfactory domain generalization. In particular, from the model evaluations on different material data sets it can be deduced, that for materials with a more pronounced domain gap to the source domain, i.e. martensite, a different and more elaborate route is required.

6.2.3 Multi-domain training

In this section, the combined ferritic and martensitic steel training is assessed. The combined training (test #2 in table 5.3) improved the IoU_e on martensite over the domain generalization values (table 5.2) by a large margin. However, the combined training did not reach the performance achieved by sole martensitic steel training (test #1 in table 5.3). This suggests that the prevalence of ferritic steel damage in the combined data set adversely affects performance on the martensite testing set. Note that a pronounced imbalance arises since martensite is not only underrepresented in terms of training tile amount but also the damage area (individual damage instances are significantly smaller), see table 3.1. The hypothesis was substantiated since imbalance correction alleviated this issue and led to significant improvements, cf. table 5.3 tests #3–5. Therefore, the imbalance correction represented an essential step to eliminate the network's inclination to learn only extensive protrusions that covered a larger pixel fraction in the data set. Additional martensite data rendered the training set more representative of the variety of damage locations that exist. Hence, a more general representation of extrusions could be learned which manifests in a more balanced multi-domain segmentation performance. This especially held true after imbalance correction. Then the combined model's performance on individual domains approximately matched the performances of either network dedicated to the respective domains.

It can be observed that this conjunct model is capable of assessing the connectivity of close proximity tongue-like extrusions significantly better than solely ferrite-trained models. In particular, adding martensite extrusions to the training suppresses the merging of tongue-like extrusions. The balancing methods slightly decreasing the copper performance could be attributed to strongly weighted tongue-like extrusions shifting the focus away from protrusions which make up the largest portion of the damaged area in copper. To conclude, these observations suggest that there is an optimal balance of ferrite (protrusions) and martensite (tongue-like extrusions) data at which the segmentation performance is maximized for copper.

The accurate segmentation of damage instances is important when the damage instances are linked to specific microstructural entities such as grains, their boundaries, or other defects. Especially since slip markings are often localized at grain boundaries, sometimes impinging on them, sometimes transitioning through them, and image registration is non-ideal, there are high requirements on damage segmentation accuracy. An improved segmentation model which performs reliable segmentation of

tongue-like extrusions can be utilized if the overall workflow (see figure 1.1) is supposed to be applied to hierarchically-structured materials with very localized damage instances.

There is still potential for improvement of model #6 in table 5.3, e.g. by optimizing the augmentation pipeline for this multi-domain setting and including the copper data in training. Nonetheless, this model presents a promising starting point, as it achieves satisfactory performance on multiple domains and can be expected to transfer well to a multitude of other metallic materials and imaging conditions. Larger and more expressive segmentation models potentially could optimize for multiple domains simultaneously with fewer compromises between the material domains. The evaluation of the conjunct model's transferability to further material domains will be the subject of future works.

6.3 Damage modeling use cases

6.3.1 Validation of microstructure-sensitive simulations

Data generated by the workflow presented here was utilized to assess the spatial correlation between actual damage locations and ones predicted by a phenomenological crystal plasticity fatigue model. Detailed results of this collaborative study are published elsewhere [38]. Crystal plasticity simulations take the crystallographic orientation and slip systems in polycrystals into account to compute load path-dependent strengthening and thereby provide a prediction on how the microstructure modulates mechanical fields. Stress and strain fields from the simulation are then translated into a damage accommodation tendency through so-called fatigue indicator parameters (FIP).

The constitutive model assumed plasticity through dislocation slip on $\{1\,1\,0\}\langle 1\,1\,1\rangle$ as the sole mechanism of plastic deformation, Schmid behavior (only shear components on permitted slip systems contribute to plastic deformation), a local model, and small strains. In this context 'local model' refers to the assumption that the critical resolved shear stress (CRSS) which needs to be exceeded to trigger dislocation movement exclusively depends on the deformation state of the individual element [318]. In the simulation, a sub-modeling approach was followed where first the macroscopic boundary conditions were imposed at the specimen level to compute a continuum response. Subsequently, this continuum response was transferred onto the actual specimen microstructure. This facilitated a direct comparison of FIP fields with actual microstructural damage locations, see figure 6.2a. As reference damage locations were semantically segmented by deep learning (see section 6.2), the data sets complement existing HR-DIC-based validation approaches. While HR-DIC can measure plastic strains originating from slip bands quantitatively and represents a crucial form of validation [319], segmentation methods consider image texture to detect actual damage locations. Therefore, comparison to damage segmentation maps provides a form of validation that complements existing approaches well.

Moreover, four instances of microstructurally short crack growth observed in the data were utilized to validate a phenomenological CPFEM-based crack growth model [39], see figure 6.2b. In this effort, the crack initiation location was provided as an individual element boundary condition. The predicted 2D crack paths resembled those observed in our tests [39]. However, when observing figure 6.2b, which represents the crack instance extensively discussed in section 6.1, it is apparent that the intragranular alteration of the crack path is not captured by the CPFEM model. This arises since the formulation assumes that cracks traverse through the whole grain without changing the slip plane. However, even when intragranular redirection would be permitted, this local model presumably could not capture comparatively long-range effects such as strain fields originating from micro plasticity in the vicinity or dislocation emission from the crack tip into grain boundaries. While crack paths could be predicted decently well, hot spots in computed fatigue indicator parameters did not show a satisfying spatial correlation with actual protrusion locations [38]. Among the applied Fatemi-Socie (see figure 6.2), accumulated plastic slip, and dissipated energy FIPs, none seemed to coincide with the protrusion locations particularly well. There are numerous potential deviations between the phenomenological model and known physical driving forces for protrusion formation. For instance, aspects such as the development of dislocation structures and diffusion of point defects are disregarded but are known to play a central role in protrusion formation. However, also the assumption of sole $\{1\,1\,0\}\langle 1\,1\,1\rangle$ slip [69], prismatic grain shapes

[320], Schmid behaviour [62] likely introduce deviations. The exact causes for the mediocre agreement were difficult to pinpoint since the data processing, constitutive modeling, and FIP formulations are all potentially contributing to the deviations. In this case, complementary HR-DIC measurements can provide intermediate validation of plastic strain fields and slip bands to exclude error sources systematically. Comparisons as illustrated here between computational approaches and experimental data can help validate models, improve their parameterization efforts, and give insights into missing modeling aspects. The unsatisfactory protrusion prediction performance of this phenomenological CPFE model and the incomplete knowledge of protrusion formation motivate the use of data-driven approaches which can learn correlations in the data.

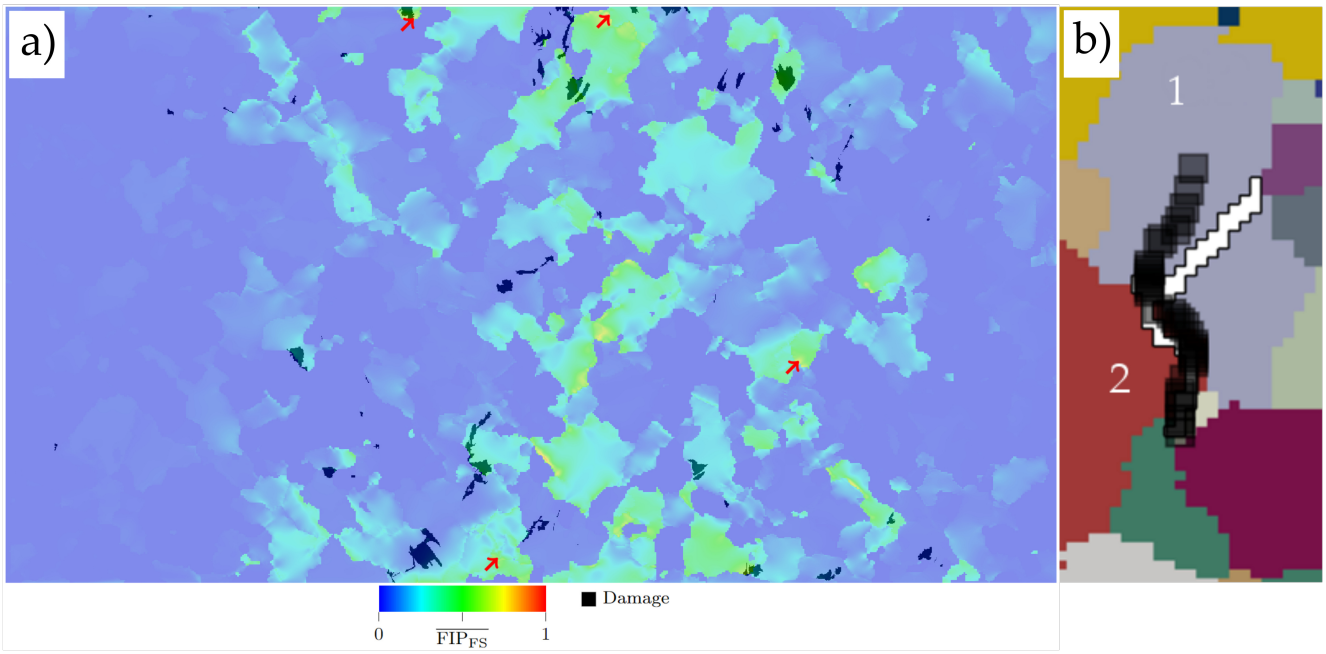

Figure 6.2: a) Normalized FIP_{FS} distribution superimposed with damage locations colored in black. The red arrows indicate FIP values that are greater than 80% of the maximum FIP value found with a search radius of 10 element edge lengths. **b)** The actual (black) and predicted crack path (white) are superimposed onto a unique grain color image. This crack instance corresponds to the one previously analyzed in figure 5.9 and 5.10. Subfigures **a)** and **b)** are adopted from [38] and [39], respectively.

6.3.2 Damage classification

In this work, for the cause of protrusion prediction, the microstructure is represented by clustering pixel regions corresponding to the physical concept of grains. Then each grain poses an individual data instance and is assigned multidimensional features describing the grain and its local environment, see table 5.4. This approach exhibits similarities to and was inspired by efforts by Orme et al. [137], and Mangal [139] who used tree-based classifiers with instances representing grains to predict deformation twinning in magnesium alloys and load hot spots in crystal plasticity simulation, respectively. In the work of Sharma et al. [321] which represented an extension of [137] focusing on GB regions as instances, the performance saturated reaching approximately 85% accuracy at 15.700 considered GB training instances corresponding to 770 twinning events. In contrast to their work on twinning where large strains of 7.5% in an AZ31 magnesium alloy were considered, the data set at hand is substantially more imbalanced and the underlying statistics are more prone to microstructural details. This indicates that more data would be beneficial for the problem at hand. Additionally, while the coupling between damage instances is deemed mostly negligible due to the HCF loading and the relevant specimen and microstructure scales, it can not be excluded entirely. In the work of Mangal [322], an autocorrelation analysis was applied using two-point statistics in order to explore and rule out mutual interference between hot spot instances. A similar approach can be applied here using the high-fidelity damage maps when carefully taking the non-uniform stress state into consideration. As opposed to the work of Mangal, the work addressed here represents an effort based on experimental data, where the underlying mechanisms are comparatively

more diverse resulting in higher variance and the presence of confounding variables. When it comes to the coupling between grains, the approach in this work was to enrich the feature vector with some information about neighbor grains, see the AG column in table 5.4 rather than modeling grain adjacency explicitly. Intuitively, the modeling of fatigue damage initiation should rely on nuanced interactions between adjacent grains and their descriptors. This can not be captured with the current approach as grains are assumed to be independent of each other and the grain instances are not informed by the neighbor grain's properties. In very few instances, slip traces propagated through a grain boundary. This puts the separate treatment of grain instances and such a tabular approach into question.

Considering the complexity of the prediction task, incomplete feature space, and inherent data scarcity, the binary balanced random forest classifier achieved a decent performance in identifying grains that would develop damage. While the performance is substantially better than a naïve random classifier, an appropriate baseline performance is currently missing, e.g., from crystal plasticity. A high number of false positives remains and needs to be mitigated going onward. Nonetheless, in a reliability setting, false positive predictions are generally preferable over false negatives. Visualizing the prediction performance on a map containing microstructural information can provide insights into the characteristics of the classifier. For instance, the observations made based on figure 5.14 indicate that the classifier devotes its attention mainly to grain size and local continuum FE von Mises stress. The preliminary observations on feature importance are largely in line with insights granted by the mean decrease in impurity metric (see section 2.3). It pointed to grain size-related metrics being the most important. Aside from that, the Young's modulus of each grain in the specimen axis direction ranked among the most important features. Moreover, the mean grain-wise von Mises stress from continuum FE simulation contributes to a major extent. More nuanced relations in the data on the other hand such as the interaction between microtextural and grain boundary descriptors are apparently not inferred during training. This could be either ascribed to their inappropriate parameterization or the model's inability to find such correlations despite permitting a high depth of individual decision trees. Indeed, decision tree and random forest models require discriminative features and do not combine individual features. Another potentially important aspect is that during the construction of the individual trees of the bagging classifier, the feature grouping (i.e. three Euler angles) was not respected.

Prospectively, separating the instances with differing damage types further before training classifiers could be interesting. Protrusions that emerge at different time points or at defects can be distinguished efficiently by the outlined workflow through the post-processing of the in-situ image series data and the spatially-correlated high-resolution pore information, respectively. These different types of damage instances, such as protrusions at grain boundaries or at pores, have distinct underlying damaging mechanisms and statistics. Training multiple classifiers for the different damage types therefore can prove beneficial. The current approach did consider only limited contextual information. While for instance the maximum disorientation or the maximum Young's modulus difference across grain boundaries to *any* neighbor grain are considered, their spatial localization is discarded. However, most protrusions were observed to not cover the whole grain but rather to be localized at specific grain boundaries. This indicates that spatial information and a comprehensive description of individual grain boundaries are required. Thus, in the future, different methods to incorporate contextual information as described in the first part of the discussion should be tested. Enriching the feature space, aside from incorporating contextual information, is one of the aspects which should be attributed the highest significance going onward. This could range from the incorporation of easily accessible features such as grain boundary triple points known as regions of high incompatibility up to consideration of features from additional modalities as elaborated at the beginning of the discussion section. On this end, diffraction contrast tomography data for 3D microtexture and phase-contrast CT data for internal MnS-inclusion information was collected and will be utilized in future works.

While the produced data set does not exhaustively capture all fatigue influence factors, it represents a relatively large quantity damage data set with unprecedented information on the local damage evolution. It stores the reality for future validations and allows for sensitivity studies to purposefully incorporate relevant features into the model. An exemplary data set for one specimen was published [323]. The underlying methods of the workflow including deep learning damage segmentation and multimodal data registration are assumed to be applicable to a wide range of materials. Arguably, deep learning networks trained in a supervised fashion pose the limiting factor in this regard since their generalization capability across data sets was demonstrated to be compromised. However, novel machine learning concepts such as unsupervised domain adaptation [324]

can render such damage segmentation models transferable to a wider range of materials and damage characteristics without requiring annotated data for new target materials.

6.4 Prospects of machine learning in mechanics of materials

The ongoing digital transformation of materials, which focuses on storing, semantically describing, and connecting materials data along with its meta data, and background knowledge, will undoubtedly fuel future machine learning efforts. It is safe to state, that increased availability of structured data will facilitate the usage of models with comparatively higher representation power such as vision transformers which rely on less inductive bias [207]. While nowadays, owing to limited data connectivity, mostly short-range tasks, e.g., detecting phases in images, are tackled, linking data throughout process chains will presumably enable addressing more challenging tasks, such as predicting properties directly from process parameters. Aside from that, also unstructured data can be utilized in the context of deep learning, such as for the training of self-supervised models [147].

In materials science and mechanics of materials, there is a wealth of computational simulation techniques with a range of underlying assumptions, characteristic scales, and simplifications. These comprise density functional theory, molecular dynamics, particle methods, phase field approaches, discrete dislocation dynamics, crystal plasticity, and multi-physics FE to name a few. Increasing the interoperability between the corresponding synthetic data promotes enriching the data space with complementing implicit knowledge on specific mechanisms and scales. The same applies to experimental data. For a few smartly sampled grains, experimental information such as GB segregation, full GB character, or precipitation character can be supplemented by applying the aforementioned characterization techniques. Such grains can act as high-fidelity data in a multi-fidelity setting to simultaneously achieve improved generalizability and prediction accuracy. In literature, such a multi-fidelity approach has been proposed using different ab initio simulation data to jointly train a graph-based neural network regression model for band gap prediction [136]. Accordingly, such experimental multimodal and multi-fidelity data can reduce epistemic uncertainty in fatigue damage prediction in the future. In contrast to rule-based simulations, experimental data tends to capture a more comprehensive mechanistic description. While experimental limitations consist of comparatively poorly defined boundary conditions, restricted resolutions, and disturbing physical phenomenons through the measurement itself, simulations are limited by computational constraints and incomplete knowledge. This dichotomy potentially can be exploited to effectively augment the data space with complementary features and alleviate shortcomings of the other data type, respectively.

Another promising candidate to exploit the existing knowledge and reduce annotated data requirements are *physics-informed* or *physics-constrained* DL models [325]. Rather than supplying a multitude of input-output pairs, conditions that represent domain knowledge are imposed on the output space. In such cases, explicit domain knowledge is typically encoded into the loss function by softly constraining it. For microstructure inference, laws from thermodynamics including different crystal growth or segregation/precipitate formation models potentially can condition DL models. A yield strength prediction model could utilize the well-known Hall-Petch relation.

Knowledge graphs currently focus on capturing factual knowledge in triples (e.g., “twin boundary–is a–grain boundary”), transcribed in the resource description framework (RDF) or extensions such as RDF*. Recent research attempts to also include common sense knowledge as well as cause-effect relationships. For the description of processes, the notion of events or temporal information, in general, is important. Concise modeling of these concepts facilitates deductive and inductive reasoning to continuously extend the graph representations [326]. Computer vision enhanced with knowledge graphs can extract contextual knowledge better [327] and enable complete scene comprehension, e.g., “A crack initiates at a grain boundary due to impinging slip markings”. In principle, various types of *literals* can be used along with knowledge contained in graphs in so-called knowledge graph embedding (KGE) approaches [328].

Given the large number of different materials and processes combined with the typically time-intensive generation of data sets for many tasks, data scarcity in materials science is likely to subsist. As a consequence of emerging high-speed image acquisition technologies, raw data can in principle available in abundance. However, manual annotation processes often pose

the bottleneck in the creation of representative experimental data sets. This hampers the paradigm of supervised learning in the materials science domain. Advances in correlative microscopy with EBSD measurements, as presented in [305], promote routines for automatically generating annotations which renders it less labor-intensive and improves the quality of the presumed ground truth. Nonetheless, strategies for data frugal training as well as model generalization to alternate materials or processing conditions are indispensable to cope with materials diversity. Fortunately, machine learning underwent rapid advances in the past two decades and now provides a vast number of corresponding approaches which can be adopted in materials research.

One example of data frugal techniques includes *semi-supervised learning*. It tackles the task of additionally learning from unlabeled data. This is often achieved by so-called pseudo labeling, consistency regularization, as well as model distillation [329] and reaches performances approaching full supervision with mere fractions of the annotated data for many tasks. Rather than providing pixel-wise annotations for training a segmentation network, in a *weakly-supervised learning* setting, e.g., image-wise annotations are used. There are different annotation abstraction levels ranging from bounding boxes [330] to naming the classes present in an image [331]. Typical methodologies rely on classification networks which provide seeds for the segmentation network, and constrained seed region growing to respect object boundaries [331, 332]. In recent years a leap in weakly-supervised segmentation performance was achieved [333], rendering it a promising method for phase fractions. This is affirmed since well-contrasted grain boundaries presumably can pose distinct and suitable borders for region growing. In particular, for metallographic segmentation tasks in which target phases are often dispersed across the whole image, pixel-wise annotation is cumbersome. Here it can be particularly worthwhile to replace manual pixel-wise annotations with appropriate weak labels.

Alternate techniques called *semi- or unsupervised domain adaption* evolve around the idea that for a specific task (e.g., segmentation) annotated data of one source domain (e.g., material A) can be used together with non-annotated or minimally annotated data of a target domain (e.g., material B) to produce meaningful predictions in latter. The methods achieving this rely on feature matching between both domains, and self-training to provide pseudo labels or generative networks to produce target data [334]. The range of materials and processes that can be covered with such techniques in material scientific challenges is yet to be unveiled. In this regard, the usage of unsupervised domain adaptation to transfer between distinct materials processing routes and image modalities was explored recently [324]. This represents an extension of the phase segmentation work presented here and was part of a master thesis supervised by the author. Moreover, also in this context of domain transfers, the materials science domain can profit from its longstanding experience in rule-based, realistic simulation techniques. The resulting synthetic data can be exploited in domain adaptation to obtain annotated data in a source domain [32, 335] or for pre-training. In [336], the error rate of a domain discriminator network was utilized to quantify the domain gap relying only on unlabeled input images of both domains. Further, it was illustrated that the domain gap in conjunction with the error rate of the source task can provide an estimate of how well a network trained in one domain transfers to another domain. This estimate in the future could provide an indication of, whether re-purposing a trained network for another target material is feasible or domain adaptation is necessary.

Undoubtedly, there are large gaps in our knowledge when it comes to how the microstructural features quantitatively interact to affect the material's local response to cyclic loading. However, recent developments in the field of machine learning can provide us with statistical tools to accelerate research and unravel this high-dimensional problem.

References

[1] H. Eickenbusch and O. Krauss, "Werkstoffinnovationen für nachhaltige Mobilität und Energieversorgung," *VDI-Platz*, vol. 1, p. 40468, 2014.

[2] M. Yellishetty, P. Ranjith, and A. Tharumarajah, "Iron ore and steel production trends and material flows in the world: Is this really sustainable?" *Resources, conservation and recycling*, vol. 54, no. 12, pp. 1084–1094, 2010.

[3] M. Fischedick, J. Marzinkowski, P. Winzer, and M. Weigel, "Techno-economic evaluation of innovative steel production technologies," *Journal of Cleaner Production*, vol. 84, pp. 563–580, 2014.

[4] National Science and Technology Council (US), *Materials genome initiative for global competitiveness*. Executive Office of the President, National Science and Technology Council, 2011.

[5] S. O'Meara, "Materials science is helping to transform China into a high-tech economy," *Nature*, vol. 567, no. 7748, pp. S1–S1, 2019.

[6] "Innovations-Plattform MaterialDigital," https://www.materialdigital.de/, visited 2021-05-09.

[7] C. Eberl, M. Niebel, E. Bitzek, T. Dahmen, F. Fritzen, P. Gumbsch, T. Hickel, S. Klein, F. Mücklich, M. S. Müller *et al.*, "Consortium proposal NFDI-MatWerk," 2021.

[8] T. Zhou, Z. Song, and K. Sundmacher, "Big data creates new opportunities for materials research: A review on methods and applications of machine learning for materials design," *Engineering*, vol. 5, no. 6, pp. 1017–1026, 2019.

[9] R. O. Ritchie, "Mechanisms of fatigue-crack propagation in ductile and brittle solids," *International journal of Fracture*, vol. 100, no. 1, pp. 55–83, 1999.

[10] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne *et al.*, "The FAIR guiding principles for scientific data management and stewardship," *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.

[11] J. Elmer, S. Allen, and T. Eagar, "Microstructural development during solidification of stainless steel alloys," *Metallurgical transactions A*, vol. 20, no. 10, pp. 2117–2131, 1989.

[12] L. Renversade, R. Quey, W. Ludwig, D. Menasche, S. Maddali, R. M. Suter, and A. Borbély, "Comparison between diffraction contrast tomography and high-energy diffraction microscopy on a slightly deformed aluminium alloy," *IUCrJ*, vol. 3, no. 1, pp. 32–42, 2016.

[13] R. Pokharel, "3D bulk grain evolution in polycrystalline Cu: comparison between HEDM observation and FFT based crystal plasticity simulations," *In A. Bajaj, P. Zavattieri, M. Koslowski, T. Siegmund (Eds.). Proceedings of the Society of Engineering Science 51st Annual Technical Meeting*, 2014.

[14] R. Pokharel, J. Lind, A. K. Kanjarla, R. A. Lebensohn, S. F. Li, P. Kenesei, R. M. Suter, and A. D. Rollett, "Polycrystal plasticity: comparison between grain-scale observations of deformation and simulations," *Annu. Rev. Condens. Matter Phys.*, vol. 5, no. 1, pp. 317–346, 2014.

[15] A. D. Spear, S. F. Li, J. F. Lind, R. M. Suter, and A. R. Ingraffea, "Three-dimensional characterization of microstructurally small fatigue-crack evolution using quantitative fractography combined with post-mortem X-ray tomography and high-energy X-ray diffraction microscopy," *Acta materialia*, vol. 76, pp. 413–424, 2014.

[16] D. Naragani, M. D. Sangid, P. A. Shade, J. C. Schuren, H. Sharma, J.-S. Park, P. Kenesei, J. V. Bernier, T. J. Turner, and I. Parr, "Investigation of fatigue crack initiation from a non-metallic inclusion via high energy X-ray diffraction microscopy," *Acta Materialia*, vol. 137, pp. 71–84, 2017.

[17] J. Miao, T. M. Pollock, and J. W. Jones, "Microstructural extremes and the transition from fatigue crack initiation to small crack growth in a polycrystalline nickel-base superalloy," *Acta Materialia*, vol. 60, no. 6-7, pp. 2840–2854, 2012.

[18] M. N. Batista, M. C. Marinelli, and I. Alvarez-Armas, "Effect of initial microstructure on surface relief and fatigue crack initiation in aisi 410 ferritic-martensitic steel," *Fatigue & Fracture of Engineering Materials & Structures*, vol. 42, no. 1, pp. 61–68, 2019.

[19] B. Chen, J. Jiang, and F. P. Dunne, "Is stored energy density the primary meso-scale mechanistic driver for fatigue crack nucleation?" *International Journal of Plasticity*, vol. 101, pp. 213–229, 2018.

[20] Z. Chen, W. Lenthe, J. Stinville, M. Echlin, T. Pollock, and S. Daly, "High-resolution deformation mapping across large fields of view using scanning electron microscopy and digital image correlation," *Experimental Mechanics*, vol. 58, no. 9, pp. 1407–1421, 2018.

[21] M. Charpagne, J. Stinville, P. Callahan, D. Texier, Z. Chen, P. Villechaise, V. Valle, and T. Pollock, "Automated and quantitative analysis of plastic strain localization via multi-modal data recombination," *Materials Characterization*, vol. 163, p. 110245, 2020.

[22] A. Blug, D. J. Regina, S. Eckmann, M. Senn, A. Bertz, D. Carl, and C. Eberl, "Real-time GPU-based digital image correlation sensor for marker-free strain-controlled fatigue testing," *Applied Sciences*, vol. 9, no. 10, p. 2025, 2019.

[23] N. O'Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova, D. Riordan, and J. Walsh, "Deep learning vs. traditional computer vision," in *Science and Information Conference*. Springer, 2019, pp. 128–144.

[24] S. Chakraborty, R. Tomsett, R. Raghavendra, D. Harborne, M. Alzantot, F. Cerutti, M. Srivastava, A. Preece, S. Julier, R. M. Rao *et al.*, "Interpretability of deep learning models: A survey of results," in *2017 IEEE smartworld, ubiquitous intelligence & computing, advanced & trusted computed, scalable computing & communications, cloud & big data computing, Internet of people and smart city innovation*. IEEE, 2017, pp. 1–6.

[25] S. Beery, G. Van Horn, and P. Perona, "Recognition in terra incognita," in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 456–473.

[26] A. Torralba and A. A. Efros, "Unbiased look at dataset bias," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 1521–1528.

[27] P. Welinder, M. Welling, and P. Perona, "A lazy man's approach to benchmarking: Semisupervised classifier evaluation and recalibration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3262–3269.

[28] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a few examples: A survey on few-shot learning," *ACM Computing Surveys (CSUR)*, vol. 53, no. 3, pp. 1–34, 2020.

[29] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *International conference on machine learning*. PMLR, 2015, pp. 1180–1189.

[30] G. A. Gesese, R. Biswas, and H. Sack, "A comprehensive survey of knowledge graph embeddings with literals: Techniques and applications." in *DL4KG@ ESWC*, 2019, pp. 31–40.

[31] R. Stewart and S. Ermon, "Label-free supervision of neural networks with physics and domain knowledge," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[32] J.-A. Bolte, M. Kamp, A. Breuer, S. Homoceanu, P. Schlicht, F. Huger, D. Lipinski, and T. Fingscheidt, "Unsupervised domain adaptation to improve image segmentation quality both in the source and target domain," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 1404–1413.

[33] Y.-H. Tsai, W.-C. Hung, S. Schulter, K. Sohn, M.-H. Yang, and M. Chandraker, "Learning to adapt structured output space for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 06 2018, pp. 7472–7481.

[34] T. Straub, M. Berwind, T. Kennerknecht, Y. Lapusta, and C. Eberl, "Small-scale multiaxial setup for damage detection into the very high cycle fatigue regime," *Experimental Mechanics*, vol. 55, no. 7, pp. 1285–1299, 2015.

[35] M. Buck, T. Straub, and C. Eberl, "Experimental investigation of damage detection and crack initiation up to the very high cycle fatigue regime," in *Fatigue of Materials at Very High Numbers of Loading Cycles*. Springer, 2018, pp. 365–393.

[36] A. Durmaz, N. Hadzic, T. Straub, C. Eberl, and P. Gumbsch, "Efficient experimental and data-centered workflow for microstructure-based fatigue data: towards a data basis for predictive ai models," *Experimental Mechanics*, vol. 61, pp. 1489–1502, 2021.

[37] A. Thomas, A. R. Durmaz, T. Straub, and C. Eberl, "Automated Quantitative Analyses of Fatigue-Induced Surface Damage by Deep Learning," *Materials*, vol. 13, no. 15, p. 3298, 2020.

[38] A. R. Durmaz, E. Natkowski, N. Arnaudov, P. Sonnweber-Ribic, S. Weihe, S. Münstermann, C. Eberl, and P. Gumbsch, "Micromechanical fatigue experiments for validation of microstructure-sensitive fatigue simulation models," *International Journal of Fatigue*, vol. 160, p. 106824, 2022.

[39] E. Natkowski, A. R. Durmaz, P. Sonnweber-Ribic, and S. Münstermann, "Fatigue lifetime prediction with a validated micromechanical short crack model for the ferritic steel en 1.4003," *International Journal of Fatigue*, vol. 152, p. 106418, 2021.

[40] D. R. Steinmetz, T. Jäpel, B. Wietbrock, P. Eisenlohr, I. Gutierrez-Urrutia, A. Saeed-Akbari, T. Hickel, F. Roters, and D. Raabe, "Revealing the strain-hardening behavior of twinning-induced plasticity steels: Theory, simulations, experiments," *Acta Materialia*, vol. 61, no. 2, pp. 494–510, 2013.

[41] F.-D. Fischer, G. Reisner, E. Werner, K. Tanaka, G. Cailletaud, and T. Antretter, "A new view on transformation induced plasticity (TRIP)," *International Journal of Plasticity*, vol. 16, no. 7-8, pp. 723–748, 2000.

[42] F. Maresca, E. Polatidis, M. Šmíd, H. Van Swygenhoven, and W. A. Curtin, "Measurement and prediction of the transformation strain that controls ductility and toughness in advanced steels," *Acta Materialia*, vol. 200, pp. 246–255, 2020.

[43] S. Dworak, H. Rechberger, and J. Fellner, "How will tramp elements affect future steel recycling in Europe? – A dynamic material flow model for steel in the EU-28 for the period 1910 to 2050," *Resources, Conservation and Recycling*, vol. 179, p. 106072, 2022.

[44] C. M. Zener and S. Siegel, "Elasticity and anelasticity of metals." *The Journal of Physical Chemistry*, vol. 53, no. 9, pp. 1468–1468, 1949.

[45] H. Ledbetter and R. Reed, "Elastic Properties of Metals and Alloys, I. Iron, Nickel, and Iron-Nickel Alloys," *Journal of Physical and Chemical Reference Data*, vol. 2, pp. 531–618, 1973.

[46] W. H. Miller, *A treatise on crystallography*. Deighton, 1839.

[47] V. Randle and O. Engler, *Introduction to texture analysis: macrotexture, microtexture and orientation mapping*. CRC press, 2000.

[48] G. S. Rohrer, D. M. Saylor, B. El Dasher, B. L. Adams, A. D. Rollett, and P. Wynblatt, "The distribution of internal interfaces in polycrystals," *International Journal of Materials Research*, vol. 95, no. 4, pp. 197–214, 2004.

[49] D. M. Saylor, A. Morawiec, and G. S. Rohrer, "Distribution of grain boundaries in magnesia as a function of five macroscopic parameters," *Acta Materialia*, vol. 51, no. 13, pp. 3663–3674, 2003.

[50] A. Sutton, E. Banks, and A. Warwick, "The five-dimensional parameter space of grain boundaries," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 471, no. 2181, p. 20150442, 2015.

[51] V. Randle, "The coincidence site lattice and the 'sigma enigma'," *Materials Characterization*, vol. 47, no. 5, pp. 411–416, 2001.

[52] A. Stukowski, D. Cereceda, T. D. Swinburne, and J. Marian, "Thermally-activated non-Schmid glide of screw dislocations in W using atomistically-informed kinetic Monte Carlo simulations," *International Journal of Plasticity*, vol. 65, pp. 108–130, 2015.

[53] E. Schmid and W. Boas, *Kristallplastizität: mit besonderer Berücksichtigung der Metalle*. Springer-Verlag, 2013, vol. 17.

[54] W. Püschl, "Models for dislocation cross-slip in close-packed crystal structures: a critical review," *Progress in materials science*, vol. 47, no. 4, pp. 415–461, 2002.

[55] F. Kroupa, "Dislocation dipoles and dislocation loops," *Le Journal de Physique Colloques*, vol. 27, no. C3, pp. C3–154, 1966.

[56] D. Raabe, M. Sachtleber, Z. Zhao, F. Roters, and S. Zaefferer, "Micromechanical and macromechanical effects in grain scale polycrystal plasticity experimentation and simulation," *Acta materialia*, vol. 49, no. 17, pp. 3433–3441, 2001.

[57] J. C. Li, "Petch relation and grain boundary sources," *Transactions of the Metallurgical Society of AIME*, vol. 227, no. 1, p. 239, 1963.

[58] H. Mughrabi, "Cyclic slip irreversibilities and the evolution of fatigue damage," *Metallurgical and Materials Transactions B*, vol. 40, no. 4, pp. 431–453, 2009.

[59] J. Polák, "On the role of point defects in fatigue crack initiation," *Materials Science and Engineering*, vol. 92, pp. 71–80, 1987.

[60] U. Essmann, U. Gösele, and H. Mughrabi, "A model of extrusions and intrusions in fatigued metals I. point-defect production and the growth of extrusions," *Philosophical Magazine A*, vol. 44, no. 2, pp. 405–426, 1981.

[61] H. Mughrabi, K. Herz, and X. Stark, "Cyclic deformation and fatigue behaviour of α-iron mono-and polycrystals," *International Journal of fracture*, vol. 17, no. 2, pp. 193–220, 1981.

[62] D. Cereceda, M. Diehl, F. Roters, D. Raabe, J. M. Perlado, and J. Marian, "Unraveling the temperature dependence of the yield strength in single-crystal tungsten using atomistically-informed crystal plasticity calculations," *International Journal of Plasticity*, vol. 78, pp. 242–265, 2016.

[63] V. Vitek, "Core structure of screw dislocations in body-centred cubic metals: relation to symmetry and interatomic bonding," *Philosophical Magazine*, vol. 84, no. 3-5, pp. 415–428, 2004.

[64] J. Man, M. Petrenec, K. Obrtlík, and J. Polak, "AFM and TEM study of cyclic slip localization in fatigued ferritic x10cral24 stainless steel," *Acta Materialia*, vol. 52, no. 19, pp. 5551–5561, 2004.

[65] K. Pohl, P. Mayr, and E. Macherauch, "Shape and structure of persistent slip bands in iron carbon alloys," in *Defects, fracture and fatigue*. Springer, 1983, pp. 147–159.

[66] ——, "Persistent slip bands in the interior of a fatigued low carbon steel," *Scripta Metallurgica*, vol. 14, no. 11, pp. 1167–1169, 1980.

[67] A. Weidner, R. Beyer, C. Blochwitz, C. Holste, A. Schwab, and W. Tirschler, "Slip activity of persistent slip bands in polycrystalline nickel," *Materials Science and Engineering: A*, vol. 435, pp. 540–546, 2006.

[68] H. Mughrabi and C. Wüthrich, "Asymmetry of slip and shape changes during cyclic deformation of α-iron single crystals," *Philosophical Magazine*, vol. 33, no. 6, pp. 963–984, 1976.

[69] P. Franciosi, L. Le, G. Monnet, C. Kahloun, and M.-H. Chavanne, "Investigation of slip system activity in iron at room temperature by SEM and AFM in-situ tensile and compression tests of iron single crystals," *International Journal of Plasticity*, vol. 65, pp. 226–249, 2015.

[70] L. R. Lopes and J. Charlier, "Effect of grain size and intergranular stresses on the cyclic behaviour of a ferritic steel," *Materials Science and Engineering: A*, vol. 169, no. 1-2, pp. 67–77, 1993.

[71] S. Ratanaphan, D. L. Olmsted, V. V. Bulatov, E. A. Holm, A. D. Rollett, and G. S. Rohrer, "Grain boundary energies in body-centered cubic metals," *Acta Materialia*, vol. 88, pp. 346–354, 2015.

[72] M. D. Sangid, T. Ezaz, H. Sehitoglu, and I. M. Robertson, "Energy of slip transmission and nucleation at grain boundaries," *Acta materialia*, vol. 59, no. 1, pp. 283–296, 2011.

[73] A. P. Sutton, R. W. Balluffi, and A. P. Sutton, "Interfaces in crystalline materials," 1995.

[74] T. Lee, I. Robertson, and H. Birnbaum, "Prediction of slip transfer mechanisms across grain boundaries," *Scripta metallurgica*, vol. 23, no. 5, pp. 799–803, 1989.

[75] ——, "TEM in situ deformation study of the interaction of lattice dislocations with grain boundaries in metals," *Philosophical Magazine A*, vol. 62, no. 1, pp. 131–153, 1990.

[76] ——, "An in situ transmission electron microscope deformation study of the slip transfer mechanisms in metals," *Metallurgical Transactions A*, vol. 21, no. 9, pp. 2437–2447, 1990.

[77] E. Bayerschen, A. McBride, B. Reddy, and T. Böhlke, "Review on slip transmission criteria in experiments and crystal plasticity models," *Journal of materials science*, vol. 51, no. 5, pp. 2243–2258, 2016.

[78] Z. Shen, R. Wagoner, and W. Clark, "Dislocation pile-up and grain boundary interactions in 304 stainless steel," *Scripta metallurgica*, vol. 20, no. 6, pp. 921–926, 1986.

[79] W. Soer and J. T. M. De Hosson, "Detection of grain-boundary resistance to slip transfer using nanoindentation," *Materials Letters*, vol. 59, no. 24-25, pp. 3192–3195, 2005.

[80] T. Britton, D. Randman, and A. Wilkinson, "Nanoindentation study of slip transfer phenomenon at grain boundaries," *Journal of Materials Research*, vol. 24, no. 3, pp. 607–615, 2009.

[81] D. Raabe, M. Herbig, S. Sandlöbes, Y. Li, D. Tytko, M. Kuzmina, D. Ponge, and P.-P. Choi, "Grain boundary segregation engineering in metallic alloys: A pathway to the design of interfaces," *Current Opinion in Solid State and Materials Science*, vol. 18, no. 4, pp. 253–261, 2014.

[82] S. Nishijima and K. Kanazawa, "Stepwise sn curve and fish-eye failure in gigacycle fatigue," *Fatigue & fracture of engineering materials & structures (Print)*, vol. 22, no. 7, pp. 601–607, 1999.

[83] M. Klesnil, "The degree of damage at the French curve and at the fatigue limit during oscillating bend loading," *Metal Treat. Drop Forging*, vol. 55, pp. 55–63, 1965.

[84] D. McDowell and F. Dunne, "Microstructure-sensitive computational modeling of fatigue crack formation," *International journal of fatigue*, vol. 32, no. 9, pp. 1521–1542, 2010.

[85] H. Mughrabi, "Microstructural mechanisms of cyclic deformation, fatigue crack initiation and early crack growth," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 373, no. 2038, p. 20140132, 2015.

[86] P. Forsyth, "A two stage process of fatigue crack growth," 1997.

[87] Y. Furuya, "Specimen size effects on gigacycle fatigue properties of high-strength steel under ultrasonic fatigue testing," *Scripta Materialia*, vol. 58, no. 11, pp. 1014–1017, 2008.

[88] S. Burger, C. Eberl, A. Siegel, A. Ludwig, and O. Kraft, "A novel high-throughput fatigue testing method for metallic thin films," *Science and technology of advanced materials*, 2011.

[89] Y. H. Zhang and L. Edwards, "On the blocking effect of grain boundaries on small crystallographic fatigue crack growth," *Materials Science and Engineering: A*, vol. 188, no. 1-2, pp. 121–132, 1994.

[90] N. Thompson, N. Wadsworth, and N. Louat, "The origin of fatigue fracture in copper," *Philosophical Magazine*, vol. 1, no. 2, pp. 113–126, 1956.

[91] B. L. Boyce, J. R. Michael, and P. G. Kotula, "Fatigue of metallic microdevices and the role of fatigue-induced surface oxides," *Acta Materialia*, vol. 52, no. 6, pp. 1609–1619, 2004.

[92] G. Ebi and A. McEvily, "Effect of processing on the high temperature low cycle fatigue properties of modified 9Cr-1Mo ferritic steel," *Fatigue & Fracture of Engineering Materials & Structures*, vol. 7, no. 4, pp. 299–314, 1984.

[93] S. Lynch, "Hydrogen embrittlement phenomena and mechanisms," *Corrosion reviews*, vol. 30, no. 3-4, pp. 105–123, 2012.

[94] E. Hall, "The deformation and ageing of mild steel: III discussion of results," *Proceedings of the Physical Society. Section B*, vol. 64, no. 9, p. 747, 1951.

[95] N. J. Petch, "The cleavage strength of polycrystals," *Journal of the Iron and Steel Institute*, vol. 174, pp. 25–28, 1953.

[96] D. Raabe, M. Herbig, S. Sandlöbes, Y. Li, D. Tytko, M. Kuzmina, D. Ponge, and P.-P. Choi, "Grain boundary segregation engineering in metallic alloys: A pathway to the design of interfaces," *Current Opinion in Solid State and Materials Science*, vol. 18, no. 4, pp. 253–261, 2014.

[97] J. Hochhalter, D. Littlewood, R. Christ, M. Veilleux, J. Bozek, A. Ingraffea, and A. Maniatty, "A geometric approach to modeling microstructurally small fatigue crack formation: II. physically based modeling of microstructure-dependent slip localization and actuation of the crack nucleation mechanism in AA 7075-T651," *Modelling and Simulation in Materials Science and Engineering*, vol. 18, no. 4, p. 045004, 2010.

[98] Y. Cheng, M. Mrovec, and P. Gumbsch, "Atomistic simulations of interactions between the 1/2< 111> edge dislocation and symmetric tilt grain boundaries in tungsten," *Philosophical Magazine*, vol. 88, no. 4, pp. 547–560, 2008.

[99] D. E. Spearot and M. D. Sangid, "Insights on slip transmission at grain boundaries from atomistic simulations," *Current Opinion in Solid State and Materials Science*, vol. 18, no. 4, pp. 188–195, 2014.

[100] J. J. Möller and E. Bitzek, "On the influence of crack front curvature on the fracture behavior of nanoscale cracks," *Engineering Fracture Mechanics*, vol. 150, pp. 197–208, 2015.

[101] L.-X. Cao and C.-Y. Wang, "Atomistic simulation for configuration evolution and energetic calculation of crack in body-centered-cubic iron," *Journal of materials research*, vol. 21, no. 10, pp. 2542–2549, 2006.

[102] Y. Cui, G. Po, P. Srivastava, K. Jiang, V. Gupta, and N. Ghoniem, "The role of slow screw dislocations in controlling fast strain avalanche dynamics in body-centered cubic metals," *International journal of plasticity*, vol. 124, pp. 117–132, 2020.

[103] J. Zghal, H. Gmati, C. Mareau, and F. Morel, "A crystal plasticity based approach for the modelling of high cycle fatigue damage in metallic materials," *International Journal of Damage Mechanics*, vol. 25, no. 5, pp. 611–628, 2016.

[104] H. Proudhon, J. Li, F. Wang, A. Roos, V. Chiaruttini, and S. Forest, "3D simulation of short fatigue crack propagation by finite element crystal plasticity and remeshing," *International Journal of fatigue*, vol. 82, pp. 238–246, 2016.

[105] K. Pierson, A. Rahman, and A. D. Spear, "Predicting microstructure-sensitive fatigue-crack path in 3D using a machine learning framework," *Jom*, vol. 71, no. 8, pp. 2680–2694, 2019.

[106] A. Rovinelli, M. D. Sangid, H. Proudhon, and W. Ludwig, "Using machine learning and a data-driven approach to identify the small fatigue crack driving force in polycrystalline materials," *npj Computational Materials*, vol. 4, no. 1, pp. 1–10, 2018.

[107] K. Tanaka and T. Mura, "A Dislocation Model for Fatigue Crack Initiation," *J. appl. Mech*, vol. 48, no. 1, pp. 97–103, 1980.

[108] T. Mura, "A theory of fatigue crack initiation," *Materials Science and Engineering: A*, vol. 176, no. 1-2, pp. 61–70, 1994.

[109] K. S. Chan, "A microstructure-based fatigue-crack-initiation model," *Metallurgical and Materials Transactions A*, vol. 34, no. 1, pp. 43–58, 2003.

[110] T. Mura, "A theory of fatigue crack initiation," *Materials Science and Engineering A*, vol. 176, no. 1-2, pp. 61–70, 1994.

[111] A. Saxena and S. D. Antolovich, "Low cycle fatigue, fatigue crack propagation and substructures in a series of polycrystalline Cu-Al alloys," *Metallurgical Transactions A*, vol. 6, no. 9, pp. 1809–1828, 1975.

[112] G. Venkataraman, Y.-W. Chung, and T. Mura, "Application of minimum energy formalism in a multiple slip band model for fatigue—II. crack nucleation and derivation of a generalised Coffin-Manson law," *Acta metallurgica et materialia*, vol. 39, no. 11, pp. 2631–2638, 1991.

[113] A. M. Korsunsky, D. Dini, F. P. Dunne, and M. J. Walsh, "Comparative assessment of dissipated energy and other fatigue criteria," *International Journal of Fatigue*, vol. 29, no. 9-11, pp. 1990–1995, 2007.

[114] A. Fatemi and D. F. Socie, "A critical plane approach to multiaxial fatigue damage including out-of-phase loading," *Fatigue & Fracture of Engineering Materials & Structures*, vol. 11, no. 3, pp. 149–165, 1988.

[115] V. V. Wan, D. W. Maclachlan, and F. P. Dunne, "A stored energy criterion for fatigue crack nucleation in polycrystals," *International Journal of Fatigue*, vol. 68, pp. 90–102, 2014.

[116] N. K. Lassen, D. J. Jensen, and K. Conradsen, "Image processing procedures for analysis of electron back scattering patterns," *Scanning microscopy*, vol. 6, no. 1, pp. 115–121, 1992.

[117] T. B. Britton, J. Jiang, P. S. Karamched, and A. J. Wilkinson, "Probing deformation and revealing microstructural mechanisms with cross-correlation-based, high-resolution electron backscatter diffraction," *Jom*, vol. 65, no. 9, pp. 1245–1253, 2013.

[118] D. Jha, S. Singh, R. Al-Bahrani, W.-k. Liao, A. Choudhary, M. De Graef, and A. Agrawal, "Extracting grain orientations from EBSD patterns of polycrystalline materials using convolutional neural networks," *Microscopy and Microanalysis*, vol. 24, no. 5, pp. 497–502, 2018.

[119] S. I. Wright and M. M. Nowell, "EBSD image quality mapping," *Microscopy and microanalysis*, vol. 12, no. 1, p. 72, 2006.

[120] D. P. Field, "Recent advances in the application of orientation imaging," *Ultramicroscopy*, vol. 67, no. 1-4, pp. 1–9, 1997.

[121] S. I. Wright, M. M. Nowell, S. P. Lindeman, P. P. Camus, M. De Graef, and M. A. Jackson, "Introduction and comparison of new EBSD post-processing methodologies," *Ultramicroscopy*, vol. 159, pp. 81–94, 2015.

[122] J.-Y. Kang, S.-J. Park, and M.-B. Moon, "Phase analysis on dual-phase steel using band slope of electron backscatter diffraction pattern," *Microscopy and Microanalysis*, vol. 19, no. S5, pp. 13–16, 2013.

[123] L. Ryde, "Application of EBSD to analysis of microstructures in commercial steels," *Materials Science and Technology*, vol. 22, no. 11, pp. 1297–1306, 2006.

[124] S. Zaefferer, P. Romano, and F. Friedel, "EBSD as a tool to identify and quantify bainite and ferrite in low-alloyed Al-TRIP steels," *Journal of microscopy*, vol. 230, no. 3, pp. 499–508, 2008.

[125] S. J. Randolph, J. Filevich, A. Botman, R. Gannon, C. Rue, and M. Straw, "In situ femtosecond pulse laser ablation for large volume 3D analysis in scanning electron microscope systems," *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena*, vol. 36, no. 6, p. 06JB01, 2018.

[126] D. An, T. Griffiths, P. Konijnenberg, S. Mandal, Z. Wang, and S. Zaefferer, "Correlating the five parameter grain boundary character distribution and the intergranular corrosion behaviour of a stainless steel using 3D orientation microscopy based on mechanical polishing serial sectioning," *Acta Materialia*, vol. 156, pp. 297–309, 2018.

[127] M. Ge, F. Su, Z. Zhao, and D. Su, "Deep learning analysis on microscopic imaging in materials science," *Materials Today Nano*, vol. 11, p. 100087, 2020.

[128] A. Agrawal and A. Choudhary, "Deep materials informatics: Applications of deep learning in materials science," *MRS Communications*, vol. 9, no. 3, pp. 779–792, 2019.

[129] K. Choudhary, B. DeCost, C. Chen, A. Jain, F. Tavazza, R. Cohn, C. WooPark, A. Choudhary, A. Agrawal, S. J. Billinge *et al.*, "Recent advances and applications of deep learning methods in materials science," *arXiv preprint arXiv:2110.14820*, 2021.

[130] S. M. Azimi, D. Britz, M. Engstler, M. Fritz, and F. Mücklich, "Advanced steel microstructural classification by deep learning methods," *Scientific reports*, vol. 8, no. 1, pp. 1–14, 2018.

[131] J. Gola, J. Webel, D. Britz, A. Guitar, T. Staudt, M. Winter, and F. Mücklich, "Objective microstructure classification by support vector machine (SVM) using a combination of morphological parameters and textural features for low carbon steels," *Computational Materials Science*, vol. 160, pp. 186–196, 2019.

[132] B. Ma, X. Wei, C. Liu, X. Ban, H. Huang, H. Wang, W. Xue, S. Wu, M. Gao, Q. Shen *et al.*, "Data augmentation in microscopic images for material data mining," *npj Computational Materials*, vol. 6, no. 1, pp. 1–9, 2020.

[133] B. L. DeCost, B. Lei, T. Francis, and E. A. Holm, "High throughput quantitative metallography for complex microstructures using deep learning: A case study in ultrahigh carbon steel," *Microscopy and Microanalysis*, vol. 25, no. 1, pp. 21–29, 2019.

[134] G. Roberts, S. Y. Haile, R. Sainju, D. J. Edwards, B. Hutchinson, and Y. Zhu, "Deep learning for semantic segmentation of defects in advanced stem images of steels," *Scientific reports*, vol. 9, no. 1, pp. 1–12, 2019.

[135] Y.-J. Cha, W. Choi, and O. Büyüköztürk, "Deep learning-based crack damage detection using convolutional neural networks," *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 5, pp. 361–378, 2017.

[136] C. Chen, Y. Zuo, W. Ye, X. Li, and S. P. Ong, "Learning properties of ordered and disordered materials from multi-fidelity data," *Nature Computational Science*, vol. 1, no. 1, pp. 46–53, 2021.

[137] A. D. Orme, I. Chelladurai, T. M. Rampton, D. T. Fullwood, A. Khosravani, M. P. Miles, and R. K. Mishra, "Insights into twinning in Mg AZ31: A combined EBSD and machine learning study," *Computational Materials Science*, vol. 124, pp. 353–363, 2016.

[138] R. Sharma, I. Chelladurai, A. D. Orme, M. P. Miles, C. Giraud-Carrier, and D. T. Fullwood, "A step towards intelligent EBSD microscopy: machine-learning prediction of twin activity in MgAZ31," *Journal of Microscopy*, vol. 272, no. 1, pp. 67–78, 2018.

[139] A. Mangal and E. A. Holm, "Applied machine learning to predict stress hotspots i: Face centered cubic materials," *International Journal of Plasticity*, vol. 111, pp. 122–134, 2018.

[140] J. R. Mianroodi, N. H Siboni, and D. Raabe, "Teaching solid mechanics to artificial intelligence—a fast solver for heterogeneous materials," *npj Computational Materials*, vol. 7, no. 1, pp. 1–10, 2021.

[141] N. N. Vlassis, R. Ma, and W. Sun, "Geometric deep learning for computational mechanics part i: Anisotropic hyperelasticity," *Computer Methods in Applied Mechanics and Engineering*, vol. 371, p. 113299, 2020.

[142] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[143] K. He, R. B. Girshick, and P. Dollár, "Rethinking imagenet pre-training," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 4917–4926.

[144] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[145] A. Newell and J. Deng, "How useful is self-supervised pretraining for visual tasks?" in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 7345–7354.

[146] M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, and A. Joulin, "Emerging properties in self-supervised vision transformers," *arXiv preprint arXiv:2104.14294*, 2021.

[147] P. Goyal, M. Caron, B. Lefaudeux, M. Xu, P. Wang, V. Pai, M. Singh, V. Liptchinsky, I. Misra, A. Joulin *et al.*, "Self-supervised pretraining of visual features in the wild," *arXiv preprint arXiv:2103.01988*, 2021.

[148] A. Saeed, T. Ozcelebi, and J. Lukkien, "Multi-task self-supervised learning for human activity detection," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 2, pp. 1–30, 2019.

[149] T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," *Artificial intelligence*, vol. 267, pp. 1–38, 2019.

[150] Z. C. Lipton, "The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery." *Queue*, vol. 16, no. 3, pp. 31–57, 2018.

[151] M. Robnik-Šikonja and M. Bohanec, "Perturbation-based explanations of prediction models," *Human and Machine Learning: Visible, Explainable, Trustworthy and Transparent*, pp. 159–175, 2018.

[152] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[153] D. Bau, B. Zhou, A. Khosla, A. Oliva, and A. Torralba, "Network dissection: Quantifying interpretability of deep visual representations," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 6541–6549.

[154] L. Deng, "A tutorial survey of architectures, algorithms, and applications for deep learning," *APSIPA transactions on Signal and Information Processing*, vol. 3, 2014.

[155] N. Segev, M. Harel, S. Mannor, K. Crammer, and R. El-Yaniv, "Learn on source, refine on target: A model transfer learning framework with random forests," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 9, pp. 1811–1824, 2016.

[156] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.

[157] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *Advances in Computer Vision and Pattern Recognition*, vol. 17, no. 9783319583464, pp. 189–209, 2017.

[158] F. Pérez-Cruz, "Kullback-Leibler divergence estimation of continuous distributions," pp. 1666–1670, 2008.

[159] R. E. Schapire, "The strength of weak learnability," *Machine learning*, vol. 5, no. 2, pp. 197–227, 1990.

[160] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[161] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[162] J. Ali, R. Khan, N. Ahmad, and I. Maqsood, "Random forests and decision trees," *International Journal of Computer Science Issues (IJCSI)*, vol. 9, no. 5, p. 272, 2012.

[163] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on knowledge and data engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.

[164] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[165] C. Chen, A. Liaw, and L. Breiman, "Using Random Forest to Learn Imbalanced Data," *Discovery*, no. 1999, pp. 1–12.

[166] I.-l. Lemaître, "A python toolbox to tackle the curse of imbalanced datasets in machine learning," *J. Mach. Learn. Res*, no. 18, p. 1.

[167] B. Hosseini and B. Hammer, "Interpretable discriminative dimensionality reduction and feature selection on the manifold," *arXiv preprint arXiv:1909.09218*, 2019.

[168] G. Louppe, L. Wehenkel, A. Sutera, and P. Geurts, "Understanding variable importances in forests of randomized trees," *Advances in neural information processing systems*, vol. 26, 2013.

[169] Y. L. Pavlov, "Random forests," in *Random Forests*. De Gruyter, 2019.

[170] C. Strobl, A.-L. Boulesteix, A. Zeileis, and T. Hothorn, "Bias in random forest variable importance measures: Illustrations, sources and a solution," *BMC bioinformatics*, vol. 8, no. 1, pp. 1–21, 2007.

[171] C. Strobl, A.-L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis, "Conditional variable importance for random forests," *BMC bioinformatics*, vol. 9, no. 1, pp. 1–11, 2008.

[172] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[173] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *The Annals of statistics*, vol. 32, no. 2, pp. 407–499, 2004.

[174] L. McCann and R. E. Welsch, "Robust variable selection using least angle regression and elemental set sampling," *Computational Statistics & Data Analysis*, vol. 52, no. 1, pp. 249–257, 2007.

[175] S. Weisberg, "Discussion of "Least angle regression"by Efron et al." *arXiv preprint math/0406473*, 2004.

[176] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the royal statistical society: series B (statistical methodology)*, vol. 67, no. 2, pp. 301–320, 2005.

[177] F. R. Bach, "Bolasso: model consistent lasso estimation through the bootstrap," pp. 33–40, 2008.

[178] H. Zare, G. Haffari, A. Gupta, and R. R. Brinkman, "Scoring relevancy of features based on combinatorial analysis of lasso with application to lymphoma diagnosis," in *BMC genomics*, vol. 14, no. 1. Springer, 2013, pp. 1–9.

[179] A. Mangal and E. A. Holm, "A comparative study of feature selection methods for stress hotspot classification in materials," *Integrating Materials and Manufacturing Innovation*, vol. 7, no. 3, pp. 87–95, 2018.

[180] D. Li, J. Hu, C. Wang, X. Li, Q. She, L. Zhu, T. Zhang, and Q. Chen, "Involution: Inverting the inherence of convolution for visual recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 12 321–12 330.

[181] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

[182] I. Freeman, L. Roese-Koerner, and A. Kummert, "Effnet: An efficient structure for convolutional neural networks," pp. 6–10, 2018.

[183] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.

[184] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," *arXiv preprint arXiv:1603.07285*, 2016.

[185] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Icml*, 2010.

[186] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," pp. 315–323, 2011.

[187] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," pp. 448–456, 2015.

[188] H. Li, Z. Xu, G. Taylor, C. Studer, and T. Goldstein, "Visualizing the loss landscape of neural nets," *Advances in neural information processing systems*, vol. 31, 2018.

[189] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How does batch normalization help optimization?" *Advances in neural information processing systems*, vol. 31, 2018.

[190] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," *arXiv preprint arXiv:1607.06450*, 2016.

[191] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *arXiv preprint arXiv:1607.08022*, 2016.

[192] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," in *International conference on artificial neural networks*. Springer, 2010, pp. 92–101.

[193] W. Luo, Y. Li, R. Urtasun, and R. Zemel, "Understanding the effective receptive field in deep convolutional neural networks," *Advances in Neural Information Processing Systems*, no. Nips, pp. 4905–4913, 2016.

[194] Z. Wojna, V. Ferrari, S. Guadarrama, N. Silberman, L.-C. Chen, A. Fathi, and J. Uijlings, "The devil is in the decoder," *International Journal of Computer Vision*, vol. 127, no. 11, pp. 1694–1706, 2019.

[195] P. Thévenaz, T. Blu, and M. Unser, "Image interpolation and resampling," *Handbook of Medical Image Processing and Analysis*, pp. 465–493, 2009.

[196] A. Odena, V. Dumoulin, and C. Olah, "Deconvolution and checkerboard artifacts," *Distill*, vol. 1, no. 10, p. e3, 2016.

[197] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," pp. 234–241, 2015.

[198] M. Yi-de, L. Qing, and Q. Zhi-Bai, "Automated image segmentation using improved pcnn model based on cross-entropy," pp. 743–746, 2004.

[199] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," pp. 3431–3440, 2015.

[200] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," pp. 2980–2988, 2017.

[201] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[202] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.

[203] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," pp. 424–432, 2016.

[204] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[205] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation," pp. 11–19, 2017.

[206] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.

[207] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[208] H. Wang, Y. Zhu, B. Green, H. Adam, A. Yuille, and L.-C. Chen, "Axial-deeplab: Stand-alone axial-attention for panoptic segmentation," pp. 108–126, 2020.

[209] H. Guo, H. Liu, C. Wu, W. Zhi, Y. Xiao, and W. She, "Logistic discrimination based on G-mean and F-measure for imbalanced problem," *Journal of Intelligent & Fuzzy Systems*, vol. 31, no. 3, pp. 1155–1166, 2016.

[210] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets," *PloS one*, vol. 10, no. 3, p. e0118432, 2015.

[211] T. Straub, M. Berwind, T. Kennerknecht, Y. Lapusta, and C. Eberl, "Small-scale multiaxial setup for damage detection into the very high cycle fatigue regime," *Experimental Mechanics*, vol. 55, no. 7, pp. 1285–1299, 2015.

[212] F. Povolo, A. Canzian, and E. Favret, "Unidirectional laser oblique illumination (ULOI): A comparison with partial dark field illumination (PDFI) in the optical microscopy of metals," *Optics and Laser Technology*, vol. 29, no. 2, pp. 67–73, 1997.

[213] F. Jünger, P. v. Olshausen, and A. Rohrbach, "Fast, label-free super-resolution live-cell imaging using rotating coherent scattering (ROCS) microscopy," *Scientific reports*, vol. 6, no. 1, pp. 1–11, 2016.

[214] H. Bennett and J. Porteus, "Relation between surface roughness and specular reflectance at normal incidence," *JOSA*, vol. 51, no. 2, pp. 123–129, 1961.

[215] J. Chalfoun, M. Majurski, T. Blattner, K. Bhadriraju, W. Keyrouz, P. Bajcsy, and M. Brady, "MIST: accurate and scalable microscopy image stitching tool with stage modeling and error minimization," *Scientific reports*, vol. 7, no. 1, pp. 1–10, 2017.

[216] S. Preibisch, S. Saalfeld, and P. Tomancak, "Globally optimal stitching of tiled 3D microscopic image acquisitions," *Bioinformatics*, vol. 25, no. 11, pp. 1463–1465, 2009.

[217] S. Maraghechi, J. Hoefnagels, R. Peerlings, O. Rokoš, and M. Geers, "Correction of scanning electron microscope imaging artifacts in a novel digital image correlation framework," *Experimental mechanics*, vol. 59, no. 4, pp. 489–516, 2019.

[218] Y. Lu, X. Zhang, and H. Li, "A simplified focusing and astigmatism correction method for a scanning electron microscope," *AIP Advances*, vol. 8, no. 1, p. 015124, 2018.

[219] T. Straub, *Experimental investigation of crack initiation in face-centered cubic materials in the high and very high cycle fatigue regime*. KIT Scientific Publishing, 2016, vol. 55.

[220] E. Polatidis, W.-N. Hsu, M. Šmíd, and H. Van Swygenhoven, "A high resolution digital image correlation study under multiaxial loading," *Experimental Mechanics*, vol. 59, no. 3, pp. 309–317, 2019.

[221] J. Man, K. Obrtlik, and J. Polak, "Extrusions and intrusions in fatigued metals. part 1. state of the art and history," *Philosophical Magazine*, vol. 89, no. 16, pp. 1295–1336, 2009.

[222] G. Seidametova, J.-B. Vogt, and I. P. Serre, "The early stage of fatigue crack initiation in a 12% Cr martensitic steel," *International Journal of Fatigue*, vol. 106, pp. 38–48, 2018.

[223] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellyn, and W. Eubank, "Nonrigid multimodality image registration," in *Medical imaging 2001: image processing*, vol. 4322. Spie, 2001, pp. 1609–1620.

[224] E. H. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. Ogden, "Pyramid methods in image processing," *RCA engineer*, vol. 29, no. 6, pp. 33–41, 1984.

[225] C. Smith, "Two microscopes are better than one," *Nature*, vol. 492, no. 7428, pp. 293–297, 2012.

[226] O. Gomes, "Multimodal microscopy: automatic registration of images from optical microscopy and SEM using Fiji/ImageJ," *X Stermat*, pp. 60–63, 2016.

[227] A. A. Goshtasby, *2-D and 3-D image registration: for medical, remote sensing, and industrial applications*. John Wiley & Sons, 2005.

[228] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and vision computing*, vol. 21, no. 11, pp. 977–1000, 2003.

[229] O. Gomes and S. Paciornik, "Multimodal microscopy for ore characterization," in *Scanning Electron Microscopy*. IntechOpen, 2012.

[230] D. Britz, J. Webel, J. Gola, and F. Mücklich, "A correlative approach to capture and quantify substructures by means of image registration," *Practical Metallography*, vol. 54, no. 10, pp. 685–696, 2017.

[231] D. G. Lowe, "Object recognition from local scale-invariant features," vol. 2, pp. 1150–1157, 1999.

[232] S. Rahunathan, D. Stredney, P. Schmalbrock, and B. D. Clymer, "Image registration using rigid registration and maximization of mutual information," in *13th Annu. Med. Meets Virtual Reality Conf*, 2005.

[233] S. Nandish, G. Prabhu, and K. V. Rajagopal, "Multiresolution image registration for multimodal brain images and fusion for better neurosurgical planning," *biomedical journal*, vol. 40, no. 6, pp. 329–338, 2017.

[234] G. Nolze, "Geometrically caused image distortion effects and their influence on interpretation of EBSD measurements," *Materials Science and Technology*, vol. 22, no. 11, pp. 1343–1351, 2006.

[235] F. Ram, S. Zaefferer, T. Jäpel, and D. Raabe, “Error analysis of the crystal orientations and disorientations obtained by the classical electron backscatter diffraction technique,” *Journal of Applied Crystallography*, vol. 48, no. 3, pp. 797–813, 2015.

[236] G. Nolze, “Image distortions in SEM and their influences on EBSD measurements,” *Ultramicroscopy*, vol. 107, no. 2-3, pp. 172–183, 2007.

[237] J. P. Kapur and D. P. Casasent, “Geometric correction of SEM images,” in *Hybrid Image and Signal Processing VII*, vol. 4044. International Society for Optics and Photonics, 2000, pp. 165–176.

[238] C. Wu, B. Adams, C. Bauer, D. Casasent, A. Morawiec, S. Ozdemir, and A. Talukder, “Mapping the mesoscale interface structure in polycrystalline materials,” *Ultramicroscopy*, vol. 93, no. 2, pp. 99–109, 2002.

[239] C. O. S. Sorzano, P. Thévenaz, and M. Unser, “Elastic registration of biological images using vector-spline regularization,” *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 4, pp. 652–663, 2005.

[240] R. Szeliski and H.-Y. Shum, “Motion estimation with quadtree splines,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 18, no. 12, pp. 1199–1210, 1996.

[241] C. Sanchez, G. Cristóbal, G. Bueno, S. Blanco, M. Borrego-Ramos, A. Olenici, A. Pedraza, and J. Ruiz-Santaquiteria, “Oblique illumination in microscopy: A quantitative evaluation,” *Micron*, vol. 105, pp. 47–54, 2018.

[242] R. Sugimoto, R. Maruyama, Y. Tamada, H. Arimoto, and W. Watanabe, “Contrast enhancement by oblique illumination microscopy with an LED array,” *Optik*, vol. 183, pp. 92–98, 2019.

[243] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[244] W. Kim and C. Jung, “Illumination-invariant background subtraction: Comparative review, models, and prospects,” *IEEE Access*, vol. 5, pp. 8369–8384, 2017.

[245] K. Maag, M. Rottmann, S. Varghese, F. Hüger, P. Schlicht, and H. Gottschalk, “Improving video instance segmentation by light-weight temporal uncertainty estimates,” in *2021 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2021, pp. 1–8.

[246] N. Gates and A. Fatemi, “Friction and roughness induced closure effects on shear-mode crack growth and branching mechanisms,” *International Journal of Fatigue*, vol. 92, pp. 442–458, 2016.

[247] M. Toyosada, K. Gotoh, and T. Niwa, “Fatigue crack propagation for a through thickness crack: a crack propagation law considering cyclic plasticity near the crack tip,” *International Journal of Fatigue*, vol. 26, no. 9, pp. 983–992, 2004.

[248] R. H. Chan, C.-W. Ho, and M. Nikolova, “Salt-and-pepper noise removal by median-type noise detectors and detail-preserving regularization,” *IEEE Transactions on image processing*, vol. 14, no. 10, pp. 1479–1485, 2005.

[249] F. E. Nicodemus, “Directional reflectance and emissivity of an opaque surface,” *Applied optics*, vol. 4, no. 7, pp. 767–775, 1965.

[250] T. F. Coleman and Y. Li, “An interior trust region approach for nonlinear minimization subject to bounds,” *SIAM Journal on optimization*, vol. 6, no. 2, pp. 418–445, 1996.

[251] S. Burger, *High Cycle Fatigue of Al and Cu Thin Films by a Novel High-Throughput Method*. KIT Scientific Publishing, 2013, vol. 23.

[252] N. Arnaudov, *Micromechanical simulation of fatigue crack initiation under hydrogen influence*. Stuttgart: Materialprüfungsanstalt (MPA), Universität Stuttgart, 2021.

[253] C. Mattheck, R. Kappel, and A. Sauer, "Shape optimization the easy way: the 'method of tensile triangles'," *International Journal of Design & Nature and Ecodynamics*, vol. 2, no. 4, pp. 301–309, 2007.

[254] E.-S. Lee, "Machining characteristics of the electropolishing of stainless steel (STS316L)," *The International Journal of Advanced Manufacturing Technology*, vol. 16, no. 8, pp. 591–599, 2000.

[255] A. Bhuyan, B. Gregory, H. Lei, S. Y. Yee, and Y. B. Gianchandani, "Pulse and dc electropolishing of stainless steel for stents and other devices," in *SENSORS, 2005 IEEE*. IEEE, 2005, pp. 314–317.

[256] G. Yang, B. Wang, K. Tawfiq, H. Wei, S. Zhou, and G. Chen, "Electropolishing of surfaces: theory and applications," *Surface Engineering*, vol. 33, no. 2, pp. 149–166, 2017.

[257] A. D. Kammers and S. Daly, "Self-assembled nanoparticle surface patterning for improved digital image correlation in a scanning electron microscope," *Experimental Mechanics*, vol. 53, no. 8, pp. 1333–1341, 2013.

[258] N. De Vincentis and D. Field, "Factors affecting confidence index in EBSD analysis," *Ultramicroscopy*, vol. 225, p. 113269, 2021.

[259] F. Bachmann, R. Hielscher, and H. Schaeben, "Grain detection from 2D and 3D EBSD data —- specification of the MTEX algorithm," *Ultramicroscopy*, vol. 111, no. 12, pp. 1720–1733, 2011.

[260] ——, "Texture analysis with MTEX – free and open source software toolbox," in *Solid state phenomena*, vol. 160. Trans Tech Publ, 2010, pp. 63–68.

[261] R. Bergmann, R. H. Chan, R. Hielscher, J. Persch, and G. Steidl, "Restoration of manifold-valued images by half-quadratic minimization," *arXiv preprint arXiv:1505.07029*, 2015.

[262] P. Lehto, "Adaptive domain misorientation approach for the EBSD measurement of deformation induced dislocation sub-structures," *Ultramicroscopy*, vol. 222, p. 113203, 2021.

[263] P. Klapetek, D. Nečas, A. Campbellová, A. Yacoot, and L. Koenders, "Methods for determining and processing 3D errors and uncertainties for AFM data analysis," *Measurement Science and Technology*, vol. 22, no. 2, p. 025501, 2011.

[264] F. Gołek, P. Mazur, Z. Ryszka, and S. Zuber, "AFM image artifacts," *Applied surface science*, vol. 304, pp. 11–19, 2014.

[265] B. W. Erickson, S. Coquoz, J. D. Adams, D. J. Burns, and G. E. Fantner, "Large-scale analysis of high-speed atomic force microscopy data sets using adaptive image processing," *Beilstein journal of nanotechnology*, vol. 3, no. 1, pp. 747–758, 2012.

[266] A. Lin, "Binary search algorithm," *WikiJournal of Science*, vol. 2, no. 1, pp. 1–13, 2019.

[267] H. T. Banks and D. Inman, "On damping mechanisms in beams," 1991.

[268] M. Arjovsky, "Out of distribution generalization in machine learning," Ph.D. dissertation, New York University, 2020.

[269] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of big data*, vol. 6, no. 1, pp. 1–48, 2019.

[270] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: fast and flexible image augmentations," *Information*, vol. 11, no. 2, p. 125, 2020.

[271] R. Liaw, E. Liang, R. Nishihara, P. Moritz, J. E. Gonzalez, and I. Stoica, "Tune: A Research Platform for Distributed Model Selection and Training," no. 2012, 2018. [Online]. Available: http://arxiv.org/abs/1807.05118

[272] J. Ma, "Segmentation loss odyssey," *arXiv preprint arXiv:2005.13449*, 2020.

[273] G. Bonny, D. Terentyev, and L. Malerba, "On the α–α' miscibility gap of Fe–Cr alloys," *Scripta materialia*, vol. 59, no. 11, pp. 1193–1196, 2008.

[274] G. Bonny, R. Pasianot, E. Zhurkin, and M. Hou, "Determination of the phase diagram from interatomic potentials: The iron–chromium case," *Computational materials science*, vol. 50, no. 7, pp. 2216–2220, 2011.

[275] J. Farrar, "Inclusions and susceptibility to lamellar tearing of welded structural steels," *Welding J*, vol. 35, pp. 321–331, 1974.

[276] X. Xiao, G. Liu, B. Hu, J. Wang, and W. Ma, "Microstructure stability of v and ta microalloyed 12% Cr reduced activation ferrite/martensite steel during long-term aging at 650° C," *Journal of Materials Science & Technology*, vol. 31, no. 3, pp. 311–319, 2015.

[277] M. Meyers, O. Vöhringer, and V. Lubarda, "The onset of twinning in metals: a constitutive description," *Acta materialia*, vol. 49, no. 19, pp. 4025–4039, 2001.

[278] T. Taniguchi, Y. Kaneko, and S. Hashimoto, "Fatigue lives of a ferritic stainless steel containing deformation twins," in *Key Engineering Materials*, vol. 353. Trans Tech Publ, 2007, pp. 283–286.

[279] S. Pal, M. Alam, S. Maloy, D. Hoelzer, and G. Odette, "Texture evolution and microcracking mechanisms in as-extruded and cross-rolled conditions of a 14YWT nanostructured ferritic alloy," *Acta Materialia*, vol. 152, pp. 338–357, 2018.

[280] J. Mackenzie, "Second paper on statistics associated with the random disorientation of cubes," *Biometrika*, vol. 45, no. 1-2, pp. 229–240, 1958.

[281] I. Muto, Y. Izumiyama, and N. Hara, "Microelectrochemical measurements of dissolution of mns inclusions and morphological observation of metastable and stable pitting on stainless steel," *Journal of the Electrochemical Society*, vol. 154, no. 8, p. C439, 2007.

[282] Y. Kaneko, S. Hashimoto, and S. Miura, "Cracking along mechanical twin boundaries during fatigue deformation in a ferritic stainless steel crystal," *Philosophical magazine letters*, vol. 72, no. 5, pp. 297–301, 1995.

[283] D. Brandon, "The structure of high-angle grain boundaries," *Acta metallurgica*, vol. 14, no. 11, pp. 1479–1484, 1966.

[284] A. Ojha, H. Sehitoglu, L. Patriarca, and H. Maier, "Twin nucleation in Fe-based bcc alloys—modeling and experiments," *Modelling and Simulation in Materials Science and Engineering*, vol. 22, no. 7, p. 075010, 2014.

[285] K. Ma, J. Wang, V. Singh, B. Tamersoy, Y.-J. Chang, A. Wimmer, and T. Chen, "Multimodal image registration with deep context reinforcement learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 240–248.

[286] Y. Hu, M. Modat, E. Gibson, W. Li, N. Ghavami, E. Bonmati, G. Wang, S. Bandula, C. M. Moore, M. Emberton *et al.*, "Weakly-supervised convolutional neural networks for multimodal image registration," *Medical image analysis*, vol. 49, pp. 1–13, 2018.

[287] H. Schaeben, "A simple standard orientation density function: The hyperspherical de la Vallée Poussin kernel," *physica status solidi (b)*, vol. 200, no. 2, pp. 367–376, 1997.

[288] J. Han, H. Li, Z. Zhu, L. Jiang, H. Xu, and L. Ma, "Effects of processing optimisation on microstructure, texture, grain boundary and mechanical properties of Fe–17Cr ferritic stainless steel thick plates," *Materials Science and Engineering: A*, vol. 616, pp. 20–28, 2014.

[289] J. Luster and M. Morris, "Compatibility of deformation in two-phase ti-al alloys: Dependence on microstructure and orientation relationships," *Metallurgical and Materials Transactions A*, vol. 26, no. 7, pp. 1745–1756, 1995.

[290] C. Blochwitz, R. Richter, W. Tirschler, and K. Obrtlik, "The effect of local textures on microcrack propagation in fatigued fcc metals," *Materials Science and Engineering: A*, vol. 234, pp. 563–566, 1997.

[291] A. Castillo and S. R. Kalidindi, "A bayesian framework for the estimation of the single crystal elastic parameters from spherical indentation stress-strain measurements," *Frontiers in Materials*, vol. 6, p. 136, 2019.

[292] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," *Advances in neural information processing systems*, vol. 30, 2017.

[293] S. Torquato, "Statistical description of microstructures," *Annual review of materials research*, vol. 32, no. 1, pp. 77–111, 2002.

[294] N. Lubbers, T. Lookman, and K. Barros, "Inferring low-dimensional microstructure representations using convolutional neural networks," *Physical Review E*, vol. 96, no. 5, p. 052111, 2017.

[295] P. Du, A. Zebrowski, J. Zola, B. Ganapathysubramanian, and O. Wodo, "Microstructure design using graphs," *npj Computational Materials*, vol. 4, no. 1, pp. 1–7, 2018.

[296] J. Feng, Q. Teng, X. He, L. Qing, and Y. Li, "Reconstruction of three-dimensional heterogeneous media from a single two-dimensional section via co-occurrence correlation function," *Computational Materials Science*, vol. 144, pp. 181–192, 2018.

[297] N. Hadzic, A. R. Durmaz, F. Jünger, A. Rohrbach, and C. Eberl, "High resolution in-situ fatigue damage analysis and extrusion kinetics with ROCS microscopy," Poster presented at the SCHARF project meeting, Freiburg, Germany. INATECH, 2019.

[298] H. Abdolvand and A. J. Wilkinson, "Assessment of residual stress fields at deformation twin tips and the surrounding environments," *Acta Materialia*, vol. 105, pp. 219–231, 2016.

[299] A. Kundu and D. P. Field, "Geometrically necessary dislocation density evolution in interstitial free steel at small plastic strains," *Metallurgical and Materials Transactions A*, vol. 49, no. 8, pp. 3274–3282, 2018.

[300] W. Pantleon, "Resolving the geometrically necessary dislocation content by conventional electron backscattering diffraction," *Scripta Materialia*, vol. 58, no. 11, pp. 994–997, 2008.

[301] T. Ruggles, T. Rampton, A. Khosravani, and D. Fullwood, "The effect of length scale on the determination of geometrically necessary dislocations via EBSD continuum dislocation microscopy," *Ultramicroscopy*, vol. 164, pp. 1–10, 2016.

[302] J. Kysar, Y. Saito, M. Oztop, D. Lee, and W. Huh, "Experimental lower bounds on geometrically necessary dislocation density," *International Journal of Plasticity*, vol. 26, no. 8, pp. 1097–1123, 2010.

[303] J. Jiang, T. Britton, and A. Wilkinson, "Measurement of geometrically necessary dislocation density with high resolution electron backscatter diffraction: effects of detector binning and step size," *Ultramicroscopy*, vol. 125, pp. 1–9, 2013.

[304] W. v. Aarle and W. Ludwig, "X-ray diffraction contrast tomography in micro-CT lab source systems," European synchrotron radiation facility Grenoble (France), Tech. Rep., 2014.

[305] A. R. Durmaz, M. Müller, B. Lei, A. Thomas, D. Britz, E. A. Holm, C. Eberl, F. Mücklich, and P. Gumbsch, "A deep learning approach for complex microstructure inference," *Nature communications*, vol. 12, no. 1, pp. 1–15, 2021.

[306] M. D. Sangid, H. J. Maier, and H. Sehitoglu, "The role of grain boundaries on fatigue crack initiation–an energy approach," *International Journal of Plasticity*, vol. 27, no. 5, pp. 801–821, 2011.

[307] K. Differt, U. Essmann, and H. Mughrabi, "A model of extrusions and intrusions in fatigued metals II. surface roughening by random irreversible slip," *Philosophical Magazine A*, vol. 54, no. 2, pp. 237–258, 1986.

[308] C. Laird, "The influence of metallurgical structure on the mechanisms of fatigue crack propagation," in *Fatigue crack propagation*. ASTM International, 1967.

[309] P. Neumann, "Coarse slip model of fatigue," *Acta Metallurgica*, vol. 17, no. 9, pp. 1219–1225, 1969.

[310] P. Neumann, "The geometry of slip processes at a propagating fatigue crack—II," *Acta Metallurgica*, vol. 22, no. 9, pp. 1167–1178, 1974.

[311] A. A. Korda, Y. Mutoh, Y. Miyashita, T. Sadasue, and S. Mannan, "In situ observation of fatigue crack retardation in banded ferrite–pearlite microstructure due to crack branching," *Scripta Materialia*, vol. 54, no. 11, pp. 1835–1840, 2006.

[312] I. Ovid'Ko and A. Sheinerman, "Grain size effect on crack blunting in nanocrystalline materials," *Scripta Materialia*, vol. 60, no. 8, pp. 627–630, 2009.

[313] D. Swenson, "Transition between stage I and stage II modes of fatigue crack growth," *Journal of Applied Physics*, vol. 40, no. 9, pp. 3467–3475, 1969.

[314] P. Grünewald, J. Rauber, M. Marx, C. Motz, and F. Schaefer, "Fatigue crack growth in micro specimens as a tool to measure crack–microstructure interactions," *Fatigue & Fracture of Engineering Materials & Structures*, vol. 43, no. 12, pp. 3037–3049, 2020.

[315] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 07, 2020, pp. 13 001–13 008.

[316] R. Wu, S. Yan, Y. Shan, Q. Dang, and G. Sun, "Deep image: Scaling up image recognition," *arXiv preprint arXiv:1501.02876*, vol. 7, no. 8, 2015.

[317] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" *Advances in neural information processing systems*, vol. 27, 2014.

[318] O. Sedaghat and H. Abdolvand, "A non-local crystal plasticity constitutive model for hexagonal close-packed polycrystals," *International Journal of Plasticity*, vol. 136, p. 102883, 2021.

[319] Z. Zhang, D. Lunt, H. Abdolvand, A. J. Wilkinson, M. Preuss, and F. Dunne, "Quantitative investigation of micro slip and localization in polycrystalline materials under uniaxial tension," *International Journal of Plasticity*, vol. 108, pp. 88–106, 2018.

[320] A. Zeghadi, S. Forest, A.-F. Gourgues, and O. Bouaziz, "Ensemble averaging stress–strain fields in polycrystalline aggregates with a constrained surface microstructure–part 2: Crystal plasticity," *Philosophical Magazine*, vol. 87, no. 8-9, pp. 1425–1446, 2007.

[321] R. Sharma, I. Chelladurai, A. D. Orme, M. P. Miles, C. Giraud-Carrier, and D. T. Fullwood, "A step towards intelligent EBSD microscopy: machine-learning prediction of twin activity in MgAZ31," *Journal of Microscopy*, vol. 272, no. 1, pp. 67–78, 2018.

[322] A. Mangal, "Applied machine learning to predict stress hotspots in materials," Ph.D. dissertation, Carnegie Mellon University, 2018.

[323] A. R. Durmaz and E. Natkowski, "Experimental and computational micromechanical fatigue damage initiation data," 2021. [Online]. Available: http://dx.doi.org/10.24406/fordatis/151

[324] A. Goetz, A. R. Durmaz, M. Müller, A. Thomas, D. Britz, P. Kerfriden, and C. Eberl, "Addressing materials' microstructure diversity using transfer learning," *npj Computational Materials*, vol. 8, no. 1, pp. 1–13, 2022.

[325] R. Stewart and S. Ermon, "Label-free supervision of neural networks with physics and domain knowledge," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[326] L. Bellomarini, E. Sallinger, and S. Vahdati, "Knowledge graphs: The layered perspective," in *Knowledge Graphs and Big Data Processing*. Springer, Cham, 2020, pp. 20–34.

[327] S. Sharifzadeh, S. M. Baharlou, and V. Tresp, "Classification by attention: Scene graph classification with prior knowledge," *arXiv preprint arXiv:2011.10084*, 2020.

[328] G. A. Gesese, R. Biswas, M. Alam, and H. Sack, "A survey on knowledge graph embeddings with literals: Which model links better literal-ly?" *Semantic Web*, vol. 12, no. 4, pp. 617–647, 2021.

[329] Y. Zhang, T. Xiang, T. M. Hospedales, and H. Lu, "Deep mutual learning," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4320–4328.

[330] J. Carreira and C. Sminchisescu, "CPMC: Automatic object segmentation using constrained parametric min-cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1312–1328, 2011.

[331] A. Kolesnikov and C. H. Lampert, "Seed, expand and constrain: Three principles for weakly-supervised image segmentation," in *European conference on computer vision*. Springer, 2016, pp. 695–711.

[332] Z. Huang, X. Wang, J. Wang, W. Liu, and J. Wang, "Weakly-supervised semantic segmentation network with deep seeded region growing," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7014–7023.

[333] J. Lee, E. Kim, S. Lee, J. Lee, and S. Yoon, "Ficklenet: Weakly and semi-supervised semantic image segmentation using stochastic inference," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5267–5276.

[334] T.-H. Vu, H. Jain, M. Bucher, M. Cord, and P. Pérez, "Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation," pp. 2517–2526, 2019.

[335] S. Sankaranarayanan, Y. Balaji, A. Jain, S. N. Lim, and R. Chellappa, "Learning from synthetic data: Addressing domain shift for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3752–3761.

[336] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Machine learning*, vol. 79, no. 1, pp. 151–175, 2010.

[337] M. D. Sangid, "The physics of fatigue crack initiation," *International journal of fatigue*, vol. 57, pp. 58–72, 2013.

A Appendix

A.1 Assessment of variance in damage segmentation data sets

In all three materials considered for damage segmentation, the fact that the surface of the mesoscale specimens is slightly curved leads to a brightness gradient resulting in shading, i.e. an additional brightness variance among tiles. This is attributed to the specimen surface orientation dependence of the SE2 detector and leads to concealed extrusions or cracks in over- or underexposed regions at the specimen edges. Since the ferritic steel tiles originate from a comparatively large number of specimens, they contain a substantial brightness variance owed to the brightness and contrast settings during acquisition, see figure 3.5. Moreover, in the course of crack growth within the ferritic steel, debris caused by friction of the crack surfaces is driven out of the crack obscuring some relevant areas.

At the specimen surface of the ferritic and martensitic steels, shallow scratches and residual particle agglomerates from colloidal silica polishing clients (OP-S) exist. Another feature, apparent for instance in figure 3.5e, are the sudden gray value jumps that can be ascribed to grain boundaries. These pose a frequent location for damage to emerge [337]. In contrast to the remaining data sets, the surface finish of the copper material was carried out with an electropolishing step. Here, the characteristic grain structure-induced image background texture, such as discontinuity in gray values, is absent and a wavy topography is noticeable in some regions of the surface, see figure 3.6d.

In figure 3.5b, an instance of poor stitching correction is shown. The adjacent tiles are shifted against each other which results in unrealistic outliers in the data set.

A.2 Quantitative assessment of damage to grain boundary distance

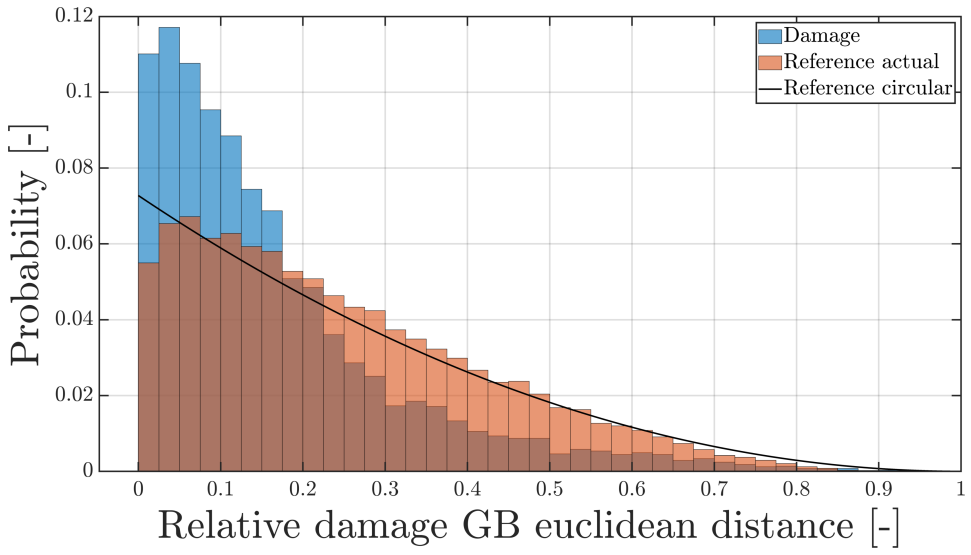

Figure A.1: Probability histogram distribution of the Euclidean distance between damaged/statistical reference pixels and its nearest grain boundary segments. While the actual reference was computed by assigning damage to EBSD pixels randomly, the circular reference considered a single unit circle grain. Adopted from [36].

A.3 Slip trace investigation

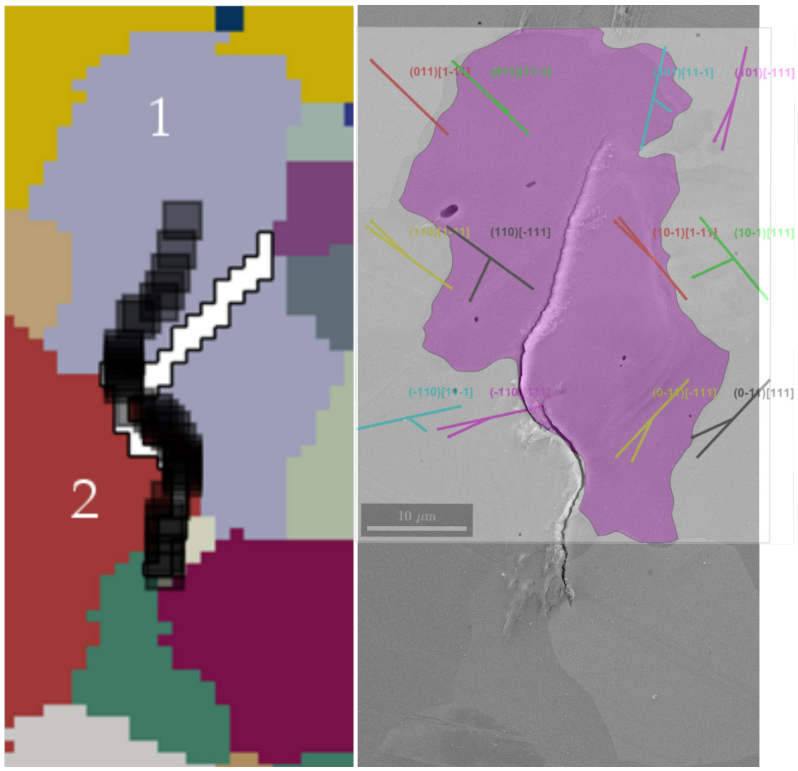

Figure A.2: Slip trace investigation. Comparison of actual and CPFEM-predicted crack paths (left) in black and white, respectively, with the slip traces of all potential slip systems (right). Adopted from [36].