Safeguarding AI-based autonomous driving against distribution shifts induced by wavefront aberrations of the windshield

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN (Dr. rer. nat.)

von der KIT-Fakultät für
Bauingenieur-, Geo- und Umweltwissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

Dominik Werner Wolf
geboren in Rheinfelden (Baden)

Karlsruhe 2025

Tag der mündlichen Prüfung:	5. Juni 2025
Referent:	Prof. Dr.-Ing. Markus Ulrich Institut für Photogrammetrie und Fernerkundung Karlsruher Institut für Technologie
1. Korreferent:	Prof. Dr. rer. nat. Alexander Braun Fakultät für Elektro- und Informationstechnik Hochschule Düsseldorf
2. Korreferent:	Prof. Dr.-Ing. Stefan Hinz Institut für Photogrammetrie und Fernerkundung Karlsruher Institut für Technologie

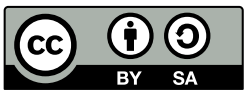
“Absolute certainty is a privilege of uneducated minds and fanatics. - It is, for scientific folk, an unattainable ideal.“ – quote by Keyser [94].

Dominik Werner Wolf

Safeguarding AI-based autonomous driving against distribution shifts induced by wavefront aberrations of the windshield

Doctoral Thesis
Date of examination: 05.06.2025
Referees:
Prof. Dr.-Ing. Markus Ulrich
Prof. Dr. rer. nat. Alexander Braun
Prof. Dr.-Ing. Stefan Hinz

Karlsruhe Institute of Technology
Department of Civil Engineering, Geo and Environmental Sciences
Institute of Photogrammetry and Remote Sensing
Kaiserstr. 12
76131 Karlsruhe



Abstract

The autonomous mobility of tomorrow should make the traffic flow more efficient, thereby saving resources and increasing road safety. By reducing greenhouse gas emissions, autonomous driving can contribute to fulfill the Paris climate goals if and only if the potential of operational effects (e.g., eco-driving, platooning, and intersection connectivity) [54] are unlocked. To enable autonomous driving, camera systems are used which are installed behind the vehicle windshield. The performance of AI-based perception modules is influenced by the optical properties of the glazing. The two main components are distortion and blurring.

The impact of these optical effects on the prediction accuracy of AI-based perception modules is tackled by two different strategies. On the one hand, the distortion will be actively corrected using high-resolution deflection measurements and, on the other hand, the blurring is suppressed to negligibleness by adequate quality requirements for the windshield. In order to realize this strategy, the correlations between glass quality and AI system performance must first be investigated so that the safety requirements from the AI world can be mapped to optical characteristics of the windshield.

Many AI systems are opaque [187], which makes the relationship between image-based input data and AI prediction difficult for the user to grasp. This poses a major problem, especially for security-relevant applications. For example, determining performance limits for an AI system within a required confidence level is extremely difficult or even impossible. Findings from this dissertation can help to increase the reliability of AI systems for autonomous driving and provide a scientific baseline for decision makers at international standardization institutes (e.g. ISO6041).

In addition, this work contributes to increasing the robustness of AI-based systems against optical aberrations of the windshield based on Fourier optical principles, novel optical measurement technologies and state-of-the-art machine learning methods. The resulting trustworthiness of the system is essential in order to establish the necessary social acceptance to realize the added value of autonomous driving on a large scale and thus sustainably increase safety and resource efficiency in mobility.

Kurzfassung

Die autonome Mobilität von morgen soll den Verkehrsfluss effizienter gestalten und damit Ressourcen einsparen und die Verkehrssicherheit erhöhen. Durch die Reduktion der Treibhausgasemissionen kann das autonome Fahren einen Beitrag zur Erfüllung der Pariser Klimaziele leisten, sofern die Potentiale der operativen Effekte – wie zum Beispiel des eco-driving, platooning, und die intersection connectivity [54] – vollumfänglich genutzt werden. Um autonomes Fahren zu ermöglichen, werden Kamerasysteme eingesetzt, welche hinter der Fahrzeugverglasung angebracht sind. Die Leistungsfähigkeit KI-basierter Erkennungsmodule wird durch die optischen Eigenschaften der Verglasung beeinflusst. Die zwei Hauptkomponenten sind dabei die Verzeichnung und die Unschärfe.

Den Auswirkungen dieser optischen Effekte auf die Prädiktionsgenauigkeit KI-basierter Erkennungsmodule soll durch zwei unterschiedliche Strategien begegnet werden. Zum einen soll die Verzeichnung über hochaufgelöste Deflektionsmessungen aktiv korrigiert werden und zum anderen soll die Unschärfe durch erhöhte Qualitätsanforderungen an die Verglasung unterdrückt werden, sodass diese Aberrationskomponente vernachlässigbar wird. Um dieses Vorhaben zu realisieren, sind zunächst die Korrelationen zwischen Glasqualität und KI-System-Performance zu untersuchen, damit die Sicherheitsanforderungen aus der KI-Welt auf optische Kenngrößen abgebildet werden können.

Viele KI-Systeme weisen eine Opazität [187] auf, was den Zusammenhang zwischen Bild-Eingabedaten und KI-Ausgabe für den Nutzer nur schwer greifbar macht. Dies stellt insbesondere bei sicherheitsrelevanten Anwendungen ein großes Problem dar. Beispielsweise ist die Bestimmung von Leistungsgrenzen innerhalb eines erforderlichen Vertrauensniveaus für solche KI-Systeme nicht oder nur schwer möglich. Erkenntnisse aus dieser Dissertation können helfen, die Zuverlässigkeit von KI-Systemen für das autonome Fahren zu erhöhen und bieten eine Entscheidungsgrundlage zur internationalen Standardisierung im Rahmen der ISO6041.

Zudem leistet diese Arbeit einen Beitrag zur Steigerung der Robustheit KI-basierter Systeme gegenüber optischen Aberrationen der Windschutzscheibe auf Basis der Fourier Optik, der optischen Messtechnik und dem maschinellen Lernen. Die dadurch erzielte Vertrauenswürdigkeit des Systems ist essentiell, um die notwendige gesellschaftliche Akzeptanz zu etablieren, um den Mehrwert des autonomen Fahrens großskalig zu realisieren und damit die Sicherheit sowie die Ressourceneffizienz in der Mobilität nachhaltig zu erhöhen.

Contents

1 Introduction

This chapter includes elements with editorial changes from

[194] Dominik Werner Wolf, Markus Ulrich, and Nikhil Kapoor. "Sensitivity analysis of AI-based algorithms for autonomous driving on optical wavefront aberrations induced by the windshield". In: *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)* (Oct. 2023), pp. 4102–4111. DOI: 10.1109/ICCVW60793.2023.00443. URL: https://ieeexplore.ieee.org/document/10350923/

marked with an orange line and includes elements with editorial changes from

[190] Dominik Werner Wolf, Alexander Braun, and Markus Ulrich. *Optical aberrations in autonomous driving: Physics-informed parameterized temperature scaling for neural network uncertainty calibration*. (Under review). 2024. arXiv: 2412.13695 [cs.CV]. URL: https://arxiv.org/abs/2412.13695

marked with a cyan line.

1.1 Motivation

Autonomously driving cars perceive the environment through different sensors [81], e.g., wide-angle cameras, telephoto cameras, Light Detection and Ranging (LiDAR) sensors etc. Typically, the measured sensor signals serve as the input to a neural network, which is supposed to predict the actions required (e.g., acceleration, steering angle etc.) to reach the next state [45]. In the development phase, the neural network is trained on a training dataset that is typically captured by a small fleet of test mules. If everything goes well with the architectural design and the neural network demonstrates sufficient accuracy and generalizability, then the autonomous driving functionality might be considered as operational. As the conceptual phase is taken to serial production, the real-world performance of the AI-based driving function might differ by a huge margin from what has been observed during development. A prominent contributor to this phenomenon is the perception chain of the telephoto camera (i.e., a camera with a telephoto lens), which is distinguished by a long focal length resulting in a high pixel resolution per field-angle. As a consequence, the target application of telephoto cameras is object recognition for far-field objects, especially important on highways with high driving speeds. Unfortunately, this

benefit also has a downside, namely an increased sensitivity for optical aberrations. Every car has a windshield, and every windshield has its unique aberration pattern. This has not been an issue for standard automotive cameras because the width of the blurring kernel induced by the windshield was always smaller than a pixel-pitch of the Complementary Metal-Oxide-Semiconductor (CMOS) sensor. With the use of telephoto cameras, this does not hold true anymore and the images captured might by heavily impacted by the windshield in terms of sharpness [194]. This gives rise to a shift in image quality between the test mule recordings used for training and the car-by-car perception chain. The shift in image quality ultimately results in a distribution shift between the training dataset and the real-world images captured by the telephoto camera through the specific windshield. If we consider a static evaluation setup instead of real-time instances then the distribution shift manifests itself as a dataset shift between the training dataset and the test dataset. The distribution shift forces the underlying neural network to extrapolate, which will negatively affect the performance. Unfortunately, the trivial solution of mounting the telephoto camera on the car roof is an unfavorable option because of external effects (e.g., mud, glaciation etc.) – which would require additional systems making the car heavier and more expensive – and for aesthetic reasons. On the contrary, the windshield provides a solid coverage and all existing systems (e.g., wipers, heating etc.) can be utilized.

As a consequence, it has do be investigated how to handle the aberration-driven distribution shift induced by windshields. This thesis serves this objective and contributes towards robust and trustworthy AI-based predictions under optical aberrations. For this purpose, a detailed study of the Fourier-optical imaging process is required, as a link between physics and AI-based computer vision needs to be established. The research questions addressed in this thesis live on a manifold spanned by three disciplines: wavefront aberration driven distribution shifts, metrological characterization of the windshield, and neural network uncertainty calibration. In detail, the contributions serve as a scientific basis for tackling the burning questions in the field: "How does the windshield impact the imaging process?", "How to parameterize and measure this influence?", "How to define system- and part-level requirements for the perception chain?", and ultimately "How can prior physical knowledge be leveraged to safeguard neural network performance and calibration under optical aberrations?".

For the homologation of autonomous driving vehicles, it will be necessary to perform a holistic analysis of the entire functional chain. Optical aberrations induced by the windshield are only one aspect of the entire challenge. Other impact factors might also be critical, like weather conditions [209] (e.g., rain, snow, fog etc.), out-of distribution events or lighting conditions. Those effects are not discussed in this thesis, which does not imply a judgment on the severity. Nevertheless, it is important to emphasize that the aberration-related distribution shift has a significant effect on diverse AI-based tasks, in particular on semantic segmentation as will be shown in this thesis. Hence, the contributions of this dissertation will facilitate the deployment of telephoto cameras behind the windshield

for enabling autonomous driving under ideal conditions, which is a necessary but not a sufficient condition for ensuring reliable operation under real-world conditions.

1.2 Synopsis

The structure of the thesis is split in eight main chapters. Chapter 2 introduces into the essential theoretical concepts in order to make the remaining chapters more easily comprehensible. The theory part elaborates on the principles of Fourier optics, introduces into the machine learning framework with a focus on uncertainty quantification and calibration, and finally, the specific features of windshields are discussed.

Very fundamentally, optical aberrations will perturb the imaged scene x. The imaging process can be abstractly described by a perturbation operator $\hat{\mathcal{H}}$, which generally depends on the field angles $(\varphi,\ \theta)$ and hence on the pixel coordinates $(\zeta_x,\ \zeta_y)$. Physically, the perturbation operator $\hat{\mathcal{H}}$ can be decomposed into a tilting operator $\hat{\mathcal{T}}$ and a blurring operator $\hat{\mathcal{B}}$, which are not commutable $\left[\hat{\mathcal{B}},\ \hat{\mathcal{T}}\right] \neq 0$ [24]. The tilting operator denotes optical distortions – straight lines remain infinitesimal thin but can be mapped into lines of arbitrary shape – and the blurring operator characterizes the sharpness reduction due to curvature modulations of the nominal wavefront. As both operators describe physically distinct effects, different aberration-handling strategies exist as illustrated in Figure 1.1.

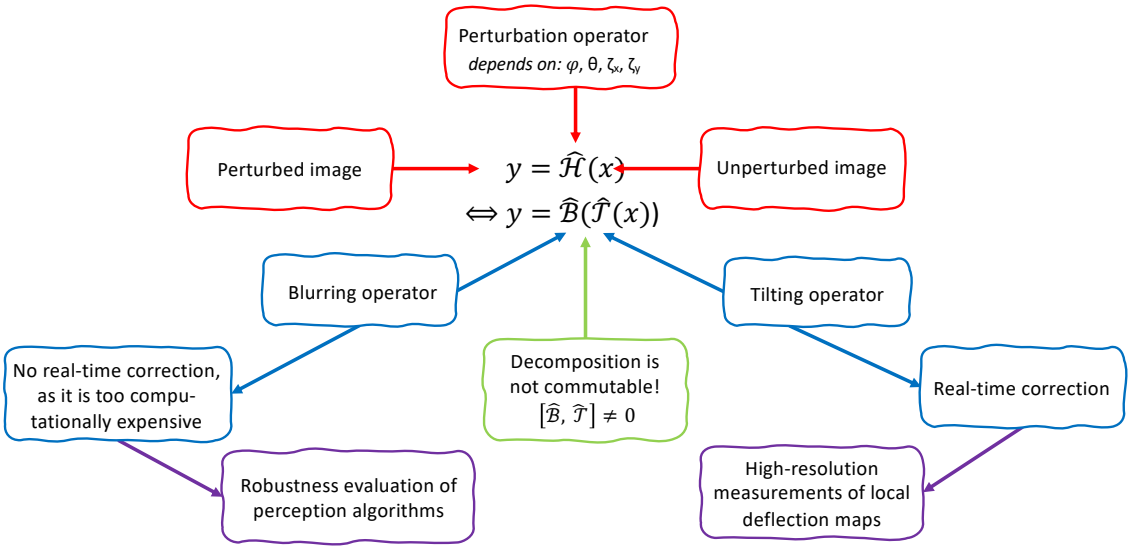


Figure 1.1: The impact of optical aberrations on the imaging process is abstractly described by the perturbation operator $\hat{\mathcal{H}}$. The decomposition of the perturbation operator into a tilting operator $\hat{\mathcal{T}}$ and a blurring operator $\hat{\mathcal{B}}$ enables us to address the problem of safeguarding level-4 functionalities against dataset shifts driven by optical aberrations of the windshield with distinct strategies.

The tilting operator is compensated by a real-time correction strategy utilizing high-resolution measurements of local deflection maps. Unfortunately, current optical distortion

measurement systems in the automotive industry are not capable of capturing the deflection maps as seen by the ADAS camera with a varying field-of-view across the image. As a consequence, we will evaluate a new measurement concept based on Background-Oriented-Schlieren (BOS) imaging in Chapter 3, which provides point-wise deflection vectors for a non-parametric, grid-based rectification of the distortion. The blurring operator has been neglected so far in the verification strategy of ADAS functionalities because it is insignificant for non-telephoto cameras as elaborated in Section 1.1. On the contrary, for perception systems equipped with telephoto cameras, it is intended to suppress the impact of the blurring operator on the imaging process by imposing sufficiently strict optical quality requirements on the windshield such that the residual degradation can be neglected. This strategy imposes a huge challenge for the glazing manufactures as the windshield is already optimized to serve different requirements simultaneously. Changing this multi-dimensional optimum in favor of higher optical quality in terms of the blurring operator might result in a worse performance regarding other windshield Key Performance Indicators (KPIs), like the head injury criterion (HIC), which depends on the internal stress distribution and determines the breaking behavior of the windshield. Nevertheless, a real-time correction of the blurring operator by applying a position-dependent deconvolution is too computationally expensive with the current on-board hardware resources. Furthermore, previous studies on the effect of dataset shifts [138] and noise corruptions [77] on image classification underpin the importance of optical robustness analyses for autonomous driving algorithms. The impact of blurring driven dataset shifts induced by optical aberrations of the windshield on traffic sign classification has already been quantified as an accuracy drop of up to ten percent [100], such that the blurring operator can not be neglected a priori. As a consequence, we will first investigate the limitations of current optical quality requirements in Chapter 4. Subsequently, we will perform a sensitivity analysis in Chapter 5 of AI-based semantic segmentation on wavefront aberrations induced by the windshield. In this work, we will focus on semantic segmentation because a pixel-wise class prediction is hypothesized to be more sensitive for optical aberrations compared to macro-level predictions in image classification or object detection. The strategy of safeguarding level-4 [87] functionalities against the blurring operator by imposing higher optical quality requirements ultimately relies on finding a bijective mapping between neural network and optical KPIs.

The conclusions of Chapter 5 will highlight that not only the performance of the semantic segmentation task is affected by optical aberrations, but also the calibration of the neural network is corrupted by the dataset shift, which leads to a lack of trustworthiness in the model's predictions. To address this issue we will introduce a novel calibration strategy for neural networks in Chapter 6, which we refer to as Physics-Informed Parameterized Temperature Scaling (PIPTS). PIPTS incorporates a physical inductive bias in the form of the Zernike coefficients of the second radial order to condition the calibration temperature on the aberration magnitude, which ultimately drives the dataset shift. For providing the relevant ground truth information about the system level Zernike coefficients of the

optical perception system for the PIPTS training, the wavefront modulation induced by the windshield – referenced to the incoming wavefront that is characterized by the non-diffraction-limited ADAS camera – must be known a priori. As this information can not be obtained by a standard Shack-Hartmann wavefront sensor, a novel measurement approach is presented in Chapter 7. This measurement approach builds upon the BOS imaging method introduced in Chapter 3, extending it to enable the extraction of part-level Zernike coefficients for the windshield from a high-resolution refractive power measurement map.

The loss functions studied in Chapter 6 reveal different minima for the mean Expected Calibration Error (mECE) and the Area Under the Reliability Error Curve (AUREC), which indicates that different calibration measures are not simultaneously optimized. As a consequence, we further investigate the coupling between several calibration measures in Chapter 8. The results demonstrate that controlling the mECE is insufficient for a holistic calibration strategy. The decoupling of neural network calibration measures generates a degree of freedom, which poses a significant risk in ensuring model trustworthiness under real world conditions. This challenge underpins the need for further research in the direction of safeguarding AI-based algorithms for autonomous driving against dataset shifts, e.g. induced by optical aberrations of the windshield.

Finally, Chapter 9 summarizes the scientific contributions of this doctoral thesis and underscores their relevance for autonomous driving.

1.3 Contributions

The contribution of this thesis to the scientific community is demonstrated by six peer-reviewed publications. These papers cover different domains as illustrated by Figure 1.2.

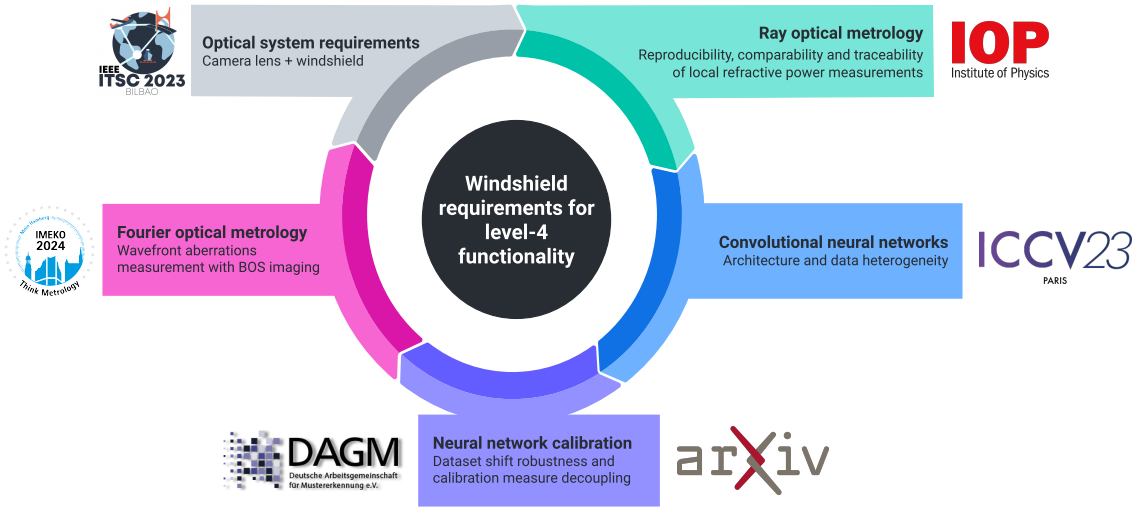


Figure 1.2: Interdisciplinary research landscape spanned by six publications.

The **ITSC** paper [193] addresses the limitations of current optical quality measures and reveals blind spots in the requirement landscape for windshields. In addition, the **Metrologia** paper [192] elaborates on utilizing BOS imaging for refractive power measurements. Furthermore, the **ICCV** paper [194] demonstrates that the Strehl ratio and the Optical Informative Gain (OIG) are superior measures than the Modulation Transfer Function (MTF) at half-Nyquist frequency. Moreover, the **ICCV** paper [194] highlights that neural network calibration is also heavily corrupted by optical aberrations of the windshield. The **arXiv** paper [190] addresses this challenge by incorporating the information about the wavefront aberrations as a physical inductive bias into the calibration process. For training the PIPTS calibrator, the Zernike coefficients of all optical components of the perception chain must be known. For this task, the **IMEKO** paper [191] presents a proof of concept study on how to adapt BOS imaging for retrieving the Zernike coefficients of the windshield. Finally, the **GCPR** paper [189] reports evidence on the decoupling of neural network calibration measures, which demonstrates the need for future research towards trustworthiness in autonomous driving. In a nutshell, the novelty of the scientific contributions of this doctoral thesis has been demonstrated and the aforementioned publications are subsequently listed in chronological order:

[192] Dominik Werner Wolf, Markus Ulrich, and Alexander Braun. "Novel developments of refractive power measurement techniques in the automotive world". In: *Metrologia* 60 (6 Sept. 2023). ISSN: 0026-1394. DOI: 10.1088/1681-7575/acf1a4. URL: https://iopscience.iop.org/article/10.1088/1681-7575/acf1a4.

[193] Dominik Werner Wolf, Markus Ulrich, and Alexander Braun. "Windscreen Optical Quality for AI Algorithms: Refractive Power and MTF not Sufficient". In: *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)* (Sept. 2023), pp. 5190–5197. DOI: 10.1109/ITSC57777.2023.10421970. URL: https://ieeexplore.ieee.org/document/10421970/.

[194] Dominik Werner Wolf, Markus Ulrich, and Nikhil Kapoor. "Sensitivity analysis of AI-based algorithms for autonomous driving on optical wavefront aberrations induced by the windshield". In: *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)* (Oct. 2023), pp. 4102–4111. DOI: 10.1109/ICCVW60793.2023.00443. URL: https://ieeexplore.ieee.org/document/10350923/.

[191] Dominik Werner Wolf, Boris Thielbeer, Markus Ulrich, and Alexander Braun. "Wavefront aberration measurements based on the Background Oriented Schlieren method". In: *Measurement: Sensors* (2024), p. 101509. ISSN: 2665-9174. DOI: https://doi.org/10.1016/j.measen.2024.101509. URL: https://www.sciencedirect.com/science/article/pii/S2665917424004859.

[189] Dominik Werner Wolf, Prasannavenkatesh Balaji, Alexander Braun, and Markus Ulrich. "Decoupling of neural network calibration measures". In: *German Conference on Pattern Recognition (GCPR)* (2024).

[190] Dominik Werner Wolf, Alexander Braun, and Markus Ulrich. *Optical aberrations in autonomous driving: Physics-informed parameterized temperature scaling for neural network uncertainty calibration*. (Under review). 2024. arXiv: 2412.13695 [cs.CV]. URL: https://arxiv.org/abs/2412.13695.

2 Theory and supporting methodological contributions

This chapter includes elements with editorial changes from

[192] Dominik Werner Wolf, Markus Ulrich, and Alexander Braun. "Novel developments of refractive power measurement techniques in the automotive world". In: *Metrologia* 60 (6 Sept. 2023). ISSN: 0026-1394. DOI: 10.1088/1681-7575/acf1a4. URL: https://iopscience.iop.org/article/10.1088/1681-7575/acf1a4

marked with a purple line and includes elements with editorial changes from

[194] Dominik Werner Wolf, Markus Ulrich, and Nikhil Kapoor. "Sensitivity analysis of AI-based algorithms for autonomous driving on optical wavefront aberrations induced by the windshield". In: *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)* (Oct. 2023), pp. 4102–4111. DOI: 10.1109/ICCVW60793.2023.00443. URL: https://ieeexplore.ieee.org/document/10350923/

marked with an orange line and includes elements with editorial changes from

[189] Dominik Werner Wolf, Prasannavenkatesh Balaji, Alexander Braun, and Markus Ulrich. "Decoupling of neural network calibration measures". In: *German Conference on Pattern Recognition (GCPR)* (2024)

marked with a mint green line and includes elements with editorial changes from

[190] Dominik Werner Wolf, Alexander Braun, and Markus Ulrich. *Optical aberrations in autonomous driving: Physics-informed parameterized temperature scaling for neural network uncertainty calibration*. (Under review). 2024. arXiv: 2412.13695 [cs.CV]. URL: https://arxiv.org/abs/2412.13695

marked with a cyan line.

This chapter aims to provide a comprehensive theoretical foundation necessary for understanding the scientific contributions presented extensively in Chapters 3 through 9. To this end, essential theoretical concepts utilized throughout this dissertation will be introduced and appropriately tailored to align with the research objectives. This leads to supporting methodological contributions, which are presented in this chapter with the headline prefix: *"From theory to practice"*.

Section 2.1 will provide an introduction to Fourier optics inspired by Goodman [61]. The intention of Section 2.1 is twofold: firstly, to elucidate the governing equations required for the construction of the Fourier optical degradation model presented in Section 2.1.2, and secondly, to establish an intuitive comprehension of central optical quality measures examined in this work (MTF, Strehl ratio, and OIG). The optical threat model in Section 2.1.2 is a technical realization of the Fourier optical principles and lays the methodological foundation for conducting the experiments, which will lead to the scientific results in Chapter 5 and 6. For the parameterization of the optical threat model, we will utilize the Zernike coefficients, which will be introduced in Section 2.1.1.

Section 2.2 will highlight specific features of windshields, which is essential to grasp the overall context of this thesis. In Section 2.2.1, a brief overview of the windshield manufacturing process will be provided, alongside a discussion of how different windshield components influence the wavefront aberration pattern. This equips the reader with a fundamental understanding of the reasons behind the manufacturing-related limitations in the optical quality of windshields. Subsequently, we will derive the Kerkhof model in Section 2.2.2. The Kerkhof model will be employed for demonstrating the comparability of BOS imaging-based refractive power measurements in Chapter 3. As the derivation entails several assumptions and approximations, it is beneficial to recap the full derivation of the Kerkhof model, especially since the Kerkhof model is not well prevalent in standard literature. The implications of these simplifications and their impact on the model's effectiveness are critically evaluated in Section 2.2.3, thereby confirming the validity of the methodology applied in Chapter 3.

Section 2.3 will concisely present the essential pillars of metrology. Those principles will be utilized across the thesis, as the requirements for a valid measurement are very similar to those for a trustworthy AI-based prediction.

Section 2.4 will explore the fundamental concepts of machine learning. We will first examine the baseline encoder-decoder structure in Section 2.4.1, a widely used design in Convolutional Neural Networks (CNNs) for dense prediction tasks such as semantic segmentation, and subsequently present the specific architectures employed in this thesis. These CNN models generate the predictions upon which the analyses in Chapters 5, 6, and 8 are based. Furthermore, the corresponding datasets for training and evaluation will be briefly described in Section 2.4.2. Additionally, in Section 2.4.3 we will define the evaluation measures employed for assessing the neural network predictive performance. Following metrological principles [140, 1], predictions are only meaningful when accompanied by an unbiased estimate of uncertainty. As a consequence, we will elaborate comprehensively on uncertainty quantification in machine learning in Section 2.4.4. Subsequently, Section 2.4.5 will shine a light on how uncertainty estimation biases can be measured, a safety-critical issue that will be addressed in the research conducted in Chapters 5, 6, and 8. Moreover, we will touch on the topic of explainable AI in Section 2.4.6, as we aim to quantify the

local sensitivity of neural network predictions on changes in the input feature space in Chapter 5.

Finally, we will introduce a mathematical tool in Section 2.5 for assessing non-linear correlations, and we will demonstrate the superiority of the Chatterjee rank correlation measure over the Pearson correlation measure in Section 2.5.1 by studying an artificial test function designed for illustrative purposes. The Chatterjee rank correlation measure will enable us to deduce precise conclusions in Chapter 5 regarding which optical quality metric demonstrates the strongest correlation with the neural network performance and is thus most suitable for deriving system-level functional requirements for the perception chain.

2.1 Principles of Fourier optics

Fourier optics denotes the theory of how light propagates taking into account its wave nature. Hence, Fourier optical methods can be used to analyze how an incoming light distribution propagates through an optical instrument (e.g., a microscope) even if interference effects are a matter of concern. This is the advantage of Fourier optics over geometrical optics, which is an intensity-based theory. The limitations of geometrical optics become evident if we consider the concept of the focal point. If no optical aberrations are considered then all light rays would intersect at the same point implying an infinite energy density. This is obviously wrong, wherefore the focal point is in reality a circle of least confusion [76]. In order to properly calculate the shape of the intensity distribution in the focal plane across the transversal coordinates, Fourier optical considerations are required. In this section, we will analytically describe the imaging process from the Fourier optics point of view and introduce the concept of transfer functions, which is an essential principle utilized throughout this thesis. The derivation is inspired by Goodman [61] but the framework is condensed and tailored to serve the needs of the optical threat model presented in Section 2.1.2.

Very fundamentally, the excitation of the electromagnetic field can be described by a monochromatic light wave $u(\vec{x},\ t)$ in the form of:

$$\begin{aligned} & u(\vec{x},\ t) = \hat{\rho}(\vec{x}) \cdot \cos\left(2\pi f t + \phi(\vec{x})\right) \\ \Leftrightarrow\ & u(\vec{x},\ t) = \mathrm{Re}\left[\rho(\vec{x}) \cdot \exp\left(-2\pi \mathrm{i} f t\right)\right] \quad \text{with:} \quad \rho(\vec{x}) := \hat{\rho}(\vec{x}) \cdot \exp\left(-\mathrm{i}\phi(\vec{x})\right), \end{aligned} \tag{2.1}$$

with $\hat{\rho}$ denoting the electromagnetic field amplitude, f indicates the frequency of the harmonic excitation, and ϕ labels the local phase depending on the coordinate vector $\vec{x} \in \mathbb{R}^3$.

As a consequence of Maxwell's equations, the light wave $u(\vec{x},\ t)$ must satisfy the scalar wave equation [76, 61]:

$$\Box u(\vec{x},\ t) = 0\ , \tag{2.2}$$

where $\Box := c^{-2}\partial_t^2 - \triangle$ denotes the d'Alembert operator [8] and c refers to the speed of light in vacuum. Substituting the light wave from Equation (2.1) into the wave Equation (2.2) results in the time-independent Helmholtz equation for the phasor $\rho(\vec{x})$:

$$\begin{aligned} (\triangle + k^2)\rho(\vec{x}) = 0 \quad &\text{with:} \quad k = \frac{2\pi f}{c} = \frac{2\pi}{\lambda} \\ \Rightarrow \left[\partial_r^2 + \frac{2}{r}\partial_r + k^2\right]\rho(\vec{x}) = 0 \quad &\text{if: Rotational symmetry is assumed .} \end{aligned} \tag{2.3}$$

Here, the propagation of the light wave is assumed in free space with light speed c, wave number k and corresponding wavelength λ. A unit amplitude spherical wave, converging towards the point $\vec{x}_o$, satisfies the Helmholtz Equation (2.3) and is commonly known as the free-space Green's function [61]. Generally, Green's functions are the physical version of the impulse response function in control systems engineering and are applicable to linear differential operators. If the system can be characterized by a Green's function then the system output is given by the convolution of the driving term or input signal with the Green's function. This theoretical mechanism is the causal reason for the validity of the superposition principle in optics.

In a nutshell, the free-space Green's function G is given by a unit amplitude spherical wave spreading out from an optical field disturbance at $\vec{x}_o$:

$$G(r) = \frac{\exp\left(\mathrm{i}kr\right)}{r} \quad \text{with:} \quad r := |\vec{x} - \vec{x}_o|\ . \tag{2.4}$$

For any complex-valued functions $G(r)$ and $\rho(r)$ the Green's theorem [61]:

$$\iiint_V \left[G(r)\ \triangle\rho(r) - \rho(r)\ \triangle G(r)\right]\mathrm{d}x^3 = \oiint_A \left[G(r)\ \partial_n\rho(r) - \rho(r)\ \partial_n G(r)\right]\mathrm{d}x^2 \tag{2.5}$$

applies if their first and second partial derivatives are single-valued and continuous [61]. The partial derivative ∂_n applies in the outward normal direction, corresponding to the surface normal vector $\vec{n}$, at each point on the surface A. The requirement of continuity for Green's theorem requires us to restrict the integration domain V such that the discontinuity induced by the field disturbance at $\vec{x}_o$ is excluded. If this constraint is considered appropriately, then the Green's function $G(r)$ (2.4) and the electromagnetic field $\rho(r)$ (2.1) satisfy the Helmholtz Equation (2.3), wherefore the two terms within the volume integral cancel each other out reducing the Green's theorem to:

$$\oiint_A \left[G(r)\ \partial_n\rho(r) - \rho(r)\ \partial_n G(r)\right]\mathrm{d}x^2 = 0\ . \tag{2.6}$$

As a result of the constraint on the integration domain V, the surface integration over the domain A consists of an inner A_{in} and an outer part A_{out} as illustrated in Figure 2.1, which encloses the volume V:

$$\begin{aligned}\oiint_A \left[G(r)\,\partial_n\rho(r) - \rho(r)\,\partial_n G(r)\right] \mathrm{d}x^2 = &\oiint_{A_{\text{in}}} \left[G(r)\,\partial_n\rho(r) - \rho(r)\,\partial_n G(r)\right] \mathrm{d}x^2 \\ &+ \oiint_{A_{\text{out}}} \left[G(r)\,\partial_n\rho(r) - \rho(r)\,\partial_n G(r)\right] \mathrm{d}x^2 \,.\end{aligned} \tag{2.7}$$

As the total volume vanishes according to Equation (2.6), the inner and outer surface integral must be equal in magnitude:

$$-\oiint_{A_{\text{in}}} \left[G(r)\,\partial_n\rho(r) - \rho(r)\,\partial_n G(r)\right] \mathrm{d}x^2 \overset{(2.6)}{=} \oiint_{A_{\text{out}}} \left[G(r)\,\partial_n\rho(r) - \rho(r)\,\partial_n G(r)\right] \mathrm{d}x^2 \,. \tag{2.8}$$

Since the origin of the spherical wave at $\vec{x}_o$ is non-dimensional, the limit $A_{\text{in}} \to 0$ determines the integral on the left-hand side, which is given by the product of the infinitesimal surface area $4\pi\varepsilon^2$ and the integrand evaluated at the location $\vec{x}_o$ of the optical field disturbance in the limit of $\varepsilon \to 0$:

$$\begin{aligned}-\lim_{\varepsilon\to 0} \oiint_{A_{\text{in}}} &\left[G(r)\,\partial_n\rho(r) - \rho(r)\,\partial_n G(r)\right] \mathrm{d}x^2 \\ &= \lim_{\varepsilon\to 0} \left[4\pi\varepsilon^2 \left[\rho(r)\,\partial_n G(r) - G(r)\,\partial_n\rho(r)\right]\big|_{r=\varepsilon}\right] \,.\end{aligned} \tag{2.9}$$

Here, ε characterizes the radius of the infinitesimal volume element, as depicted in Figure 2.1. Substituting the Green's function $G(r)$ with its explicit form given by Equation (2.4) results in:

$$\begin{aligned}&\lim_{\varepsilon\to 0} \left[4\pi\varepsilon^2 \left[\rho(r)\,\partial_n G(r) - G(r)\,\partial_n\rho(r)\right]\big|_{r=\varepsilon}\right] \\ &\overset{(2.4)}{=} \lim_{\varepsilon\to 0} \left[4\pi\varepsilon^2 \left[\rho(\varepsilon) \left(\frac{\vec{r}\cdot\vec{n}}{||\vec{r}||\cdot||\vec{n}||}\right) \left(\mathrm{i}k - \frac{1}{\varepsilon}\right) \frac{\exp(\mathrm{i}k\varepsilon)}{\varepsilon} - \frac{\exp(\mathrm{i}k\varepsilon)}{\varepsilon}\,\partial_n\,\rho(r)\big|_{r=\varepsilon}\right]\right] \\ &\overset{\vec{r}\,\downarrow\!\downarrow\,\vec{n}}{=} \lim_{\varepsilon\to 0} \left[4\pi\varepsilon^2 \left[\rho(\varepsilon) \left(\frac{1}{\varepsilon} - \mathrm{i}k\right) \frac{\exp(\mathrm{i}k\varepsilon)}{\varepsilon} - \frac{\exp(\mathrm{i}k\varepsilon)}{\varepsilon}\,\partial_n\,\rho(r)\big|_{r=\varepsilon}\right]\right] \,.\end{aligned} \tag{2.10}$$

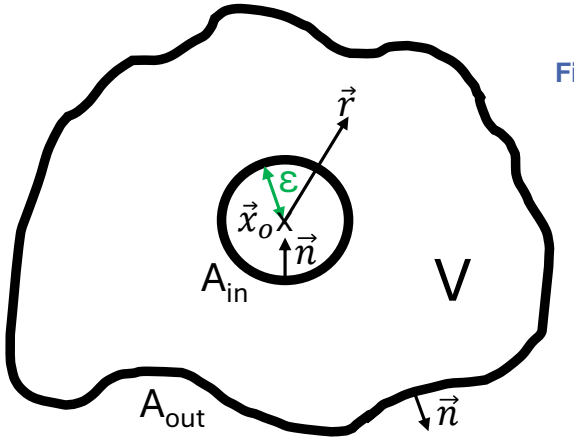


Figure 2.1: The integration domain V around the field disturbance at $\vec{x}_o$ is depicted. The integration domain is tailored to factor out the discontinuity induced by the field disturbance. Consequently, the volume V is enclosed by an outer surface A_{out} and an inner surface A_{in}. The corresponding surface normal vectors are denoted as $\vec{n}$. Finally, the infinitesimal radius of the sphere around $\vec{x}_o$ is labeled as ε.

The scalar product evaluates to -1 because the surface normal vector $\vec{n}$ of A_{in} points towards the field disturbance at location $\vec{x}_o$, whereas $\vec{r}$ points away from the excitation center. If the limit of $\varepsilon \to 0$ is determined term by term, then the left-hand side of Equation (2.8) reduces to:

$$-\lim_{\varepsilon \to 0} \oiint_{A_{\text{in}}} \left[G(r)\, \partial_n \rho(r) - \rho(r)\, \partial_n G(r) \right] \mathrm{d}x^2 \overset{(2.10)}{=} 4\pi \rho(\vec{x}_o) \,. \tag{2.11}$$

Putting everything together results in the integral theorem of Helmholtz and Kirchhoff:

$$\begin{aligned} 4\pi \rho(\vec{x}_o) &\overset{(2.8,\,2.11)}{=} \oiint_{A_{\text{out}}} \left[G(r)\, \partial_n \rho(r) - \rho(r)\, \partial_n G(r) \right] \mathrm{d}x^2 \\ \Leftrightarrow \rho(\vec{x}_o) &\overset{(2.4)}{=} \frac{1}{4\pi} \oiint_{A_{\text{out}}} \left[\left(\frac{\exp\left(\mathrm{i}kr\right)}{r} \right) \partial_n \rho(r) - \rho(r)\, \partial_n \left(\frac{\exp\left(\mathrm{i}kr\right)}{r} \right) \right] \mathrm{d}x^2 \,. \end{aligned} \tag{2.12}$$

The directional derivative of the Green's function along the normal vector $\vec{n}$ of the surface A_{out} is given by:

$$\partial_n G(r) = \frac{\partial G(r)}{\partial r} \cdot \frac{\partial r}{\partial n} = \left(\mathrm{i}k - \frac{1}{r} \right) \frac{\exp\left(\mathrm{i}kr\right)}{r} \cdot \cos\left(\measuredangle \{ \vec{r},\, \vec{n} \} \right) \,. \tag{2.13}$$

Since the distance r from the field disturbance at location $\vec{x}_o$ to an arbitrary point $\vec{x}$ is usually larger than many optical wavelengths, it holds that: $\lambda << r \Rightarrow k >> r^{-1}$. Consequently, the directional derivative can be approximated by:

$$\partial_n G(r) \approx \mathrm{i}k \cdot \frac{\exp\left(\mathrm{i}kr\right)}{r} \cdot \cos\left(\measuredangle \{ \vec{r},\, \vec{n} \} \right) \,. \tag{2.14}$$

Substituting this approximation into the integral theorem of Helmholtz and Kirchhoff from Equation (2.12) leads to:

$$\rho(\vec{x}_o) \overset{(2.12,\,2.14)}{\approx} \frac{1}{4\pi} \oiint_{A_{\text{out}}} \frac{\exp\left(\mathrm{i}kr\right)}{r} \left[\partial_n \rho(r) - \mathrm{i}k \rho(r) \cos\left(\measuredangle \{ \vec{r},\, \vec{n} \} \right) \right] \mathrm{d}x^2 \,. \tag{2.15}$$

The integral theorem of Helmholtz and Kirchhoff permits the calculation of the electromagnetic field ρ at location $\vec{x}_o$ based on the boundary values of the wave defined over any closed surface encircling that point [61]. Hence, the optical field disturbance at any point of interest $\vec{x}_o$ in the observation plane is solely determined by evaluating its values on the surface A_{out}.

If we focus specifically on diffraction induced by an aperture Σ enclosed by an infinite screen $\mathcal{S}$, as depicted in Figure 2.2, then the integration domain A_{out} can be further split into a spherical surface fragment and the intersection surface, which is given by the planar screen surface cutting the spherical wave. When we consider the surface element $\mathrm{d}x^2$ in

spherical coordinates with fixed radius R, then the integral in Equation (2.15) vanishes over the spherical surface fragment if and only if:

$$\lim_{R \mapsto \infty} R \left[\partial_n \rho(r)|_{r=R} - \mathrm{i}k\rho(R) \right] \stackrel{!}{=} 0 \,. \tag{2.16}$$

This requirement on the electromagnetic field ρ is known as the Sommerfeld radiation condition [61].

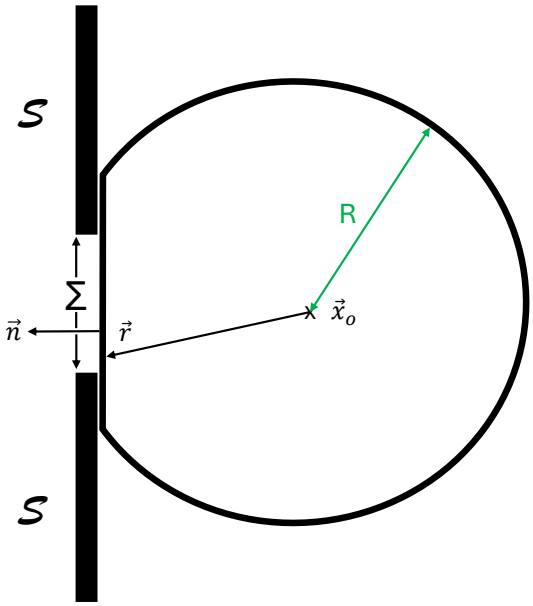


Figure 2.2: Diffraction of light by an aperture Σ enclosed by an infinite screen $\mathcal{S}$. Here, the wave propagates from the left to the right. A light impulse on the left hand side hits the aperture Σ with normal vector $\vec{n}$ and is diffracted. The observation point $\vec{x}_o$ characterizes the point of interest, where we want to determine the electromagnetic field ρ. The mapping from $\vec{x}_o$ to an arbitrary point on the aperture surface Σ is determined by the vector $\vec{r}$. Finally, a sphere with radius R is defined to enclose $\vec{x}_o$, as required by the integral theorem of Helmholtz and Kirchhoff.

Additionally, the integral over the planar screen surface $\mathcal{S}$, visualized in Figure 2.2, can be reduced to the domain of the aperture Σ if two boundary conditions are imposed:

$$\begin{aligned} &\textcircled{1} \quad \rho(r) = \partial_n \rho(r) = 0 \qquad \forall\, r \in \mathcal{S} \setminus \Sigma \,, \\ &\textcircled{2} \quad \left.\begin{aligned} \rho(r) &= \rho_s(r) \\ \partial_n \rho(r) &= \partial_n \rho_s(r) \end{aligned}\right\} \quad \forall\, r \in \Sigma \,. \end{aligned} \tag{2.17}$$

Here, $\rho_s(r)$ denotes the incident electromagnetic field emitted by an arbitrary source, wherefore it is assumed that the electromagnetic field and its directional derivative on the aperture Σ are unaffected by the screen $\mathcal{S}$. Those boundary conditions on ρ are typically referred to as the Kirchhoff boundary conditions [61].

If the Sommerfeld radiation condition and the Kirchhoff boundary conditions are fulfilled, then the integral theorem of Helmholtz and Kirchhoff from Equation (2.15) can be further simplified as:

$$\rho(\vec{x}_o) \stackrel{(2.16,\,2.17)}{\approx} \frac{1}{4\pi} \iint_{\Sigma} \frac{\exp(\mathrm{i}kr)}{r} \left[\partial_n \rho(r) - \mathrm{i}k\rho(r) \cos\left(\measuredangle \{\vec{r},\, \vec{n}\} \right) \right] \mathrm{d}x^2 \,, \tag{2.18}$$

which is referred to as the Fresnel-Kirchhoff diffraction formula [61].

Despite its experimental success and widespread practical use, Fresnel-Kirchhoff's diffraction theory builds upon fundamental mathematical inconsistencies [61]. In detail, the electromagnetic field ρ has to fulfill the Kirchhoff boundary conditions (2.17) but simultaneously underlies the Helmholtz Equation (2.3). The latter imposes a recursion relation between the Taylor-expansion coefficients of ρ, such that the expansion coefficient of order $(n+2)$ is proportional to the negative expansion coefficient of order n. As a consequence, the Kirchhoff boundary conditions for $n = 0$ and $n = 1$, which suppress the corresponding Taylor-expansion coefficients to zero, are propagated by the recursion relation to arbitrary high orders causing the electromagnetic field ρ to vanish across the entire space. Hence, only the trivial solution can be correctly described by the Fresnel-Kirchhoff diffraction formula [61].

In order to alleviate this situation, Sommerfeld utilized an alternative Green function to reduce the Kirchhoff boundary conditions [61]. If the Green's function G from Equation (2.4) is replaced by:

$$G(r,\, \tilde{r}) = \frac{\exp\left(\mathrm{i}k|\vec{x} - \vec{x}_o|\right)}{|\vec{x} - \vec{x}_o|} - \frac{\exp\left(\mathrm{i}k|\vec{x} - \vec{x}_{\tilde{o}}|\right)}{|\vec{x} - \vec{x}_{\tilde{o}}|} := \frac{\exp\left(\mathrm{i}kr\right)}{r} - \frac{\exp\left(\mathrm{i}k\tilde{r}\right)}{\tilde{r}}\,, \tag{2.19}$$

then the Kirchhoff boundary condition on the electromagnetic field ρ can be softened because the Green's function implicitly evaluates to zero for $r \in \mathcal{S}$. As a result, the first boundary condition defined in Expression (2.17) on ρ and its directional derivative is lifted. Hence, the mathematical inconsistencies arising from the Kirchhoff boundary conditions can be avoided by introducing the mirror image of the observation point at $\vec{x}_{\tilde{o}}$.

The substitution of the Green's function in the Fresnel-Kirchhoff diffraction formula (2.18) by the Sommerfeld's Green's function (2.19) requires us to recapitulate a few steps. First of all, the directional derivative of $G(r,\, \tilde{r})$ is given by:

$$\begin{aligned}
&\partial_n G(r,\, \tilde{r}) = \frac{\partial G(r,\, \tilde{r})}{\partial r} \cdot \frac{\partial r}{\partial n} + \frac{\partial G(r,\, \tilde{r})}{\partial \tilde{r}} \cdot \frac{\partial \tilde{r}}{\partial n} \\
\Leftrightarrow\; &\partial_n G(r,\, \tilde{r}) = \left(\mathrm{i}k - \frac{1}{r}\right) \frac{\exp\left(\mathrm{i}kr\right)}{r} \cdot \left(\cos\left(\measuredangle\left\{\vec{r},\, \vec{n}\right\}\right) - \cos\left(\measuredangle\left\{\vec{\tilde{r}},\, \vec{n}\right\}\right)\right) \\
\Leftrightarrow\; &\partial_n G(r,\, \tilde{r}) = 2\left(\mathrm{i}k - \frac{1}{r}\right) \frac{\exp\left(\mathrm{i}kr\right)}{r} \cdot \cos\left(\measuredangle\left\{\vec{r},\, \vec{n}\right\}\right)\,,
\end{aligned} \tag{2.20}$$

considering the fact that $\vec{\tilde{r}}$ denotes the location vector of the mirror image of the observation point, which is equal in length and only differs by a phase shift of π w.r.t. the enclosing angle to the normal direction $\vec{n}$. If the proportions of $\lambda << r \Rightarrow k >> r^{-1}$ are acknowledged, then the directional derivative can be approximated by:

$$\partial_n G(r,\, \tilde{r}) = 2\mathrm{i}k \cdot \frac{\exp\left(\mathrm{i}kr\right)}{r} \cdot \cos\left(\measuredangle\left\{\vec{r},\, \vec{n}\right\}\right)\,. \tag{2.21}$$

Substituting Equation (2.21) into the integral theorem of Helmholtz and Kirchhoff (2.12) results in:

$$\begin{aligned} \rho(\vec{x}_o) &= \frac{1}{4\pi} \iint_{\Sigma} \left[G(r,\ \tilde{r})\ \partial_n \rho(r) - 2\mathrm{i}k\rho(r) \cdot \frac{\exp(\mathrm{i}kr)}{r} \cdot \cos\left(\measuredangle\{\vec{r},\ \vec{n}\}\right) \right] \mathrm{d}x^2 \\ \Leftrightarrow \rho(\vec{x}_o) &= \frac{1}{\mathrm{i}\lambda} \iint_{\Sigma} \rho(r) \cdot \frac{\exp(\mathrm{i}kr)}{r} \cdot \cos\left(\measuredangle\{\vec{r},\ \vec{n}\}\right) \mathrm{d}x^2 \,. \end{aligned} \tag{2.22}$$

Here, the integration domain is also reduced from the entire outer surface A_{out} to the aperture surface Σ as the Sommerfeld radiation condition from Equation (2.16) is assumed to be fulfilled. Furthermore, the properties of the Sommerfeld's Green's function are utilized to discard the first term on the planar screen surface $\mathcal{S}$ as $G(r,\ \tilde{r})$ evaluates to zero by definition. Consequently, the first Kirchhoff boundary condition (2.17) is redundant and only the second condition applies as $\rho_s(r)$ is specified independently of the screen. This result is commonly denoted as the Rayleigh-Sommerfeld diffraction formula [61].

Without loss of generality we will assume that the aperture surface is spanned in the $(x,\ y)$ plane, wherefore the surface normal vector is given by $\vec{n} = \hat{e}_z$. If this assumption is made, then the cosine term is given by:

$$\cos\left(\measuredangle\{\vec{r},\ \vec{n}\}\right) = \frac{z}{r} \,. \tag{2.23}$$

Furthermore, if we adapt rectangular coordinates we can rewrite the distance r by:

$$r = \sqrt{(x_o - x_a)^2 + (y_o - y_a)^2 + z^2} \,, \tag{2.24}$$

where x_a and y_a indicate the transversal axes in the aperture plane. Similarly, x_o and y_o denote the transversal axes in observation plane, which corresponds to the imaging plane. By adapting this notation we can rewrite Equation (2.22) to:

$$\rho(x_o,\ y_o) = \frac{z}{\mathrm{i}\lambda} \iint_{\Sigma} \rho(x_a,\ y_a) \cdot \frac{\exp\left(\mathrm{i}k\sqrt{(x_o - x_a)^2 + (y_o - y_a)^2 + z^2}\right)}{(x_o - x_a)^2 + (y_o - y_a)^2 + z^2} \,\mathrm{d}x_a \mathrm{d}y_a \,. \tag{2.25}$$

Since the squared longitudinal distance between the aperture and the observation plane is usually larger than the squared transversal extent of the aperture, the squared distance r^2 can be replaced by the squared longitudinal distance z^2 as a zeroth order approximation in the denominator of Equation (2.25). This approximation holds true as long as the squared f-number is larger than one, which is generally valid for telephoto cameras considered within this thesis. For the exponent, this approximation is not sufficient since it is rescaled

by the wavenumber k, which equals $k \approx 1.14 \cdot 10^7\ \mathrm{m}^{-1}$ for green light. Hence, the Taylor expansion of the exponent will also comprise the first order term resulting in:

$$\begin{aligned} \rho(x_o,\ y_o) &\approx \frac{1}{\mathrm{i}\lambda z} \iint_{\Sigma} \rho(x_a,\ y_a) \cdot \exp\left(\mathrm{i}kz \left[1 + \frac{1}{2}\left(\frac{x_o - x_a}{z}\right)^2 + \frac{1}{2}\left(\frac{y_o - y_a}{z}\right)^2\right]\right) \mathrm{d}x_a \mathrm{d}y_a \\ \Leftrightarrow \rho(x_o,\ y_o) &= \frac{e^{\mathrm{i}kz}}{\mathrm{i}\lambda z} \iint_{\Sigma} \rho(x_a,\ y_a) \cdot \exp\left(\frac{\mathrm{i}k}{2z}\left[(x_o - x_a)^2 + (y_o - y_a)^2\right]\right) \mathrm{d}x_a \mathrm{d}y_a\ . \end{aligned} \tag{2.26}$$

This simplification of the exponent is denoted as the Fresnel approximation [61] and physically represents the replacement of the spherical Huygens wavelets in Equation (2.25) by quadratic surfaces. The Fresnel approximation is valid if the condition:

$$z^3 \gg \frac{\pi}{4\lambda} \max_{\vec{x}_o,\, \vec{x}_a} \left\{ \left[(x_o - x_a)^2 + (y_o - y_a)^2\right]^2 \right\}, \tag{2.27}$$

is satisfied [61]. The Fresnel approximation condition (2.27), which arises from the requirement that higher-order terms in the Taylor expansion of the phase are negligible, is typically violated for our applications. Nevertheless, as Goodman [61] points out, this approximation condition is way too strict as only a small region around the stationary point of the phase contributes to the diffraction integral even if higher-order terms are not suppressed on the entire domain. The principle of stationary phases applies if rapidly oscillating functions are integrated. In this case, most of the contributions effectively annihilate each other as positive and negative parts of the wave cancel out. However, in the close proximity of points where the phase varies slowly – or is stationary with vanishing derivative – the oscillatory contributions add up constructively. These points are referred to as the stationary points of the phase. As the exponent in the diffraction integral over the aperture Σ is scaled by the wavenumber k, which is in the order of $k \approx 1.14 \cdot 10^7\ \mathrm{m}^{-1}$ for green light, the principle of stationary phases is applicable. Consequently, the diffraction integral is dominated by the contribution of the point in the aperture that directly corresponds to the observation point, as the phase becomes stationary for this scenario. As a result, even if the Fresnel approximation condition (2.27) is violated, the Fresnel diffraction integral might be applicable as higher-order terms do not need to be globally negligible.

The Fresnel diffraction integral can also be rewritten in a slightly different fashion to highlight the mapping of the electromagnetic field from the aperture to the observation point, as described by:

$$\rho(x_o,\ y_o) = \iint_{\Sigma} \rho(x_a,\ y_a) \cdot h(x_o - x_a,\ y_o - y_a)\ \mathrm{d}x_a \mathrm{d}y_a = (\rho\ *\ h)\,(x_o,\ y_o)\ . \tag{2.28}$$

This result represents the convolution of the electromagnetic field ρ with quadratic wavelets across the aperture surface, which mathematically demonstrates the Huygens-

Fresnel principle [61]. The convolution kernel h for the Fresnel diffraction integral is given by:

$$h(x,\ y) = \frac{e^{\mathrm{i}kz}}{\mathrm{i}\lambda z} \cdot \exp\left(\frac{\mathrm{i}k}{2z}\left[x^2 + y^2\right]\right). \tag{2.29}$$

So far, we exclusively studied the diffracting properties of an aperture surface. If we consider a single lens as an additional element in the optical path – which represents the most fundamental imaging system – then we have to account for the phase delay induced by the lens accordingly. The local phase transformation properties of the lens t_φ are determined by the longitudinal optical path variation across the transversal axes:

$$t_\varphi(x_a,\ y_a) = \exp\left(\mathrm{i}k\left[n\Delta(x_a,\ y_a) + (\Delta(x_a = 0,\ y_a = 0) - \Delta(x_a,\ y_a))\right]\right), \tag{2.30}$$

where the first term in the exponent denotes the phase delay induced by the lens with refractive index n and the second term refers to the free space phase delay [61]. If we consider a spherical lens surface, with inner radius R_1 and outer radius R_2, then we can derive the thickness profile $\Delta(x_a,\ y_a)$ by geometrical considerations, which results in:

$$t_\varphi(x_a,\ y_a) \approx \exp\left(\mathrm{i}kn\Delta(x_a = 0,\ y_a = 0)\right) \cdot \exp\left(-\frac{\mathrm{i}k}{2f}\left[x_a^2 + y_a^2\right]\right). \tag{2.31}$$

This expression already assumes the validity of the paraxial approximation making it justifiable to approximate the spherical surface of the lens by parabolic surfaces. The focal length f encapsulates the optical properties of the lens in a single measure defined by:

$$\frac{1}{f} := (n-1) \cdot \left(\frac{1}{R_1} - \frac{1}{R_2}\right). \tag{2.32}$$

The constant phase factor in Equation (2.31) will be dropped hereafter because ultimately we are interested in the intensity profile in the observation plane, which is typically captured by a CMOS sensor. Hence, the relative longitudinal phase delay induced by the lens is given by:

$$t_\varphi(x_a,\ y_a) \approx \exp\left(-\frac{\mathrm{i}k}{2f}\left[x_a^2 + y_a^2\right]\right). \tag{2.33}$$

If we want to determine the electromagnetic field ρ at an arbitrary observation location $\vec{x}_o$ we need to add the phase modulation term t_φ of the lens into the Fresnel diffraction integral (2.26), which leads to:

$$\rho(x_o,\ y_o) = \frac{e^{\mathrm{i}kz}}{\mathrm{i}\lambda z} \iint_\Sigma \rho(x_a,\ y_a) \cdot \exp\left(-\frac{\mathrm{i}k}{2f}\left[x_a^2 + y_a^2\right]\right) \cdot \exp\left(\frac{\mathrm{i}k}{2z}\left[(x_o - x_a)^2 + (y_o - y_a)^2\right]\right) \mathrm{d}x_a \mathrm{d}y_a\,. \tag{2.34}$$

Manipulating this expression reveals some more insight:

$$\rho(x_o,\, y_o) = \frac{e^{\mathrm{i}kz}}{\mathrm{i}\lambda z} \exp\left(\frac{\mathrm{i}k}{2z}\left[x_o^2 + y_o^2\right]\right) \iint_{\Sigma} \rho(x_a,\, y_a) \times \exp\left(\frac{\mathrm{i}k}{2}\left(\frac{1}{z} - \frac{1}{f}\right)\left[x_a^2 + y_a^2\right]\right) \times \exp\left(-\frac{\mathrm{i}k}{z}\left[x_o x_a + y_o y_a\right]\right) \mathrm{d}x_a \mathrm{d}y_a \,. \tag{2.35}$$

If we restrict ourselves to the plane of the back focal length, then $z = f$ will hold true annihilating the second term of the surface integral:

$$\rho(x_o,\, y_o) = \frac{e^{\mathrm{i}kz}}{\mathrm{i}\lambda z} \cdot \exp\left(\frac{\mathrm{i}k}{2z}\left[x_o^2 + y_o^2\right]\right) \iint_{\Sigma} \rho(x_a,\, y_a) \cdot \exp\left(-\frac{\mathrm{i}k}{z}\left[x_o x_a + y_o y_a\right]\right) \mathrm{d}x_a \mathrm{d}y_a \,. \tag{2.36}$$

As a result, the diffraction pattern in the focal plane is given by the Fourier transform of the rescaled electromagnetic field ρ across the aperture surface Σ. Hence, we observe the Fraunhofer diffraction pattern without the need of fulfilling the Fraunhofer approximation criterion [61].

The diffraction integral in Equation (2.35) allows us to draw very general conclusions for an arbitrary imaging system, characterized by the focal length f and the aperture surface Σ. From a control engineering point of view it is tempting to investigate what happens if the imaging system needs to map a point-like object at the location $\vec{x}_s$, mathematically described by the Dirac-delta distribution. The electromagnetic field ρ in the observation plane will then characterize the impulse response function h of the optical system. Therefore, we will now consider a spherical wave in the form of Equation (2.4), which propagates through the aperture and the lens. If we consider similar simplifications as done by the Fresnel approximation in Equation (2.26), then it is justifiable to replace the denominator by the longitudinal distance γ between the point source at $\vec{x}_s$ and the aperture surface. Furthermore, the exponent is simplified by the paraxial approximation in order to use parabolic surfaces as a proxy for the spherical lens surface. As a consequence, the spherical wave in the aperture plane $\rho(\vec{x}_a)$, generated by the monochromatic Dirac-delta point source at $\vec{x}_s$, is given by:

$$\rho(x_a,\, y_a) \overset{(2.4,\, 2.26)}{\approx} \frac{1}{\gamma} \cdot \exp\left(\frac{\mathrm{i}k}{2\gamma}\left[(x_a - x_s)^2 + (y_a - y_s)^2\right]\right) . \tag{2.37}$$

In order to determine the impulse response function $h(x_o,\ y_o)$ of the optical system, we need to substitute the Dirac-delta point source into the Fresnel diffraction integral of Equation (2.35), which accounts for the phase modulation of the lens:

$$\begin{aligned} h(x_o,\ y_o) \overset{(2.35,\,2.37)}{=} \frac{e^{\mathrm{i}kz}}{\mathrm{i}\lambda z\gamma} \exp\left(\frac{\mathrm{i}k}{2z}\left[x_o^2+y_o^2\right]\right) \exp\left(\frac{\mathrm{i}k}{2\gamma}\left[x_s^2+y_s^2\right]\right) \\ \times \iint_{\Sigma} \exp\left(\frac{\mathrm{i}k}{2}\left(\frac{1}{z}+\frac{1}{\gamma}-\frac{1}{f}\right)\left[x_a^2+y_a^2\right]\right) \\ \times \exp\left(-\mathrm{i}k\left[\left(\frac{x_o}{z}+\frac{x_s}{\gamma}\right)x_a+\left(\frac{y_o}{z}+\frac{y_s}{\gamma}\right)y_a\right]\right)\mathrm{d}x_a\mathrm{d}y_a\ . \end{aligned} \tag{2.38}$$

The quadratic phase factor within the surface integral will generally broaden the impulse response function h if compared to the Fourier transform as a baseline [61]. As a consequence, the imaging system is optimized as the scalar coefficient of the quadratic phase factor is suppressed. This leads to the narrowest impulse response function that is physically possible and corresponds to the diffraction-limited case. For the elimination of the quadratic phase factor, the following relationship must be fulfilled:

$$\frac{1}{z}+\frac{1}{\gamma}-\frac{1}{f} \overset{!}{=} 0\ . \tag{2.39}$$

This physical condition is also known as the thin lens equation in geometrical optics [76]. The phase factors in front of the integral can be dropped if the phase distribution in the observation plane is not the matter of interest as it is the case for intensity-based imaging sensors (e.g., CMOS chips). This argument is not sufficient for the phase factor depending on the source location $\vec{x}_s$ because the impulse response function will be convolved with the scene to obtain the image. Hence, the integration domain will encompass the entire $(x_s,\ y_s)$ plane and the phase factor will explicitly contribute to it. Nevertheless, it might be justifiable to drop this phase factor as well if the phase of the quadratic phase factor remains sufficiently constant within the scene element that determines the electromagnetic field at a particular observation point [61]. Finally, we will define a scalar factor α in order to absorb all constants in front of the integral:

$$h(x_o,\ y_o) \overset{(2.38)}{=} \alpha \iint_{\Sigma} \exp\left(-\mathrm{i}k\left[\left(\frac{x_o}{z}+\frac{x_s}{\gamma}\right)x_a+\left(\frac{y_o}{z}+\frac{y_s}{\gamma}\right)y_a\right]\right)\mathrm{d}x_a\mathrm{d}y_a\ . \tag{2.40}$$

At this point, we will introduce the aperture function $P(x_a,\ y_a)$, which quantifies the transmittance properties of the planar screen and is typically given by a 2D top-hat function, assigning one to the domain of the aperture surface Σ and zeros elsewhere. Hence, the domain of integration can be extended to infinity:

$$h(x_o,\ y_o) = \alpha \int_{-\infty}^{\infty}\int P(x_a,\ y_a) \exp\left(-\mathrm{i}k\left[\left(\frac{x_o}{z}+\frac{x_s}{\gamma}\right)x_a+\left(\frac{y_o}{z}+\frac{y_s}{\gamma}\right)y_a\right]\right)\mathrm{d}x_a\mathrm{d}y_a\ . \tag{2.41}$$

Equation (2.41) can be further simplified by defining the magnification M of the system:

$$M := -\frac{z}{\gamma} \,, \tag{2.42}$$

and by substituting the integration variables with:

$$\tilde{\vec{x}}_a := \frac{1}{\lambda z} \cdot \vec{x}_a \quad \Rightarrow \quad \tilde{\alpha} := (\lambda z)^2 \cdot \alpha \,, \tag{2.43}$$

which leads us to the following expression for the impulse response function:

$$h(x_o,\, y_o) = \tilde{\alpha} \iint_{-\infty}^{\infty} P(\lambda z \tilde{x}_a,\, \lambda z \tilde{y}_a) \exp\left(-2\pi \mathrm{i}\left[(x_o - M x_s)\,\tilde{x}_a + (y_o - M y_s)\,\tilde{y}_a\right]\right) \mathrm{d}\tilde{x}_a \mathrm{d}\tilde{y}_a \,. \tag{2.44}$$

If an arbitrary input light-distribution $\rho(x_s,\, y_s)$ is considered, then the superposition principle applies since the differential operator of the Helmholtz Equation (2.3) is linear and the output light-distribution $\rho(x_o,\, y_o)$ in the observation plane can be determined by utilizing the impulse response function:

$$\begin{aligned} \rho(x_o,\, y_o) \;&= \iint_{-\infty}^{\infty} \rho(x_s,\, y_s) \cdot h(x_o,\, y_o \mid x_s,\, y_s)\, \mathrm{d}x_s \mathrm{d}y_s \\ \Leftrightarrow \rho(x_o,\, y_o) &\overset{(2.44)}{=} \tilde{\alpha} \iint_{-\infty}^{\infty} \rho(x_s,\, y_s) \\ &\qquad \times \iint_{-\infty}^{\infty} P(\lambda z \tilde{x}_a,\, \lambda z \tilde{y}_a) e^{-2\pi \mathrm{i}[(x_o - M x_s)\tilde{x}_a + (y_o - M y_s)\tilde{y}_a]}\, \mathrm{d}\tilde{x}_a \mathrm{d}\tilde{y}_a\, \mathrm{d}x_s \mathrm{d}y_s \,. \end{aligned} \tag{2.45}$$

If we perform the following substitution regarding the source plane coordinates:

$$\tilde{\vec{x}}_s := M \vec{x}_s \quad \Rightarrow \quad \tilde{\tilde{\alpha}} := \frac{\tilde{\alpha}}{M^2} \,, \tag{2.46}$$

we can rewrite the expression for the output light-distribution as:

$$\begin{aligned} \rho(x_o,\, y_o) = \tilde{\tilde{\alpha}} \cdot &\iint_{-\infty}^{\infty} \rho\left(\frac{\tilde{x}_s}{M},\, \frac{\tilde{y}_s}{M}\right) \\ &\times \iint_{-\infty}^{\infty} P(\lambda z \tilde{x}_a,\, \lambda z \tilde{y}_a) e^{-2\pi \mathrm{i}[(x_o - \tilde{x}_s)\tilde{x}_a + (y_o - \tilde{y}_s)\tilde{y}_a]}\, \mathrm{d}\tilde{x}_a \mathrm{d}\tilde{y}_a\, \mathrm{d}\tilde{x}_s \mathrm{d}\tilde{y}_s \,. \end{aligned} \tag{2.47}$$

The integral over the transformed source plane can be identified as a convolution of the electromagnetic field of the emitter with the integrated aperture function:

$$\begin{aligned} \rho(x_o,\, y_o) &\overset{(2.47)}{=} \tilde{\tilde{\alpha}} \cdot \rho\left(\frac{x_o}{M},\, \frac{y_o}{M}\right) * \iint_{-\infty}^{\infty} P(\lambda z \tilde{x}_a,\, \lambda z \tilde{y}_a) e^{-2\pi \mathrm{i}[x_o \tilde{x}_a + y_o \tilde{y}_a]}\, \mathrm{d}\tilde{x}_a \mathrm{d}\tilde{y}_a \\ \Leftrightarrow \rho(x_o,\, y_o) &= \frac{\tilde{\tilde{\alpha}}}{(\lambda z)^2} \cdot \rho\left(\frac{x_o}{M},\, \frac{y_o}{M}\right) * \mathcal{F}[P]\left(\frac{x_o}{\lambda z},\, \frac{y_o}{\lambda z}\right) . \end{aligned} \tag{2.48}$$

Consequently, the convolution kernel determining the imaging properties is given by:

$$\mathrm{PSF}_{\mathrm{coherent}} := \mathcal{F}[P]\left(\frac{x_o}{\lambda z},\, \frac{y_o}{\lambda z}\right) , \tag{2.49}$$

which is denoted as the coherent Point Spread Function (PSF). If incoherent systems are considered then the image formation process is the result of a linear intensity mapping instead of a linear amplitude mapping of the complex electromagnetic field as described by Equation (2.48). As a result, the incoherent PSF is characterized by:

$$\mathrm{PSF}_{\mathrm{incoherent}} := \mathrm{PSF}_{\mathrm{coherent}} \cdot \mathrm{PSF}^{*}_{\mathrm{coherent}} = \left|\mathcal{F}[P]\left(\frac{x_o}{\lambda z},\, \frac{y_o}{\lambda z}\right)\right|^2 . \tag{2.50}$$

As the convolution operation is quite expensive, it is tempting to transform the imaging Equation (2.48) into the frequency domain to exploit the convolution theorem.

As a result, the convolution operation can be replaced by a simple multiplication:

$$\mathcal{F}\left[|\rho|^2\right](k_{x_o},\, k_{y_o}) \overset{(2.48,\,2.50)}{=} \left(\frac{\tilde{\tilde{\alpha}} M^2}{(\lambda z)^2}\right)^2 \cdot \mathcal{F}\left[|\rho|^2\right](M k_{x_o},\, M k_{y_o}) \cdot \mathcal{F}\left[\left|\mathcal{F}[P]\left(\frac{x_o}{\lambda z},\, \frac{y_o}{\lambda z}\right)\right|^2\right](k_{x_o},\, k_{y_o}) , \tag{2.51}$$

in the frequency domain. This gives rise to the definition of the incoherent Optical Transfer Function (OTF) given by:

$$\mathrm{OTF}(k_{x_o},\, k_{y_o}) := \frac{\mathcal{F}\left[\left|\mathcal{F}[P]\left(\frac{x_o}{\lambda z},\, \frac{y_o}{\lambda z}\right)\right|^2\right](k_{x_o},\, k_{y_o})}{\mathcal{F}\left[\left|\mathcal{F}[P]\left(\frac{x_o}{\lambda z},\, \frac{y_o}{\lambda z}\right)\right|^2\right](k_{x_o} = 0,\, k_{y_o} = 0)} , \tag{2.52}$$

which is implicitly normalized by the frequency response at $\vec{k}_{x_o} = \vec{0}$. Hence, the OTF evaluates to one for the "DC" component and is monotonically decreasing [61]. If we only care about the intensity distribution in the observation plane, then the phase information is generally not of significance. Hence, the absolute of the OTF – defined as the Modulation Transfer Function (MTF) – is a common merit function for imaging systems:

$$\mathrm{MTF}(k_{x_o},\, k_{y_o}) := \left|\mathrm{OTF}(k_{x_o},\, k_{y_o})\right| \overset{(2.52)}{=} \left|\frac{\mathcal{F}\left[\left|\mathcal{F}[P]\left(\frac{x_o}{\lambda z},\, \frac{y_o}{\lambda z}\right)\right|^2\right](k_{x_o},\, k_{y_o})}{\mathcal{F}\left[\left|\mathcal{F}[P]\left(\frac{x_o}{\lambda z},\, \frac{y_o}{\lambda z}\right)\right|^2\right](k_{x_o} = 0,\, k_{y_o} = 0)}\right| . \tag{2.53}$$

We can massage this expression in order to simplify the computational implementation:

$$\mathrm{MTF}(k_{x_o},\, k_{y_o}) = \left| \frac{\mathcal{F}\left[\mathcal{F}[P]\left(\frac{x_o}{\lambda z},\, \frac{y_o}{\lambda z}\right) \cdot \mathcal{F}[P]^*\left(\frac{x_o}{\lambda z},\, \frac{y_o}{\lambda z}\right)\right](k_{x_o},\, k_{y_o})}{(\lambda z)^2 \iint_{-\infty}^{\infty} \left|P(\xi_x,\, \xi_y)\right|^2 \, \mathrm{d}\xi_x \mathrm{d}\xi_y} \right|, \tag{2.54}$$

where we utilized the Plancherel theorem to rewrite the denominator. Furthermore, we can shift the complex conjugate into the argument of the Fourier transform:

$$\mathrm{MTF}(k_{x_o},\, k_{y_o}) = \left| \frac{\mathcal{F}\left[\mathcal{F}[P]\left(\frac{x_o}{\lambda z},\, \frac{y_o}{\lambda z}\right) \cdot \mathcal{F}[P^*]\left(-\frac{x_o}{\lambda z},\, -\frac{y_o}{\lambda z}\right)\right](k_{x_o},\, k_{y_o})}{(\lambda z)^2 \iint_{-\infty}^{\infty} \left|P(\xi_x,\, \xi_y)\right|^2 \, \mathrm{d}\xi_x \mathrm{d}\xi_y} \right|, \tag{2.55}$$

by exploiting the relationship: $\mathcal{F}[J]^*(k) = \mathcal{F}[J^*](-k)$. As a next step, we will invert the outer Fourier transform in the nominator:

$$\mathrm{MTF}(k_{x_o},\, k_{y_o}) = \left| \frac{\mathcal{F}^{-1}\left[\mathcal{F}[P]\left(-\frac{x_o}{\lambda z},\, -\frac{y_o}{\lambda z}\right) \cdot \mathcal{F}[P^*]\left(\frac{x_o}{\lambda z},\, \frac{y_o}{\lambda z}\right)\right](k_{x_o},\, k_{y_o})}{(\lambda z)^2 \iint_{-\infty}^{\infty} \left|P(\xi_x,\, \xi_y)\right|^2 \, \mathrm{d}\xi_x \mathrm{d}\xi_y} \right|, \tag{2.56}$$

by recalling the relationship: $\mathcal{F}[J](k) = \mathcal{F}^{-1}[J](-k)$, which holds true if and only if the argument J is real-valued. The nominator can now be simplified by applying the convolution theorem:

$$\mathrm{MTF}(k_{x_o},\, k_{y_o}) = \left| \frac{\iint_{-\infty}^{\infty} P\left(-\frac{x_o}{\lambda z},\, -\frac{y_o}{\lambda z}\right) \cdot P^*\left(k_{x_o} - \frac{x_o}{\lambda z},\, k_{y_o} - \frac{y_o}{\lambda z}\right) \, \mathrm{d}x_o \mathrm{d}y_o}{(\lambda z)^2 \iint_{-\infty}^{\infty} \left|P(\xi_x,\, \xi_y)\right|^2 \, \mathrm{d}\xi_x \mathrm{d}\xi_y} \right|. \tag{2.57}$$

A final variable substitution leads to the compact expression:

$$\mathrm{MTF}(k_{x_o},\, k_{y_o}) = \left| \frac{\iint_{-\infty}^{\infty} P\left(\xi_x,\, \xi_y\right) \cdot P^*\left(k_{x_o} + \xi_x,\, k_{y_o} + \xi_y\right) \, \mathrm{d}\xi_x \mathrm{d}\xi_y}{\iint_{-\infty}^{\infty} \left|P(\xi_x,\, \xi_y)\right|^2 \, \mathrm{d}\xi_x \mathrm{d}\xi_y} \right|, \tag{2.58}$$

which boils down to the normalized autocorrelation function of the aperture function P:

$$\mathrm{MTF}(k_{x_o},\, k_{y_o}) = \left| \frac{P(k_{x_o},\, k_{y_o}) \star P(k_{x_o},\, k_{y_o})}{\iint_{-\infty}^{\infty} \left|P(\xi_x,\, \xi_y)\right|^2 \, \mathrm{d}\xi_x \mathrm{d}\xi_y} \right|. \tag{2.59}$$

So far, we exclusively considered optical systems without aberrations. Since the diffraction limit can never be achieved by a real optical element, we have to introduce aberrations in the presented methodology. For this reason, we will generalize the pupil function $P(\vec{x}_a)$ into the complex domain. Optical aberrations will induce local phase perturbations $kW(\vec{x}_a)$, where $W(\vec{x}_a)$ characterizes the local Optical Path Difference (OPD). Mathematically, the generalized pupil function $\tilde{P}(\vec{x}_a)$ can be defined as:

$$\tilde{P}(\vec{x}_a) := P(\vec{x}_a) \cdot \exp\left(2\pi \mathrm{i}\, \frac{W(\vec{x}_a)}{\lambda}\right) . \tag{2.60}$$

Hence, the perturbed MTF – as it is typically the case for automotive imaging systems – is determined by:

$$\widetilde{\mathrm{MTF}}(k_{x_o},\, k_{y_o}) := \left| \frac{\tilde{P}(k_{x_o},\, k_{y_o}) \star \tilde{P}(k_{x_o},\, k_{y_o})}{\iint_{-\infty}^{\infty} \left|P(\xi_x,\, \xi_y)\right|^2 \, \mathrm{d}\xi_x \mathrm{d}\xi_y} \right| . \tag{2.61}$$

In order to evaluate the optical system response for different wavelengths λ, the monochromatic MTF has to be marginalized across the normalized Power Spectral Density (PSD) of the light source. The area under the PSD curve quantifies the likelihood of emitting a photon in the wavelength range $[\lambda,\ \lambda + \Delta\lambda]$ by the light source. As a result, the polychromatic MTF is given by:

$$\widetilde{\mathrm{MTF}}(k_{x_o},\, k_{y_o}) := \int_0^{\infty} \widetilde{\mathrm{MTF}}(k_{x_o},\, k_{y_o} \mid \lambda) \cdot \mathrm{PSD}(\lambda)\, \mathrm{d}\lambda . \tag{2.62}$$

Consequently, chromatic aberrations will potentially also influence the performance of AI-based algorithms for autonomous driving but they are not discussed in detail within this thesis.

In the recent years the MTF at half-Nyquist frequency started to gain attention in the standardization of optical quality requirements for windshields. In Chapter 4 we will theoretically motivate why this measure is insufficient as a holistic quality standard and in Chapter 5 will then present quantitative evidence for this statement. Furthermore, we will highlight more suitable alternatives. Instead of specifying only a single spatial frequency requirement for the MTF it might be advisable to consider the entire spectrum. In order to do so, the area under the OTF curve can be evaluated. The spectral integral of the OTF in relation to the diffraction-limited OTF area is defined as the Strehl ratio [61] and is given by:

$$\mathrm{SR} := \frac{\iint_{-\infty}^{\infty} \widetilde{\mathrm{OTF}}(k_{x_o},\, k_{y_o})\, \mathrm{d}k_{x_o}\mathrm{d}k_{y_o}}{\iint_{-\infty}^{\infty} \mathrm{OTF}(k_{x_o},\, k_{y_o})\, \mathrm{d}k_{x_o}\mathrm{d}k_{y_o}} \overset{(2.52)}{=} \frac{\left|\mathcal{F}[\tilde{P}]\,(x_o = 0,\, y_o = 0)\right|^2}{\left|\mathcal{F}[P]\,(x_o = 0,\, y_o = 0)\right|^2} . \tag{2.63}$$

The equivalence of the two ratios points out that the Strehl ratio is a local measure in the spatial domain of the observation plane even though it accounts for the entire frequency spectrum. Hence, there is still a drawback in the definition of the Strehl ratio because it does not incorporate the knowledge about the shape of the PSF, which entirely characterizes the optical system. Therefore, an optical merit function would be desirable that shows a dependency on higher-order moments of the PSF as well. One possible metric that considers this constraint is introduced in this thesis as the Optical Informative Gain (OIG):

$$\mathrm{OIG} := \frac{\iint_{-\infty}^{\infty} \left|\mathcal{F}[\tilde{P}]\left(\frac{x_o}{\lambda z},\ \frac{y_o}{\lambda z}\right)\right|^4 \mathrm{d}x_o \mathrm{d}y_o}{\iint_{-\infty}^{\infty} \left|\mathcal{F}[P]\left(\frac{x_o}{\lambda z},\ \frac{y_o}{\lambda z}\right)\right|^4 \mathrm{d}x_o \mathrm{d}y_o} \overset{(2.52)}{=} \frac{\iint_{-\infty}^{\infty} \left|\widetilde{\mathrm{OTF}}(k_{x_o},\ k_{y_o})\right|^2 \mathrm{d}k_{x_o} \mathrm{d}k_{y_o}}{\iint_{-\infty}^{\infty} \left|\mathrm{OTF}(k_{x_o},\ k_{y_o})\right|^2 \mathrm{d}k_{x_o} \mathrm{d}k_{y_o}} . \tag{2.64}$$

Equation (2.64) takes advantage of the Plancherel theorem [37]. If the OIG is evaluated by measurement data then the domain of the OTF is restricted by the Nyquist frequency. Hence, the empirical OIG incorporates the resolution limitation given by the image sensor and relates to the amount of photonic energy, which can be spatially discriminated in relation to the diffraction-limited case. Furthermore, the aberration pattern induced by windshields exhibits a significant anisotropy (see Section 2.2.2). Therefore, we will disentangle the axes of the measures by explicitly redefining the Strehl ratio in terms of horizontal (SR_x) and vertical slices (SR_y):

$$\mathrm{SR}_x := \frac{\int_0^{f_{\mathrm{Ny}}} \widetilde{\mathrm{MTF}}(k_{x_o},\ k_{y_o}=0)\ \mathrm{d}k_{x_o}}{\int_0^{f_{\mathrm{Ny}}} \mathrm{MTF}(k_{x_o},\ k_{y_o}=0)\ \mathrm{d}k_{x_o}} \quad \text{and:} \quad \mathrm{SR}_y := \frac{\int_0^{f_{\mathrm{Ny}}} \widetilde{\mathrm{MTF}}(k_{x_o}=0,\ k_{y_o})\ \mathrm{d}k_{y_o}}{\int_0^{f_{\mathrm{Ny}}} \mathrm{MTF}(k_{x_o}=0,\ k_{y_o})\ \mathrm{d}k_{y_o}} , \tag{2.65}$$

and we also redefine the OIG accordingly:

$$\mathrm{OIG}_x := \frac{\int_0^{f_{\mathrm{Ny}}} \left|\widetilde{\mathrm{MTF}}(k_{x_o},\ k_{y_o}=0)\right|^2\ \mathrm{d}k_{x_o}}{\int_0^{f_{\mathrm{Ny}}} \left|\mathrm{MTF}(k_{x_o},\ k_{y_o}=0)\right|^2\ \mathrm{d}k_{x_o}} \quad \text{and:} \quad \mathrm{OIG}_y := \frac{\int_0^{f_{\mathrm{Ny}}} \left|\widetilde{\mathrm{MTF}}(k_{x_o}=0,\ k_{y_o})\right|^2\ \mathrm{d}k_{y_o}}{\int_0^{f_{\mathrm{Ny}}} \left|\mathrm{MTF}(k_{x_o}=0,\ k_{y_o})\right|^2\ \mathrm{d}k_{y_o}} . \tag{2.66}$$

These definitions apply symmetrically to the horizontal and vertical canonical axes in the observation plane. The substitution of the OTF by the MTF is justifiable if the incoherent PSF is sufficiently symmetric around the optical axis. In this special case, the OTF is real-valued and equals the MTF. This condition is typically fulfilled for the aberrations induced by the windshield.

2.1.1 Zernike basis decomposition

For the parameterization of the aberration pattern in terms of the Zernike coefficients ω_n, we will introduce the Zernike basis decomposition. Very fundamentally, the decomposition of a one-dimensional function $f(x)$ into a set of orthogonal polynomials $\mathcal{L}_q(x)$ on the domain $[a, b]$ is given by [17]:

$$f(x) = \sum_{i=0}^{\infty} \alpha_i \mathcal{L}_i(x) \quad , \quad \text{with:} \quad \alpha_i = \langle f | \mathcal{L}_i \rangle := \int_a^b f(x) \mathcal{L}_i(x) \, \mathrm{d}x \, . \tag{2.67}$$

In some cases – for example if wavefront aberrations are measured by a Shack-Hartmann sensor – it is desirable to decompose the derivative of $f(x)$ into a set of derivatives of $\mathcal{L}_q(x)$, which results in [17]:

$$f'(x) = \sum_{i=0}^{\infty} \beta_i \mathcal{L}'_i(x) \quad , \quad \text{with:} \quad \vec{\beta} = \left[\left\langle \mathcal{L}'_p \middle| \mathcal{L}'_q \right\rangle \right]^{-1} \left\langle f' \middle| \vec{\mathcal{L}}' \right\rangle \, . \tag{2.68}$$

Please be aware that the orthogonality relation for $\mathcal{L}_q(x)$ does not hold true for the derivatives of $\mathcal{L}_q(x)$ by default. If Equation (2.68) is integrated with respect to x, then the decomposition coefficients α_i and β_i can be related to each other:

$$\begin{aligned} f(x) \overset{(2.68)}{=} \sum_{i=0}^{\infty} \beta_i \mathcal{L}_i(x) \quad &\Rightarrow \quad \left\langle f \middle| \mathcal{L}_i \right\rangle = \beta_i \overset{(2.67)}{=} \alpha_i \\ &\Rightarrow \quad \left\langle f \middle| \vec{\mathcal{L}} \right\rangle = \left[\left\langle \mathcal{L}'_p \middle| \mathcal{L}'_q \right\rangle \right]^{-1} \left\langle f' \middle| \vec{\mathcal{L}}' \right\rangle \, . \end{aligned} \tag{2.69}$$

The optical path difference W is usually decomposed in the basis of Zernike polynomials, which are defined by [204, 205, 10]:

$$\begin{aligned} Z_n^{m(-1)^q}(\rho, \, \phi) := \sum_{k=0}^{(n-m)/2} &\frac{(-1)^k \cdot (n-k)!}{k! \cdot \left(\frac{n+m}{2} - k \right)! \cdot \left(\frac{n-m}{2} - k \right)!} \cdot \rho^{n-2k} \\ &\times \delta_{(n-m)\%2, \, 0} \cdot \begin{cases} \cos(m\phi) \, , & \text{if: } q = 0 \\ \sin(m\phi) \, , & \text{if: } q = 1 \end{cases} \, , \end{aligned} \tag{2.70}$$

and they reproduce the aberration pattern on the unit circle obeying the orthogonality relation [204, 205, 10]:

$$\langle Z_b^a, \, Z_d^c \rangle := \int_0^{2\pi} \int_0^1 Z_b^a \cdot Z_d^c \cdot \rho \, \mathrm{d}\rho \, \mathrm{d}\phi = \pi \cdot \delta_{p, \, q} \quad \text{, with:} \quad \begin{cases} p := \dfrac{b(b+2)+a}{2} \\ q := \dfrac{d(d+2)+c}{2} \end{cases} \, . \tag{2.71}$$

Hence, the scalar product vanishes on the unit circle for $q \neq p$. With the Zernike basis we can decompose any optical path difference map W into Zernike polynomials Z_n with corresponding Zernike coefficients ω_n (in units of the wavelength λ) as [15]:

$$W(\rho,\, \phi) = \sum_{n=0}^{\infty} \omega_n Z_n(\rho,\, \phi) \quad , \quad \omega_n \overset{(2.71)}{:=} \langle W,\, Z_n \rangle \ . \tag{2.72}$$

The individual Zernike coefficients can be related to independent optical effects like astigmatism or coma. But one has to be very cautious about the orthogonality argument. As an example, it is evident that the first three Zernike coefficients for piston and tilt do not alter the wavefront curvature, wherefore the perturbed MTF, defined by Equation (2.61), is congruent to the diffraction-limited MTF. Hence, the first three Zernike coefficients do not contribute to the blurring operator $\hat{\mathcal{B}}$ introduced in Figure 1.1 of Section 1.2. Nevertheless, this does not imply that the tilting operator $\hat{\mathcal{T}}$ is exclusively characterized by these Zernike coefficients. The modulation of the slope of the wavefront by the monoms of first order (x or y) determines physically the displacement of an object point in the image and generates distortion. As a result, the Zernike coefficient of first and second order are denoted as tilt but they are not characterizing the entire magnitude of the distortions present in the optical system. For example, the Zernike polynomials for coma (Z_7 and Z_8) – and potentially other higher-order terms that contain first order monoms – will contribute to the distortion of the optical system as well. In summary, while Zernike polynomials are orthogonal and independent on the unit circle by definition, the physical evaluation of specific optical aberrations typically includes several Zernike coefficients.

Finally, we want to list the Zernike polynomials up to the third radial order in Table 2.1.

Z_i	Zernike polynomial		Harmonic
	Polar coordinates	Cartesian coordinates	
Z_0	$Z_0^0 = 1$	1	✓
Z_1	$Z_1^{-1} = 2\rho \sin\phi$	$2y$	✓
Z_2	$Z_1^1 = 2\rho \cos\phi$	$2x$	✓
Z_3	$Z_2^{-2} = \sqrt{6}\rho^2 \sin 2\phi$	$2\sqrt{6}xy$	✓
Z_4	$Z_2^0 = \sqrt{3}(2\rho^2 - 1)$	$\sqrt{3}(2x^2 + 2y^2 - 1)$	×
Z_5	$Z_2^2 = \sqrt{6}\rho^2 \cos 2\phi$	$\sqrt{6}(x^2 - y^2)$	✓
Z_6	$Z_3^{-3} = \sqrt{8}\rho^3 \sin 3\phi$	$\sqrt{8}(3x^2y - y^3)$	✓
Z_7	$Z_3^{-1} = \sqrt{8}(3\rho^3 - 2\rho) \sin\phi$	$\sqrt{8}(3x^2y + 3y^3 - 2y)$	×
Z_8	$Z_3^1 = \sqrt{8}(3\rho^3 - 2\rho) \cos\phi$	$\sqrt{8}(3x^3 + 3xy^2 - 2x)$	×
Z_9	$Z_3^3 = \sqrt{8}\rho^3 \cos 3\phi$	$\sqrt{8}(x^3 + 3xy^2)$	✓

Table 2.1: Zernike polynomials up to the third radial order [204, 205, 10].

Table 2.1 also highlights which Zernike polynomials are harmonic. Harmonic functions are the set of functions that fulfill the Laplace equation, which is given by the time-independent Helmholtz Equation (2.3) in the case of $k = 0$. The task of finding the harmonic functions

which simultaneously satisfy a given boundary condition is denoted as the Dirichlet problem [33, 99, 113]. In two dimensions the general solution of the Laplace equation on a circular domain is given by [195]:

$$\psi(\rho,\ \phi) = \sum_{n=0}^{\infty} \psi_n(\rho,\ \phi) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} \rho^n \left(a_n \cos(n\phi) + b_n \sin(n\phi)\right)\ . \tag{2.73}$$

2.1.2 From theory to practice: the optical threat model

In this section, we will present an optical threat model, which is supposed to simulate the optical aberrations induced by the windshield based on the Fourier optical concepts introduced in Section 2.1. Inspired by the work of Chang et al. [25], we extend the proposed optical threat model to 4k ADAS cameras with telephoto objective lenses. Generally, the optical threat model assumes that the wavefront aberration map $W(\vec{x}_a)$ is known in advance either by measurement or optical simulation and is parameterized by a set of Zernike coefficients $\{\omega_n\}$, as introduced in Section 2.1.1.

With the knowledge of the wavefront aberration map of the windshield and the aperture stop of the camera under consideration, the generalized aperture function $\tilde{P}$ can be constructed by applying Equation (2.60). Based on $\tilde{P}$, the incoherent, non-diffraction-limited PSF is computed according to Equation (2.50), which entirely characterizes the optical system. The perturbed image is finally given by convolving the clean image with the perturbed PSF according to Equation (2.51). Furthermore, the entire ensemble of optical merit functions can be derived from the measured wavefront aberration map and characterizes the image quality from an optical perspective. In particular, the system MTF can be determined by applying Equation (2.61), while the associated areal measures, such as the Strehl ratio and OIG, can be evaluated by employing Equations (2.65) and (2.66), respectively. In addition to the Zernike coefficients, the optical threat model relies on the parameters of the camera system of interest. For the following illustrative example, we consider the OnSemi AR0820 imager and the Sunex DSL162A telephoto objective lens, which has a nominal focal length of $f = 15.3$ mm and an f-number of $f_n = 1.6$. Furthermore, the OnSemi AR0820 imager has a resolution of 3840×2160 px and the pixel pitch is specified with $\Delta x_{\text{CMOS}} = 2.1\ \mu$m. Moreover, we assume an effective wavelength of $\lambda = 550$ nm, as chromatic aberrations are not taken into account. These imaging characteristics lead to the PSF and MTF depicted in Figure 2.3 for the diffraction-limited case ($\omega_n = 0\ \forall\ n \in \mathbb{N}$).

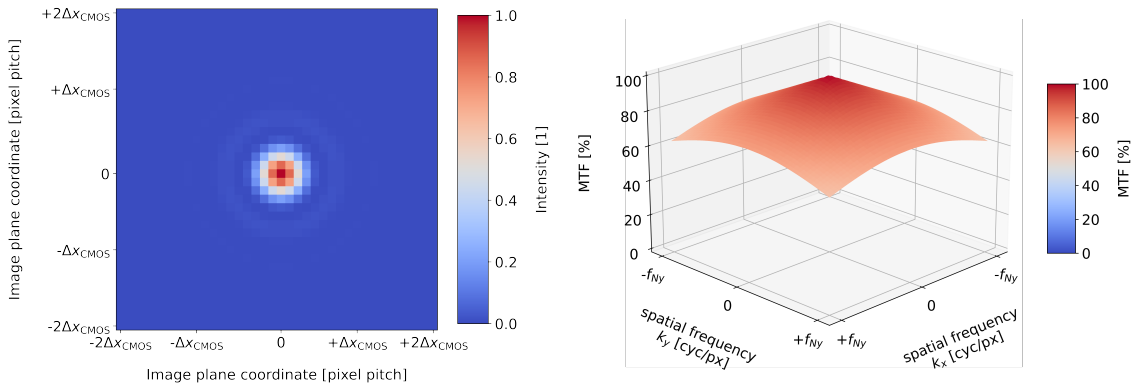


Figure 2.3: PSF (left) and MTF (right) for the diffraction-limited case.

If a windshield with typical aberrations ($\omega_3 = 0.051\ \lambda$, $\omega_4 = 0.542\ \lambda$, and $\omega_5 = -0.888\ \lambda$) is considered in the optical path, then the system PSF and the system MTF presented in Figure 2.4 result. The Zernike coefficients used in this toy example were obtained by a Shack-Hartmann wavefront measurement of a test sample windshield.

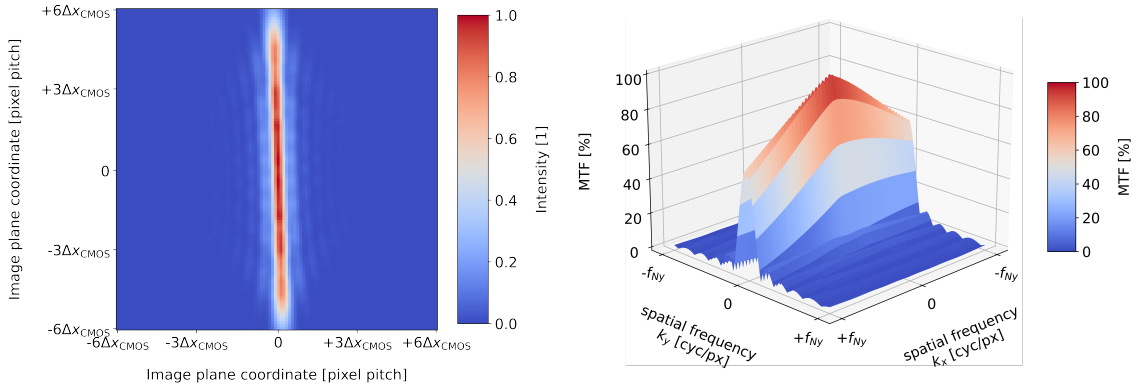


Figure 2.4: PSF (left) and MTF (right) for the aberration scenario: oblique astigmatism $\omega_3 = 0.051\ \lambda$, defocus $\omega_4 = 0.542\ \lambda$, and orthogonal astigmatism $\omega_5 = -0.888\ \lambda$.

Figure 2.5 demonstrates the effect of the PSF from Figure 2.4 on a test image.

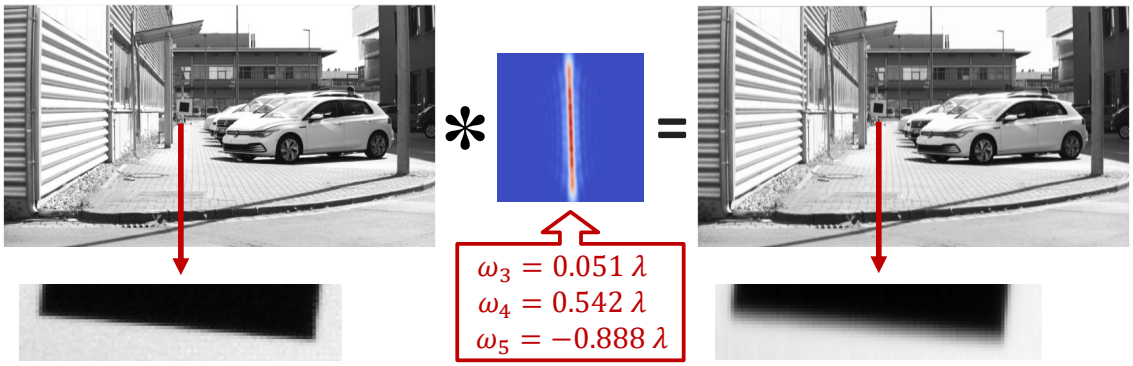


Figure 2.5: Toy example demonstrating the effect of the optical threat model applied to a real-world scene. The slanted edge targets are shown enlarged on the bottom.

The test image utilized in Figure 2.5 entails a black square target. This gives us the possibility to retrieve the MTF independently of the optical threat model by a slanted edge analysis according to ISO12233 [86]. The MTF of the aberrated image is normalized by the MTF of the unperturbed image to retrieve the net effect of the wavefront aberrations induced by the artificial windshield. This procedure is generally invalid as we will elaborate in detail in Chapter 4 but it holds true if the camera is diffraction-limited. While this assumption is applied in this test case, it is not generally valid for ADAS cameras. The normalized MTF curves for the horizontal and vertical direction are compared to the MTF parameterized by the optical threat model in Figure 2.6. In addition, the refractive power triggered by the curvature modulation of the wavefront can be evaluated by Equation (4.2), which has been benchmarked by a reference refractive power measurement using the Moiré pattern technique [192]. In conclusion, the measurement results for the physical test sample are sufficiently reflected by the optical threat model, which underpins the validity of the implemented Fourier optical approach.

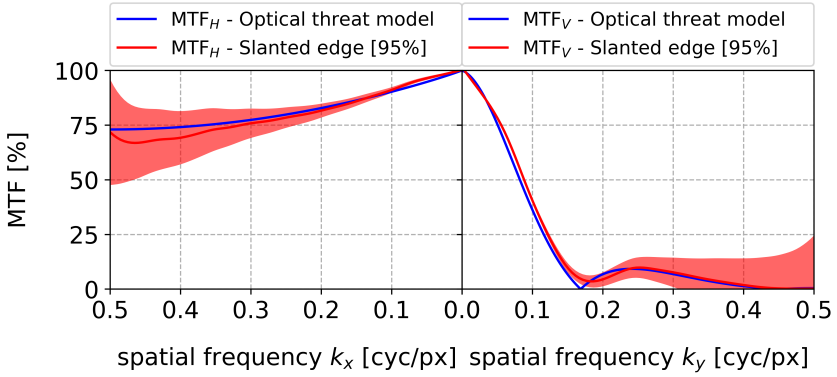


Figure 2.6: Validation of the optical threat model based on MTF measurements with the slanted edge method. The confidence bands are reflecting the Poisson noise of the image sensor.

Finally, it has to be emphasized that the optical threat model applied in this thesis is tuned for telephoto objective lenses. As a consequence, the paraxial approximation is assumed to hold true, which is not the case if wide-angle cameras are considered.

2.2 Specific features of windshields

In this Section, we want to briefly touch on the production process of windshields and introduce the Kerkhof model for describing surface imperfections of plane parallel glasses. The Kerkhof model is crucial for estimating the amplification of optical distortions under an additional tilt angle. This is important as the ADAS camera's perceptive field crosses the windshield under different angles within the field of view. On the other hand, conventional measurement systems operate with collimated light and typically measure the distortions

at the production facility under a tilt angle of $\varepsilon = 0°$. Hence, it is common practice to map these measurement results to the requirements within the windshield's camera window area by utilizing the Kerkhof model. In order to highlight the limitations of this procedure, we will evaluate the approximation error of the Kerkhof model in Subsection 2.2.3. Finally, we will employ the Kerkhof model in Chapter 3 for a comparability study of different refractive power measurement setups.

2.2.1 Material properties of automotive glass

The fabrication of a windshield starts by producing single-pane glass in the float glass process. Molten glass at a temperature of approximately 1100°C [143] is continuously poured onto a bath of molten tin, which has a much lower melting point of 232°C [103]. Due to its lower density [154, 3], the glass floats on the tin, forming a continuous layer of glass. The perfectly smooth surface of the molten tin ensures that the bottom side of the glass achieves an exceptional smoothness, unattainable through solid substrates. The molten glass cools down on its way through the tin bath and is then gradually brought to room temperature by being transported on roller conveyors passing through several cooling furnaces. Finally, the glass is cut to size.

To transform this flat glass into single-pane safety glass (also known as tempered glass [69]), the sheets are reheated and then rapidly cooled to induce internal stress, causing the glass to fracture into small, rounded fragments in case of an accident. However, this type of glass is not suitable for windshields because an impact, e.g. by a rock chip, would cause the entire pane to shatter instantly.

To address this issue, laminated safety glass is used for windshields, which consists of two congruently shaped non-tempered glass sheets, bonded together by a thin layer of polyvinyl butyral (PVB) [160]. The PVB interlayer provides elasticity and toughness, enhancing the durability of the glass. Unlike single-pane glass, glass fragments adhere to the PVB layer in case of an accident, which introduces a safety feature by preventing injuries.

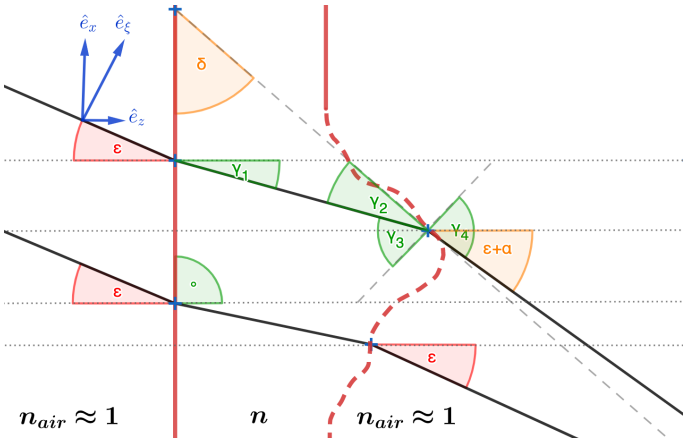
From theory to practice: the influence of the PVB foil

The magnitude of the influence of the PVB foil on the total wavefront aberration pattern of the windshield can be estimated by a back-of-the-envelope calculation. Suppose the thickness of the PVB foil is $\mu_{\mathrm{PVB}} = 0.8$ mm and the standard deviation of the PVB foil thickness is in the order of $\sigma_{\mu} = \pm 0.3\%$. Furthermore, we assume a refractive index for glass of $\mathrm{n_{glass}} = 1.5168$ [103] and for the PVB foil of $\mathrm{n_{PVB}} = 1.4850$ [105] if a wavelength of $\lambda = 550$ nm is considered. Under these assumptions, the standard deviation of the optical path difference induced by the PVB foil ($\mathrm{RMS_{PVB}}$) is approximately given by:

$\mathrm{RMS_{PVB}} = \left(\mathrm{n_{glass}} - \mathrm{n_{PVB}}\right) \sigma_\mu \mu_{\mathrm{PVB}} \approx 76.3\ \mathrm{nm} \mathrel{\widehat{=}} 0.14\ \lambda \ll 10\ \lambda$, which is a typical ballpark number for the Root Mean Square (RMS) optical path difference across the camera window area of a telephoto camera. As a consequence, optical aberrations induced by the glass layers most likely dominate over the contribution of the PVB foil.

2.2.2 Surface imperfections of plane parallel glasses - the Kerkhof model

In this section, we analyze the impact of surface imperfections on the ray propagation. The equations, which we derive by utilizing the concept of ray optics, have already been presented by Schardin [72] and Kerkhof [46]. Nevertheless, it is important to recall the underlying physical assumptions in order to understand the full picture behind the governing DIN-52305 industrial standard [40]. Figure 2.7 shows a sketch of a glass plate perturbed by surface imperfections. The angles can be derived by geometrical considerations and are given by Equation (2.74)-(2.78). In addition, Snell's law is adapted for the ray transition from air ($n_{\mathrm{air}} \approx 1$) to glass ($n_{\mathrm{glass}} \approx 1.52 =: n$) and vice versa.



$$\sin\varepsilon = n \cdot \sin\gamma_1 \tag{2.74}$$

$$\gamma_2 = \frac{\pi}{2} - \delta - \gamma_1 \tag{2.75}$$

$$\gamma_3 = \frac{\pi}{2} - \gamma_2 \tag{2.76}$$

$$n \cdot \sin\gamma_3 = \sin\gamma_4 \tag{2.77}$$

$$\varepsilon + \alpha = \gamma_4 - (\gamma_3 - \gamma_1) \tag{2.78}$$

Figure 2.7: Sketch of a plane parallel glass plate (red line) perturbed by surface imperfections (dashed red line). The incoming collimated beam (black line) is refracted by surface imperfections leading to a local focus effect.

The general idea of the Kerkhof model is to treat the perturbations of a glass plate as local prisms. Hence, the surface perturbations of the glass are translated into a map of prism wedge angles $\delta(x, y)$. For simplification, the two-dimensional problem is studied on the

condition $y \stackrel{!}{=} 0$ without loss of generality. The equations from Expression (2.74)-(2.78) can be used to derive the deflection angle α as a function of the wedge angle δ:

$$\begin{aligned}
\varepsilon + \alpha &\overset{(2.78,\,2.77,\,2.76)}{=} \arcsin\left(n \cdot \sin\gamma_3\right) - \left(\frac{\pi}{2} - \gamma_2 - \gamma_1\right) \\
\Leftrightarrow \varepsilon + \alpha &\overset{(2.76,\,2.75)}{=} \arcsin\left(n \cdot \cos\gamma_2\right) - \delta \\
\Leftrightarrow \varepsilon + \alpha + \delta &\overset{(2.75,\,2.74)}{=} \arcsin\left(n \cdot \sin\left(\arcsin\left(\frac{\sin\varepsilon}{n}\right) + \delta\right)\right) \\
\Leftrightarrow \sin\left(\varepsilon + \alpha + \delta\right) &= n \cdot \sin\left(\arcsin\left(\frac{\sin\varepsilon}{n}\right) + \delta\right) \\
\Leftrightarrow \sin\left(\varepsilon + \alpha + \delta\right) &= \sin\varepsilon \cdot \cos\delta + n \cdot \cos\left(\arcsin\left(\frac{\sin\varepsilon}{n}\right)\right) \cdot \sin\delta \\
\Leftrightarrow \sin\left(\varepsilon + \alpha + \delta\right) &= \sin\varepsilon \cdot \cos\delta + \sqrt{n^2 - \left(\sin\varepsilon\right)^2} \cdot \sin\delta \\
\Leftrightarrow \sin\left(\varepsilon + \alpha + \delta\right) &= \sin\varepsilon + \delta \cdot \sqrt{n^2 - \left(\sin\varepsilon\right)^2} + \mathcal{O}\left\{\delta^2\right\} \ .
\end{aligned} \tag{2.79}$$

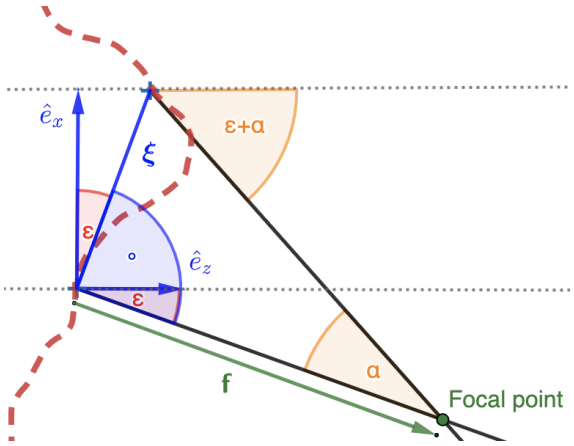
It is important to emphasize that the wedge angle δ quantifies the local imperfections of the glass surface and has an implicit spatial dependency. The deflection angle α is amplified if the optical system is tilted by the inclination angle ε. At this point, the assumption is made that the local wedge angle δ is small ($\delta \ll \pi/4$) and that the deflection angle α is small as well ($\alpha \ll \pi/4$). Hence, the last line of Expression (2.79) can be simplified:

$$\begin{aligned}
\sin\left(\varepsilon + \alpha + \delta\right) &= \sin\left(\varepsilon + \alpha\right) \cdot \cos\delta + \cos\left(\varepsilon + \alpha\right) \cdot \sin\delta \\
\Leftrightarrow \sin\left(\varepsilon + \alpha + \delta\right) &= \sin\varepsilon \cdot \cos\alpha \cdot \cos\delta + \cos\varepsilon \cdot \sin\alpha \cdot \cos\delta \\
&\quad + \cos\varepsilon \cdot \cos\alpha \cdot \sin\delta - \sin\varepsilon \cdot \sin\alpha \cdot \sin\delta \\
\Leftrightarrow \sin\left(\varepsilon + \alpha + \delta\right) &= \sin\varepsilon \cdot \cos\alpha + \cos\varepsilon \cdot \sin\alpha + \delta \cdot \cos\varepsilon \cdot \cos\alpha \\
&\quad - \delta \cdot \sin\varepsilon \cdot \sin\alpha + \mathcal{O}\left\{\delta^2\right\} \\
\Leftrightarrow \sin\left(\varepsilon + \alpha + \delta\right) &= \sin\varepsilon + \alpha \cdot \cos\varepsilon + \delta \cdot \cos\varepsilon - \alpha \cdot \delta \cdot \sin\varepsilon + \mathcal{O}\left\{\delta^2\right\} + \mathcal{O}\left\{\alpha^2\right\} \\
\Leftrightarrow \sin\left(\varepsilon + \alpha + \delta\right) &= \sin\varepsilon + \alpha \cdot \cos\varepsilon + \delta \cdot \cos\varepsilon + \mathcal{O}\left\{\left(\alpha + \delta\right)^2\right\} \ .
\end{aligned} \tag{2.80}$$

Due to the small-angle approximation, higher order terms $\mathcal{O}$ can be neglected. As a consequence, in first order approximation the deflection angle α is given by:

$$\begin{aligned}
\sin\varepsilon + \alpha \cdot \cos\varepsilon + \delta \cdot \cos\varepsilon &\overset{(2.79)}{\approx} \sin\varepsilon + \delta \cdot \sqrt{n^2 - \left(\sin\varepsilon\right)^2} \\
\Leftrightarrow \alpha &\approx \delta \cdot \left[\frac{1}{\cos\varepsilon} \cdot \sqrt{n^2 - \left(\sin\varepsilon\right)^2} - 1\right] \qquad \blacksquare \ .
\end{aligned} \tag{2.81}$$

This relationship determines the dependency of the local deflection angle $\alpha(x)$ on the local wedge angle $\delta(x)$. Furthermore, the local deflection angle can be translated to a local refractive power of the glass. In general, the refractive power quantifies the first derivative of the local deflection angle α with respect to the orthogonal off-axis coordinate ξ:



$$D := \frac{\partial \alpha}{\partial \xi} = \frac{\partial \alpha}{\partial \delta} \cdot \frac{\partial \delta}{\partial x} \cdot \frac{\partial x}{\partial \xi} =: \frac{1}{f} \,. \tag{2.82}$$

Figure 2.8: The local deflection angle α can be related to a local refractive power distortion D with focal length f.

Figure 2.8 depicts the relationship between α and ξ, which is described by the sine rule:

$$\frac{x}{\xi} = \frac{\sin\left(\frac{\pi}{2} + \alpha\right)}{\sin\left(\frac{\pi}{2} - \alpha - \varepsilon\right)} = \frac{\cos\alpha}{\cos\left(\alpha + \varepsilon\right)} \approx \frac{1}{\cos\varepsilon} \,. \tag{2.83}$$

As a result, the refractive power under an inclination angle ε, imposed by surface imperfections, is given by:

$$D_\varepsilon \overset{(2.81,2.82,2.83)}{\approx} \frac{1}{\cos\varepsilon} \cdot \left[\frac{1}{\cos\varepsilon} \cdot \sqrt{n^2 - (\sin\varepsilon)^2} - 1\right] \cdot \frac{\partial \delta}{\partial x} \,. \tag{2.84}$$

The spatial dependency in Equation (2.84) on the wedge angle profile can be eliminated if the ratio to the on-axis case is considered ($\varepsilon = 0°$). The resulting amplification factor for the refractive power of the perturbed glass depends exclusively on the inclination angle ε:

$$\frac{D_\varepsilon}{D_0} \overset{(2.84)}{\approx} \frac{1}{(n-1)\cdot\cos\varepsilon} \cdot \left[\frac{1}{\cos\varepsilon} \cdot \sqrt{n^2 - (\sin\varepsilon)^2} - 1\right] \quad \blacksquare \,. \tag{2.85}$$

As a result, we can map a refractive power measurement, taken under an inclination angle of $\varepsilon = 0°$, to an arbitrary field of view areal profile. Furthermore, Equation (2.85) paves the way for calibrating refractive power measurement devices over the entire range based on a single reference lens. By tracing back the reference lens to international standards, we can safeguard the validity of the calibration strategy and ensure that the fundamental traceability requirement for a valid measurement in metrology is fulfilled.

2.2.3 From theory to practice: the limitations of the Kerkhof model

The assumption of the Kerkhof model, given by Equation (2.85), requires plane parallel glass. Nevertheless, the Kerkhof model can be utilized to estimate the expected refractive power of a weak focusing lens under some tilting angle ε. In order to quantify the deviation from the Kerkhof model and to derive an approximation threshold for a weak

focusing lens of $\langle D \rangle = 100\ \mathrm{km}^{-1}$, we implemented a ray tracing simulation in Python. As an example, Figure 2.9 shows an ensemble of light rays refracted by the weak focusing lens under different inclination angles ε.

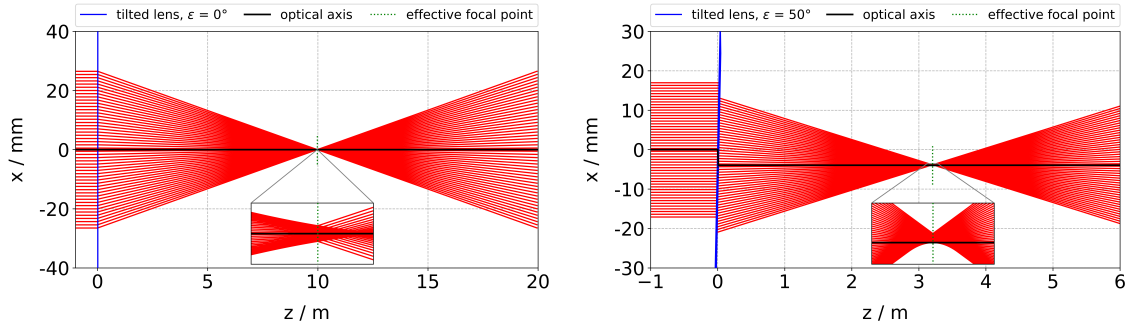
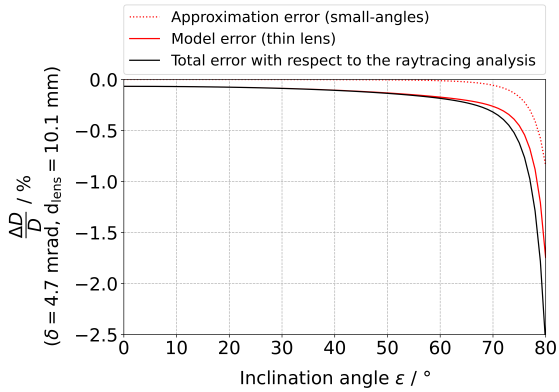
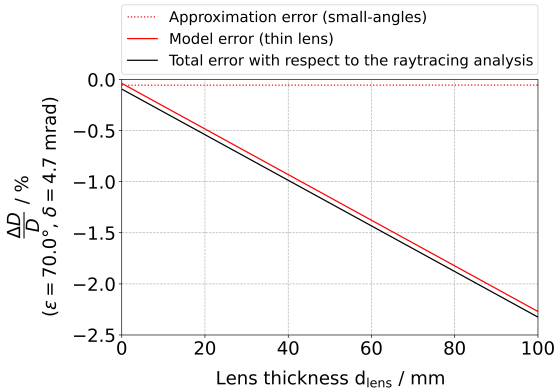


Figure 2.9: Ray tracing simulation for $\varepsilon = 0°$ (left) and $\varepsilon = 50°$ (right).

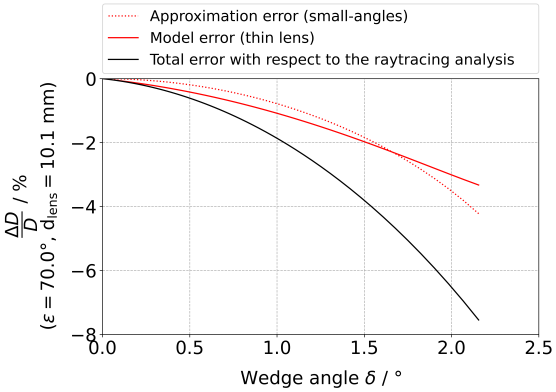
The point of least confusion determines the focal point of the lens for the corresponding inclination angle. As a side effect, the circle of least confusion increases as the inclination angle increases, which would result in an effective blurring if an imaging system is considered. This blurring can be quantitatively characterized by the MTF as discussed in Section 2.1. In addition to the inclination angle ε, there are two parameters that influence the effective refractive power of the lens. First of all, the thickness of the lens $\mathrm{d}_{\mathrm{lens}}$ has an impact because the lens surface is curved. Secondly, the wedge angle δ has to be small, such that second order terms in δ are negligible. Those three dependencies were studied in detail by executing the ray tracing simulation for different parameter settings. Figure 2.10a demonstrates that the systematic bias between the Kerkhof model and the physical simulation is less than 0.5% for $\varepsilon \leq 70°$. Typically, car windshields are inclined by less than $\varepsilon = 70°$, wherefore the Kerkhof model is sufficiently accurate for the scope of car windshield measurements. The thickness of the reference lens, produced and calibrated by ZEISS, is given by $\langle \mathrm{d}_{\mathrm{lens}} \rangle = 10.1$ mm. Thicknesses of up to 40 mm result in systematic biases of less than 1% according to Figure 2.10b. The wedge angle for the spherical lens is not constant over the entire profile of the aperture surface. Therefore, the maximum wedge angle δ was varied, which corresponds to different maximum aperture diameters. From Figure 2.10c can be concluded that the maximum aperture stop of the reference lens, which corresponds to a maximum wedge angle of $\delta = 4.7$ mrad, does not violate the assumptions of the Kerkhof model. Hence, the measurement bias is much smaller than 1% for a maximum inclination angle of $\varepsilon = 70°$. In conclusion, the Kerkhof model is valid within the refractive power measurement range of weak focusing lenses. Consequently, by periodically recalibrating the reference lens, refractive power measurement systems can be traced back to international standards and National Metrological Institutes (NMI), e.g. PTB, METAS, NIST.



(a) Exact solution versus first order, small angle approximation.



(b) Impact of lens thickness.



(c) Wedge angle dependency.

Figure 2.10: Robustness analysis of the Kerkhof model. For the parameterisation of the simulation, the worst case scenario has been considered, which is given at the edge of the reference lens aperture. The corresponding maximum wedge angle amounts to $\delta = 4.7$ mrad and the maximum inclination angle is assumed as $\varepsilon = 70°$.

2.3 The essential pillars of metrology

In general, there are three fundamental pillars for a valid measurement in metrology, namely reproducibility, comparability and traceability. The reproducibility quantifies the degree of agreement between the measurement results of experiments conducted at locations with different environmental conditions, different instrumentation, different parameterisation or different experimenter but the measurement method remains unaffected. The comparability assesses the degree of agreement between different measurements methods or principles. The traceability is essential for establishing a calibration chain and to trace back the measurement results to international standards. In other words, the reproducibility, comparability and traceability of the measurement results have to be demonstrated in order to guarantee for the scalability of industrial solutions and applications in a globalized world. Typically, the comparability is established by a round

robin test including different laboratories and the reference object under test is usually a primary standard.

2.4 Fundamentals of machine learning

In this Chapter, we want to introduce the essential concepts of machine learning to evaluate the impact of optical aberrations on the predictive performance of neural networks. The evaluations presented in Chapter 5 and Chapter 6 will require to study different neural network architectures in Section 2.4.1. These networks will be trained on different datasets, which are introduced in Section 2.4.2. Furthermore, the assessment of the predictive performance requires the definition of task-specific evaluation measures in Section 2.4.3. From a metrological perspective, a result is worth nothing without a proper uncertainty indication. This also holds true for neural-network-based predictive models, wherefore we will discuss the baseline principles of uncertainty quantification in machine learning in Section 2.4.4. As for uncertainty budgets in metrology, the confidence intervals assigned to the uncertainty estimates should match the empirical observation frequency. Unfortunately, standard uncertainty quantification measures in machine learning are biased, wherefore neural network uncertainty calibration methods are essential. The corresponding fundamentals of neural network calibration are introduced in Section 2.4.5. Finally, we want to analyze in this thesis what are the dominant aberrations influencing the prediction of a neural network. To tackle this task, we will employ local sensitivity measures from the explainable AI domain, as presented in Section 2.4.6.

2.4.1 Neural network architectures for semantic segmentation

In this work, we will exclusively consider CNN-based architectures, as they are intended to be employed in telephoto camera-based ADAS applications. Furthermore, we will focus on semantic segmentation as the primary objective of this thesis, since pixel-wise predictions are particularly sensitive to optical aberrations induced by the windshield, as the resulting blurring operator smooths out local image features.

Encoder-decoder structure

CNN-models for semantic segmentation typically consist of three key elements:

- An encoder, which is supposed to embed the input data into a latent space representation. According to the manifold hypothesis [48, 14, 123], the input data can be mapped onto a lower dimensional manifold with much fewer degrees of freedom. The general idea behind the manifold hypothesis is that not all input dimensions

are needed to describe the patterns within the data. As a consequence, it is the goal of the encoder to find a lower dimensional representation of the data set in a reduced space. A technical jargon for this reduced space is the latent space [122]. A very prominent dimensionality reduction technique is the Principle Component Analysis (PCA), which essentially does the same job by reducing the input space to the principle axes, which are distinguished by the highest variance. Hence, PCA is a linear encoder. To include nonlinear mappings on arbitrarily shaped manifolds we will consider a neural network with non-linear activation functions. The ideal case we are aiming for is to find a latent space with a dimensionality equivalent to the degree of freedom of the underlying system, which generates the data. If we make use of the terminology in Lagrangian mechanics, then a perfect encoder would be able to map the input data into the generalized coordinates [60] of the system.

- The bottleneck, which characterizes the part of the network where the information density is the highest and the data is represented in the latent space.
- An decoder, which maps from the latent space manifold to the target space. In the case of semantic segmentation, the task is to predict a 2D image of the same size as the input image containing pixel-wise classifications.

Subsequently, we will introduce three different neural network architectures, used within this work, which utilize the encoder-decoder structure. All these models will employ the cross-entropy loss function [5] for training due to its widespread use for semantic segmentation [180, 201, 27, 208].

UNET

The UNET architecture, proposed by Ronneberger et al. [149], enhances the preservation of local spatial information and alleviates the problem of vanishing gradients in the encoder-decoder structure, which arises from the multiplicative nature of the loss gradient across the layers. For this reason, skip connections are introduced between the encoder and the decoder for layers that share the same spatial fidelity. Hence, local features embedded on manifolds with different dimensionality are concatenated with the corresponding decoder reconstruction, which implicitly enhances the sensitivity for small scale patterns. Mathematically, the skip connection describes an identity mapping, which has a derivative of one. Hence, the total loss gradient is composed of several additive terms that relate to the gradient accumulation across different network parts, which reduces the risk of vanishing gradients for the total loss. This idea of reducing the effective path length for backpropagation in the UNET layout is similar to the underlying principle of the ResNet [75] architecture, which was initially introduced for image classification.

High-Resolution Network (HRNet)

The High-Resolution Network (HRNet) architecture, invented by Microsoft [179, 124, 2], builds upon the UNET architecture by increasing the information capacity of the model. This is a tempting approach for modern ADAS functionalities, which rely on 4k high-resolution cameras, because microscale patterns generated by small, far away objects (e.g. Euro-pallet in 300m distance on the highway) might indicate a safety-critical scenario. In general, a model architecture can be tuned in three dimensions: depth (e.g., more layers), wideness (e.g., more channels) or finesse (e.g., higher resolution images). The standard sequential encoder-decoder architecture in deep convolutional neural networks lacks on information capacity in the condensed low-resolution feature map. Hence, the standard encoder-decoder architecture is typically extended for highly spatially sensitive applications like autonomous driving. The HRNet tackles this challenge by switching the information propagation from serial to parallel [179]. In detail, convolutions are performed in parallel on multiple resolutions to improve the information capacity of the model architecture. Therefore, the high-resolution representation of the input information is maintained throughout the whole process. Repeated fusion steps between parallel streams of different resolutions ensure an information flow across the levels. As a consequence, the HRNet enhances the information capacity w.r.t. the UNET architecture, wherefore a superior performance is expected if microscale data patterns are of particular relevance.

Multi-Task Learning (MTL) model

The encoder-decoder structure can also be employed to tackle several tasks simultaneously. In order to do so, the encoder is trained on the entire data set over all tasks to find the most effective latent space manifold. Subsequently, the individual decoder heads are fine-tuned on the task-specific data sets, while the encoder weights are kept constant. The in-house developed Multi-Task Learning (MTL) model, which is utilized within this work as a black-box model, consists of a large shared encoder with several feature extraction layers followed by five decoder heads, each for a specific task, referred to as task head. These task heads are mainly of two types: segmentation heads and object detection heads. In detail, the parallelized decoders address the following tasks:

- **Semantic segmentation head:** Provides a pixel-wise classification across the image for several classes. The head's performance is quantified by the mIoU, defined in Equation (2.86).
- **Blockage detection head:** Provides a binary segmentation mask that detects if a certain region of the image is blocked or not. The identification of obstructed image pixels, which are thus invalid for extracting information in the segmentation head, functions as a built-in safety mechanism within the given MTL architecture.

The evaluation metric is given by the IoU, which is the single-class version of the mIoU (see Equation (2.86)).

- **Traffic Light Detection and Classification:** At first, 2D bounding boxes for traffic lights in the image are predicted. Subsequently, the pixels within a single 2D bounding box are segmented to either belong to the class "traffic light bulb" or "housing". Finally, the pixels belonging to the class "traffic light bulb" are used to classify the signal color of the corresponding traffic light. For quantifying the performance of this multi-step classification task, a head-specific combined metric is evaluated, which relies, i.a., on the average accuracy and the area under the precision-recall curve.
- **Traffic Sign Detection and Classification:** Predicts 2D bounding boxes for traffic signs within the image. Afterwards, the subimages are used to classify the corresponding traffic sign type. Similar to the traffic light classification head a combined metric is assessed for the performance of the multi-step traffic sign classification task.
- **Vehicle Detection:** Provides a categorized 3D bounding box across two types of vehicles: large vehicles (e.g., trucks, buses, etc.) and passenger cars. The head's performance is evaluated by the mean average precision metric, defined in Equation (2.90).

For a consolidated evaluation, we first determine the task-specific metrics (i.e., average precision for object detection and mIoU for semantic segmentation). However, to convey a holistic model performance, the head-specific loss functions are integrated using weighted averaging after normalization culminating into an overall combined multi-task loss, ranging from 0 (worst performance) to 1 (perfect performance). This aggregated score reflects the model's collective efficiency across all tasks. In order to prevent a single task from being dominant in the learning process, the individual, task-specific head losses can be integrated by an uncertainty-based weighting scheme to obtain a more robust combined metric [28].

2.4.2 Datasets

Typically, datasets for autonomous driving are recorded by cameras mounted behind the windshield. As a consequence, the images are inherently perturbed by optical aberrations, which leads to an unknown dataset domain shift that makes it impossible to quantitatively assess the impact of different windshield configurations without prior knowledge about the inherent aberrations. Within this thesis we will utilize three different datasets, which all have in common that the camera was mounted on the car roof:

- KITTI dataset from the KIT [55] with an unified resolution of 1224×370 px. The images were taken by a Point Grey Flea 2 (FL2-14S3C-C) camera with an Edmund Optics (NT59-917) objective lens [55]. The camera is parameterized by a focal length of $f = 8$ mm and an f-number of $f_n = 8$ [25]. Furthermore, the pixel pitch of the CMOS sensor was given by $\Delta x_{\text{CMOS}} = 4.65\ \mu$m.
- Multi-task dataset from CARIAD [194] with a resolution of 2048×1024 px. The data set was recorded by a camera with a focal length of $f = 5.04$ mm and an f-number of $f_n = 1.46$. The images were captured by a CMOS sensor with a pixel pitch of $\Delta x_{\text{CMOS}} = 2.1\ \mu$m.
- A2D2 dataset from AUDI [56] with a resolution of 1920×1208 px. The camera employed was the Sekonix SF3325-100 with the Sekonix NA6062 objective lens [56]. This camera is characterized by a focal length of $f = 5.49$ mm and an f-number of $f_n = 1.8$. Finally, the image is digitalized by the Onsemi AR0231 CMOS sensor with a pixel pitch of $\Delta x_{\text{CMOS}} = 3\ \mu$m.

2.4.3 Evaluation measures

Mean Intersection over Union (mIoU)

The governing benchmark metric for semantic segmentation is given by the Jaccard similarity coefficient or also commonly known as the Intersection over Union (IoU) [126]. In this thesis we will make use of multi-class segmentation datasets, wherefore the mean of the IoU is computed over all classes (N_c), denoted as mIoU:

$$\text{mIoU} := \frac{1}{N_c} \sum_{i=0}^{N_c} \frac{\text{G}_i \cap \text{P}_i}{\text{G}_i \cup \text{P}_i} \hat{=} \frac{1}{N_c} \sum_{i=0}^{N_c} \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i + \text{FN}_i} \,. \tag{2.86}$$

Here, G_i denotes the set of image pixels corresponding to the ground truth labels for class i, and P_i indicates the set of image pixels predicted for class i. In detail, the number of class-wise true positive predictions (TP_i) are proportioned to the total number of predictions within the cross-section domain composed by TP_i, the class-wise false positive predictions FP_i as well as the class-wise false negative predictions FN_i.

At this point, it should be emphasized that the ground truth labels are assumed to be perfect within this thesis, which ignores effects like labeling noise [51].

Mean Average Precision (mAP)

The standard benchmark metric for object detection is given by the mean Average Precision (mAP) [111]. In order to compute the mAP, we first have to define the

Precision (P), which measures the fraction of correct predictions relative to the total number of detected objects for a given class i, conditioned on a predefined confidence level τ used to classify the predictions as positive. Essentially, a positive prediction for class i is given if the predicted confidence $p_i(\hat{\boldsymbol{y}}|\hat{\boldsymbol{x}}) \geq \tau$. Mathematically, the Precision (P) is defined as [111]:

$$\mathrm{P}_i(\tau) := \frac{\mathrm{TP}_i(\tau)}{\mathrm{TP}_i(\tau) + \mathrm{FP}_i(\tau)} \,. \tag{2.87}$$

Furthermore, we need to introduce the Recall (R), which quantifies the fraction of correct predictions relative to the total number of ground truth instances for a given class i, also conditioned on a predefined confidence level τ. Mathematically, the Recall (R) is defined as [111]:

$$\mathrm{R}_i(\tau) := \frac{\mathrm{TP}_i(\tau)}{\mathrm{TP}_i(\tau) + \mathrm{FN}_i(\tau)} \,. \tag{2.88}$$

With the precision and the recall we can generate the precision-recall curve. Each point on the curve corresponds to a different confidence level τ used to classify predictions as positive. As τ decreases the recall increases and the precision typically drops. The Area Under the precision-recall Curve ($\mathrm{AUC}_i^{(\mathrm{PR})}$) can be determined by:

$$\mathrm{AUC}_i^{(\mathrm{PR})} := \int_0^1 \mathrm{P}_i(\mathrm{R}_i)\, \mathrm{d}\mathrm{R}_i = -\int_0^1 \mathrm{P}_i(\mathrm{R}_i(\tau))\, \frac{\mathrm{d}\mathrm{R}_i(\tau)}{\mathrm{d}\tau}\, \mathrm{d}\tau \,. \tag{2.89}$$

Since the derivative of the recall with respect to the confidence level is non-positive, the area under the curve, $\mathrm{AUC}_i^{(\mathrm{PR})}$, is strictly positive.

If we perform this analysis of the precision-recall curve for each class i, then we can compute the mean Average Precision (mAP) as [111]:

$$\mathrm{mAP} := \frac{1}{N_c} \sum_{i=0}^{N_c} \mathrm{AUC}_i^{(\mathrm{PR})} \,. \tag{2.90}$$

As the integral in Equation (2.89) is generally intractable in closed form, it is commonly approximated based on a discrete set of samples drawn from the precision-recall curve. In this case, the $\mathrm{AUC}_i^{(\mathrm{PR})}$ is denoted as the Average Precision (AP).

2.4.4 Uncertainty quantification in machine learning

The goal of uncertainty quantification in machine learning is to characterize the conditional probability density function (pdf) $\mathcal{P}(\hat{\boldsymbol{y}}|\hat{\boldsymbol{x}})$ of the predicted neural network output $\hat{\boldsymbol{y}}$ given the instance input $\hat{\boldsymbol{x}}$. Mathematically, this conditional pdf is characterized by marginalizing over the training data set $\{\boldsymbol{x},\ \boldsymbol{y}\}$ and the weights $\boldsymbol{w}$:

$$\mathcal{P}(\hat{\boldsymbol{y}}|\hat{\boldsymbol{x}}) = \iiint \mathcal{P}(\hat{\boldsymbol{y}}|\boldsymbol{w}, \boldsymbol{x}, \boldsymbol{y}, \hat{\boldsymbol{x}}) \cdot \mathcal{P}(\boldsymbol{w}|\boldsymbol{x}, \boldsymbol{y}) \cdot \mathcal{P}(\boldsymbol{y}|\boldsymbol{x}) \cdot \mathcal{P}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{y}\, \mathrm{d}\boldsymbol{w} \,. \tag{2.91}$$

Very fundamentally, the total predictive uncertainty of a neural network can be decomposed into an aleatoric and an epistemic uncertainty contribution [80].

The aleatoric uncertainty contribution arises if a non-deterministic prediction problem or noise is considered. In this case, the second central moment of the conditional pdf $\mathcal{P}(\boldsymbol{y}|\boldsymbol{x})$ is non-zero, which results in an irreducible uncertainty contribution to the total prediction uncertainty. On the other hand, if the underlying mapping from the input $\boldsymbol{x}$ to the output $\boldsymbol{y}$ is given by an arbitrary function and there is no noise in the data generation process, then the conditional distribution $\mathcal{P}(\boldsymbol{y}|\boldsymbol{x})$ collapses to a Dirac-delta distribution and the aleatoric uncertainty contribution annihilates. This also indicates that the aleatoric uncertainty contribution is reducible if the data quality in terms of the signal to noise ratio is enhanced. The aleatoric uncertainty component can be determined by learning the conditional distribution $\mathcal{P}(\boldsymbol{y}|\boldsymbol{x})$ of the target output quantity $\boldsymbol{y}$ across the feature domain $\boldsymbol{x}$. To do so, one possibility is to extend the output space, such that a pdf is learned instead of a point-wise estimate $\hat{\boldsymbol{y}}$. Commonly, the output pdf is approximated as a Gaussian pdf, wherefore the corresponding standard deviation serves as a measure of the aleatoric uncertainty component. The parameters of this Gaussian pdf can be obtained by maximizing the likelihood of the pdf parameters given the observed data, which is equivalent to minimizing the negative log-likelihood. As an alternative post-hoc procedure, data augmentation has been proposed by Wang et al. [177] to estimate the aleatoric uncertainty component. This method does not require architectural changes as the conditional distribution $\mathcal{P}(\boldsymbol{y}|\boldsymbol{x})$ is approximated by Monte-Carlo sampling from prior distributions on the parameters of an image acquisition model, which considers image transformations and noise corruptions.

The epistemic uncertainty contribution characterizes the approximation uncertainty, which arises from the difference between the infimum in the loss landscape and the local minimum, where the neural network training converged given a certain training data set $\{\boldsymbol{x},\ \boldsymbol{y}\}$. Generally, the epistemic uncertainty contribution is reducible by increasing the sampling rate in the dataset manifold under constant signal to noise ratio. This can be achieved by simply capturing more training data [80] or by isentropically augmenting the existing dataset via geometrical transformations for example [156]. Mathematically, the epistemic uncertainty contribution is quantified by the conditional pdf $\mathcal{P}(\boldsymbol{w}|\boldsymbol{x},\boldsymbol{y})$. Bayesian neural networks include this conditional probability into the network architecture by assigning a pdf to each individual weight between two neurons [118]. Learning a distribution on the weights is similar to having an infinite ensemble of networks. During inference, each of the weights is sampled from their corresponding pdf. The epistemic uncertainty contribution can then be estimated by performing several forward passes, which leads to the posterior distribution. In order to learn the prior weight distributions, Bayes rule can be employed. It paves the way for iteratively updating the prior according to the observed training data. The posterior could be analytically calculated if the data likelihood (aka. evidence) is known. Unfortunately, the data likelihood requires the

integration over all possible network parameterizations rendering the integral to analytical unsolvability. Therefore, the posterior distribution is commonly approximated. There are two main approximation techniques, which are variational methods and sampling-based methods.

The idea of variational inference is to learn a simpler distribution with a new set of parameters (variational parameters) to approximate the true posterior distribution [7]. In order to do so, the Kullback-Leibler divergence between the posterior distribution – utilizing the simpler distribution for the weights – and the prior distribution serves as an additional regularization loss term. The total loss is then denoted as the Evidence Lower Bound (ELBO). Unfortunately, the approximation method is associated with a high computational cost because twice as many parameters must be trained in the baseline scenario of a Gaussian prior distribution.

Monte-Carlo Dropout [52] is a promissing alternative to variational inference as it does not require to introduce additional trainable parameters, which reduces the computational cost significantly [52]. It works as the standard Dropout technique for regularization [157], where units from the network are randomly dropped, but it is employed during inference instead of training. This results in several sub-networks of the original network [52]. By sampling several times with the same input instance, an approximation of the Bayesian posterior distribution can be reconstructed. The second central moment of this approximated distribution directly estimates the epistemic uncertainty contribution [135] of the total predictive uncertainty.

Fundamentally, Monte-Carlo Dropout can also be understood as an ensemble of different networks with shared weights. This gives rise to the idea of Deep Ensembles, where an ensemble of congruent network architectures are independently trained with distinct initialization [107]. Hence, a set of different optimal weight tensors is obtained, which leads to a distribution over the prediction result for the same instance. This distribution approximates the Bayesian posterior distribution [107] in a similar fashion as Monte-Carlo Dropout. Finally, the combination of Deep Ensembles and Monte-Carlo Dropout offers an efficient way to increase the sample size for a more robust approximation of the Bayesian posterior distribution [161].

Nevertheless, a major downside of the Deep Ensembles approach is the computational overhead generated during training and inference by the utilization of several models that sample the hypothesis space [80]. This problem has been addressed recently in the work of Landgraf et al. on Deep Uncertainty Distillation using Ensembles for Semantic Segmentation (DUDES) [109]. They propose a student-teacher distillation framework where the Deep Ensemble model, referred to as the teacher, is used to guide a less complex model, known as the student, in estimating the ensemble-based predictive uncertainties. This significantly reduces the inference time because only one forward pass is required and the information about the posterior distribution is distilled into the student neural network,

wherefore the calibration quality of the predictive uncertainty estimate is maintained. This concept has been further extended by the student-teacher distillation framework for efficient multi-task uncertainties, referred to as EMUFormer [108]. They employ a Deep Ensemble of transformer-based multi-task networks for semantic segmentation and monocular depth estimation (termed as SegDepthFormer) to evaluate the predictive uncertainty. The backbone of this methodology is the idea of enhancing the generalization capabilities of a neural network by multi-task learning. With the SegDepthFormer architecture they demonstrated that this idea can be transformed to the network calibration and the results indicate less biased predictive uncertainty estimates in terms of the mean Expected Calibration Error (mECE, defined in Section 2.4.5) if multi-task learning is employed.

The superiority of sampling-based methods (particularly Deep Ensembles) over variational methods in terms of the uncertainty calibration error can be explained by the loss landscape [49]. By utilizing an ensemble of neural networks, it is possible to explore several valleys within the loss landscape, wherefore the approximated Bayesian posterior distribution contains information about several modes in the hypothesis space [80]. In contrast, variational methods capture the local uncertainty around a single mode as the prior assumption of a Gaussian distribution limits the correlation length.

The implicit limitation of the hypothesis space by the selection of a particular neural network architecture gives rise to a third uncertainty component, namely the model uncertainty. The model uncertainty is of epistemic nature as the approximation uncertainty because it is reducible by considering more sophisticated neural network architectures. Due to the implicit limitation of the hypothesis space it is even possible that the groundtruth functional relationship, underlying the data generation process, is unachievable by the neural network even in the case of infinite training data. Mathematically, we can account for the model uncertainty by marginalizing the conditional pdf $\mathcal{P}(\hat{\boldsymbol{y}}|\hat{\boldsymbol{x}})$ over the hypothesis space h:

$$\mathcal{P}(\hat{\boldsymbol{y}}|\hat{\boldsymbol{x}}) = \iiiint \mathcal{P}(\hat{\boldsymbol{y}}|h,\boldsymbol{w},\boldsymbol{x},\boldsymbol{y},\hat{\boldsymbol{x}}) \cdot \mathcal{P}(\boldsymbol{w}|h,\boldsymbol{x},\boldsymbol{y}) \cdot \mathcal{P}(\boldsymbol{y}|\boldsymbol{x}) \cdot \mathcal{P}(\boldsymbol{x}) \cdot \mathcal{P}(h)\, \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{y}\, \mathrm{d}\boldsymbol{w}\, \mathrm{d}h\,. \quad (2.92)$$

The distribution $\mathcal{P}(h)$ over the hypothesis space is sometimes referred to as a second order distribution w.r.t. the outcome $\hat{\boldsymbol{y}}$ [80]. In the simplest case, second order distributions describe the uncertainty in the parameter estimates that parameterize the pdf of an underlying random variable [21]. As an example, the sample variance represents an unbiased estimator for the true variance of a random variable. But this estimation also comes with an uncertainty, wherefore the sample variance itself underlies a pdf [192]. In detail, according to the Cochran theorem [30], the normalized sample variance follows a χ^2_ν distribution with ν degrees of freedom. In a nutshell, second order random variables generate distributions of distributions in order to reflect the uncertainty in the quantification of the parameterization of the underlying pdf. Now, in the context of

the hypothesis space h, this definition of second order distributions can be generalized to assign a probability to each possible realization of the parameter space described by the conditional pdf $\mathcal{P}(\boldsymbol{w}|h,\boldsymbol{x},\boldsymbol{y})$. This implies that the conditional pdf $\mathcal{P}(\boldsymbol{w}|h,\boldsymbol{x},\boldsymbol{y})$ is not confined to a multivariate distribution with a single parameter vector. Instead, the parameters itself are considered as distributions in the hypothesis space, which accounts for the uncertainty in selecting the right type of model based on prior background knowledge [80]. Due to the high information capacity of deep neural networks, the approximation uncertainty generally dominates over the model uncertainty [80], which reflects the universal approximation capabilities of deep neural networks theoretically motivated by the Weierstrass approximation theorem [182].

In summary, the total predictive uncertainty, composed of the aleatoric uncertainty contribution generated by the spread of the conditional pdf $\mathcal{P}(\boldsymbol{y}|\boldsymbol{x})$ and the epistemic uncertainty component induced by the non-zero width of the conditional pdf $\mathcal{P}(\boldsymbol{w}|\boldsymbol{x},\boldsymbol{y})$, can be determined by applying the law of total variance [183]:

$$\mathrm{VAR}\left[\hat{\boldsymbol{y}}\right] = \mathrm{VAR}\left[\mathrm{E}\left[\hat{\boldsymbol{y}} \mid \hat{\boldsymbol{x}}\right]\right] + \mathrm{E}\left[\mathrm{VAR}\left[\hat{\boldsymbol{y}} \mid \hat{\boldsymbol{x}}\right]\right] \ . \tag{2.93}$$

Here, the aleatoric uncertainty component $\mathrm{E}\left[\mathrm{VAR}\left[\hat{\boldsymbol{y}} \mid \hat{\boldsymbol{x}}\right]\right]$ can be estimated by predicting the variance $\mathrm{VAR}\left[\hat{\boldsymbol{y}} \mid \hat{\boldsymbol{x}}\right]$ alongside the target mean $\mathrm{E}\left[\hat{\boldsymbol{y}} \mid \hat{\boldsymbol{x}}\right]$. By averaging this variance estimate over several forward passes with the same input instance $\hat{\boldsymbol{x}}$, we can finally approximate the expected aleatoric uncertainty component $\mathrm{E}\left[\mathrm{VAR}\left[\hat{\boldsymbol{y}} \mid \hat{\boldsymbol{x}}\right]\right]$. Furthermore, the epistemic uncertainty component $\mathrm{VAR}\left[\mathrm{E}\left[\hat{\boldsymbol{y}} \mid \hat{\boldsymbol{x}}\right]\right]$ can be estimated by sampling several times from the weight distribution $\mathcal{P}(\boldsymbol{w}|\boldsymbol{x},\boldsymbol{y})$ in a Bayesian manner and by computing the variance across the set of predictions drawn from the posterior distribution. As a result, the sum of both uncertainty components provides us a statistically sound estimate for the total predictive uncertainty of the neural network [92].

So far, we have analyzed the total predictive uncertainty by computing the conditional pdf $\mathcal{P}(\hat{\boldsymbol{y}}|\hat{\boldsymbol{x}})$, which entirely determines the statistical behavior of the predicted output $\hat{\boldsymbol{y}}$ given an input instance $\hat{\boldsymbol{x}}$. From a metrological point of view, we would like to map the conditional pdf $\mathcal{P}(\hat{\boldsymbol{y}}|\hat{\boldsymbol{x}})$ to a scalar uncertainty value, which is typically done by evaluating the second central moment that quantifies the width of the distribution. But this procedure already makes implicit assumptions about the shape of the pdf (in particular single mode, Gaussian like), which is not guaranteed for highly non-linear functions modeled by neural networks. As a consequence, the total predictive uncertainty estimates from the Bayesian approach might be biased, which results in high uncertainty calibration errors. Hence, the mapping from the conditional pdf $\mathcal{P}(\hat{\boldsymbol{y}}|\hat{\boldsymbol{x}})$ to a scalar uncertainty measure should ideally account for the entire shape of the pdf. The Shannon entropy [152] provides a corresponding measure for it. The Shannon entropy is defined as the expected Shannon information and quantifies the information gain obtained by knowing the pdf relative to the case where no information about the system is known a priori (uniform distribution). In addition, the Shannon entropy has important mathematical properties: non-negativity,

reaching a maximum for the case of a uniformly distributed $\mathcal{P}(\hat{\boldsymbol{y}}|\hat{\boldsymbol{x}})$ and invariance under permutations in the label space $\mathcal{Y}$ [34]. The Shannon entropy $\mathcal{H}$ in the discrete case (normalized by $\log(K)$ - which is the entropy for a uniform distribution over K classes) is defined for each prediction instance as:

$$\mathcal{H}\left[p(\hat{\boldsymbol{y}}|\hat{\boldsymbol{x}})\right] := -\frac{1}{\log\left(K\right)} \sum_{i=1}^{N} p_i(\hat{\boldsymbol{y}}|\hat{\boldsymbol{x}}) \cdot \log\left(p_i(\hat{\boldsymbol{y}}|\hat{\boldsymbol{x}})\right) \ . \tag{2.94}$$

Here, $p_i(\hat{\boldsymbol{y}}|\hat{\boldsymbol{x}})$ denotes the i-th sample drawn from the conditional probability mass function (pmf) $p(\hat{\boldsymbol{y}}|\hat{\boldsymbol{x}})$, which is the discretized version of the conditional pdf $\mathcal{P}(\hat{\boldsymbol{y}}|\hat{\boldsymbol{x}})$.

From an information theory point of view, the total predictive uncertainty, quantified by the Shannon entropy, can also be decomposed into the aleatoric and the epistemic contribution [38, 80, 178]:

$$\begin{aligned} \mathcal{H}\left[\mathcal{P}(\hat{\boldsymbol{y}}|\hat{\boldsymbol{x}})\right] = &\int \mathcal{H}\left[\mathcal{P}(\hat{\boldsymbol{y}}|\boldsymbol{w},\hat{\boldsymbol{x}})\right] \cdot \mathcal{P}(\boldsymbol{w}) \, \mathrm{d}\boldsymbol{w} \\ &+ \int \mathcal{D}_{\mathrm{KL}}\left[\mathcal{P}(\hat{\boldsymbol{y}}|\boldsymbol{w},\hat{\boldsymbol{x}})||\mathcal{P}(\hat{\boldsymbol{y}}|\hat{\boldsymbol{x}})\right] \cdot \mathcal{P}(\boldsymbol{w}) \, \mathrm{d}\boldsymbol{w} \ , \end{aligned} \tag{2.95}$$

as done for the total variance in Equation (2.93). The first term denotes the expected conditional entropy and describes the aleatoric uncertainty contribution, which refers to the stochasticity in the underlying data-generating process. The second term is determined by the expected Kullback-Leibler divergence and characterizes the epistemic uncertainty component, which quantifies the information gain w.r.t. the model weights $\boldsymbol{w}$ if the unconditional distribution would have been known a priori. The model uncertainty component introduced in Equation (2.92) can also be accounted for in the decomposition of the Shannon entropy as a third term, which is referred to as the Bregman decomposition [128]. Since the groundtruth functional relationship between $\boldsymbol{x}$ and $\boldsymbol{y}$ is not known a priori, the Kullback-Leibler divergence between the groundtruth and the posterior distribution needs to be approximated to quantify the bias. We will propose an alternative strategy to estimate this bias term by utilizing the AUSE methodology in Chapter 8.

Unfortunately, recent studies [128, 188, 67] showed evidence on the violation of the disentanglement between the deduced uncertainty contributions (aleatoric and epistemic), which indicates an inherent inconsistency in the above-mentioned methods.

If semantic segmentation is the target application – as it is the case in this thesis – then quantifying the total predictive uncertainty of the neural network by the Shannon entropy might be questionable. The argument against utilizing the Shannon entropy can be demonstrated by a thought experiment. Consider a classification problem with three different classes. If the softmax output assigns 95% of the probability mass to the most likely outcome then, according to the Shannon entropy, the total predictive uncertainty lies in the range of [18.1%, 21.2%]. The degree of freedom in this setup lies in the allocation of the probability mass over the two remaining classes. If the trainable neural

network parameters are determined by minimizing the negative log-likelihood, which is equivalent to minimizing the cross-entropy loss during training, then only the softmax output regarding the groundtruth label is considered. The distribution of the remaining probability mass over the wrong classes does not influence the loss at all. This gives rise to the idea of exclusively utilizing the maximum softmax output as a measure of prediction confidence and the complement as an uncertainty measure respectively, as defined by the variation ratio (V) [117], given by:

$$V := 1 - \max\, p(\hat{\boldsymbol{y}}|\hat{\boldsymbol{x}}) \; . \tag{2.96}$$

By doing so, the dependency on the probability mass distribution over the remaining classes is lifted and the degree of freedom is annihilated. We will investigate the differences between point-wise predictive uncertainty estimators and entropy-based uncertainty measures for semantic segmentation in more detail in Chapter 8.

2.4.5 Neural network uncertainty calibration

For introducing the baseline concept of neural network uncertainty calibration for semantic segmentation, let us consider a multi-class setting with K classes. We will assume that a neural network model has been trained to predict pixel-wise classes $\hat{y}$ and the conditional pmf $p(\hat{y}|\hat{\boldsymbol{x}})$ is estimated by an arbitrary uncertainty quantification method as elaborated in Section 2.4.4. The predicted confidence for the most likely class $\max\, p(\hat{y}|\hat{\boldsymbol{x}}) \in [0, 1]$ is expected to match the model accuracy c if the neural network is calibrated [70]. Mathematically, this calibration condition on the model confidence can be expressed as:

$$\mathbb{P}\left(\hat{y} = \mathrm{E}[y|\hat{\boldsymbol{x}}] \mid c = \max\, p(\hat{y}|\hat{\boldsymbol{x}})\right) = c \; , \quad \forall c \in [0, 1] \; . \tag{2.97}$$

This implies that if the model prediction is made with a confidence of $\max\, p(\hat{y}|\hat{\boldsymbol{x}}) = 80\%$ for 100 repeated experiments, then the average accuracy should also amount to $c = 80\%$. Hence, the observed model accuracy serves as the expected model confidence, which should be assigned to the prediction a priori.

A complementary approach to quantify the miscalibration in uncertainty was introduced as a direct extension of Equation (2.97) by Laves et al. [110], where it was formulated that a model, which predicts an uncertainty of $\mathcal{H}\left[p(\hat{y}|\hat{\boldsymbol{x}})\right] = 20\%$ for 100 repeated experiments, would similarly lead to an average error of $q = 20\%$. This gives rise to the idea of introducing an alternative calibration condition tailored for predictive uncertainty estimates:

$$\mathbb{P}\left(\hat{y} \neq \mathrm{E}[y|\hat{\boldsymbol{x}}] \mid q = \mathcal{H}\left[p(\hat{y}|\hat{\boldsymbol{x}})\right]\right) = q \; , \quad \forall q \in [0, 1] \; . \tag{2.98}$$

Here, $\mathcal{H}$ denotes an arbitrary measure for the predictive uncertainty, but most commonly, the Shannon entropy (2.94) is employed.

Unfortunately, the predictive uncertainty measures presented in Section 2.4.4 are not perfectly calibrated a priori due to the implicit assumptions made in the uncertainty quantification process. As a consequence, additional measures are required that assess the quality of the uncertainty/confidence calibration given a particular uncertainty/confidence measure. Subsequently, we will introduce several calibration quality measures used within this work, which aim to quantify the violation of the calibration condition outlined in Equation (2.97) for confidence, and introduced in the context of uncertainty in Equation (2.98).

Expected Calibration Error (ECE)

The expected calibration error (ECE), introduced by Naeini et al. [139] is the weighted and binned average of the difference between the model accuracy and the softmax likelihood of the prediction. The ECE is defined as:

$$\mathrm{ECE} = \sum_{m=1}^{M} \frac{|B_m|}{n} \left|\mathrm{acc}\left(B_m\right) - \mathrm{conf}\left(B_m\right)\right| \ , \tag{2.99}$$

where the model's predictions from the softmax function σ are binned into M bins B_m of equal width in the range $[0, 1]$ and the total number of predictions is denoted by n. The absolute difference between the bin-wise average accuracy $\mathrm{acc}\left(B_m\right)$ and average confidence $\mathrm{conf}\left(B_m\right)$ weighted by the fraction of predictions present in each bin results in the ECE. This procedure is repeated on an image by image basis, as each instance should be analyzed independently. The mean ECE (mECE), obtained by averaging over all instances of the test dataset, is finally reported. Nixon et al. [137] expose drawbacks of the ECE in estimating the true calibration error, attributing the binning of skewed data into a fixed calibration range as one of the causes. Also, they empirically observe that the calibration loss surface has its respective minimum for the Negative Log-Likelihood (NLL) and the ECE at different temperatures, hinting at the decoupling between different metrics that Section 8 investigates in more detail.

Uncertainty Calibration Error (UCE)

Analogous to the definition of ECE in Equation (2.99) and its interpretation, the expected uncertainty calibration error (UCE) can be expressed as the weighted and binned average of the difference between the model error (modeled as the complement to the model accuracy) and the total uncertainty in the prediction (usually quantified by the Shannon entropy). Mathematically, the UCE is given by:

$$\mathrm{UCE} = \sum_{m=1}^{M} \frac{|B_m|}{n} \left|\mathrm{err}\left(B_m\right) - \mathrm{unc}\left(B_m\right)\right| \ . \tag{2.100}$$

Here, $\mathrm{err}(B_m) := 1/|B_m| \sum_{i \in B_m} \mathbf{1}(\hat{y}_i \neq \mathrm{E}[y_i|\hat{\boldsymbol{x}}_i])$ determines the binwise error and the binwise total uncertainty is characterized by $\mathrm{unc}(B_m) := 1/|B_m| \sum_{i \in B_m} \mathcal{H}[p(\hat{y}_i|\hat{\boldsymbol{x}}_i)]$.

Area Under the Reliability Error Curve (AUREC)

The area enclosed between the reliability curve and the expected diagonal in case of perfect calibration is defined as the Area Under the Reliability Error Curve (AUREC). As the area is discretized by the binning operation, the AUREC equals the ECE if the weighting factor w.r.t. the bin cardinality is replaced by the bin width.

Uncertainty Calibration Score (UCS)

The idea of assessing the areal deviation was originally introduced by Wursthorn et al. [196] for estimating uncertainty calibration in a pose estimation problem by the Uncertainty Calibration Score (UCS). As pose estimation is a regression problem, the UCS is adopted to the reliability diagram in a classification problem by utilizing the AUREC. Hence, the UCS is redefined as:

$$\mathrm{UCS} = 1 - \frac{\mathrm{AUREC}}{A_{\max}} \,. \tag{2.101}$$

The worst possible case of calibration corresponds to the maximum enclosed area of $A_{\max} = 0.5$ for classification tasks. As the calibration metrics ECE and UCE are terminologically very similar to UCS, CCQS (Confidence Calibration Quality Score) is used for the confidence-based reliability diagrams and UCQS (Uncertainty Calibration Quality Score) substitutes UCS for uncertainty-based calibration curves.

Area Under the Sparsification Error curve (AUSE)

Sparsification refers to the sequential removal of elements from the validation set, sorted in descending order based on a predictive uncertainty measure. A sparsification curve is computed as a function of sparsification. In the context of multi-class classification, if such an uncertainty measure that attempts to estimate the uncertainty works optimally as intended, then the wrongly classified instances (the sum of false positives and false negatives) are targeted first by the measure. As soon as the sparsification process has removed all wrongly classified instances, the target key performance indicator should be maximized. A sparsification curve obtained from such a flawless uncertainty measure is the *oracle* curve. It is the theoretical upper bound of all sparsification curves.

The sparsification error curve is obtained as the difference between the oracle curve and the sparsification curve utilizing a predictive uncertainty measure. It is common to use the Shannon entropy from Equation (2.94) as an uncertainty measure to sort and subsequently sparsify the predictions. Finally, a proper scoring rule like the Brier score [59] can be used

to evaluate the remaining predictions in the validation set [71]. An alternative approach is to evaluate the sparsified predictions directly with the IoU and to interpret the complement to the softmax likelihood as the predictive uncertainty measure in accordance with the Variation Ratio (2.96). Dreissig et al. [42] employ this approach since it eliminates the conundrum that arises out of the difference between the nature of uncertainty quantified by Shannon entropy and the variation ratio - while the entropy considers the entire probability vector, the variation ratio considers only the uncertainty associated with the top-1 class predicted by the model [43].

Finally, the integral of the sparsification error curve, defined by the difference between the oracle and the sparsification curve, results in a scalar calibration measure denoted as the Area Under the Sparsification Error curve (AUSE).

Temperature Scaling

The calibration quality measures introduced so far give us the possibility to evaluate the trustworthiness of an arbitrary uncertainty measure. As a next step, we want to introduce calibration methods – in particular temperature scaling – in order to correct the bias, which some uncertainty measures might entail. Post-hoc methods to calibrate a trained model's predictions are broadly classified into non-parametric and parametric calibration methods. Generally, non-parametric methods do not preserve the model accuracy, which is undesirable. Parametric approaches like Platt scaling [142] and its multi-class extension temperature scaling [70] are popular yet straight-forward approaches that employ a global scalar parameter referred to as the temperature T to rescale the logits vector produced by the penultimate layer Ξ of the model. The optimal temperature is determined by minimizing the negative log-likelihood. An advantage of temperature-scaling-based methods is that logit-rescaling does not alter the ranking of the prediction vector $p(\hat{y}|\hat{\boldsymbol{x}})$, thereby preserving the model accuracy. This invariance of the temperature scaling methodology regarding the predicted class manifests itself mathematically as:

$$\underset{\mathrm{K}}{\operatorname{argmax}}\ \sigma\left(\Xi\left(\boldsymbol{w}, \hat{\boldsymbol{x}}\right)\right) \overset{\text{acc}}{=} \underset{\mathrm{K}}{\operatorname{argmax}}\ \sigma\left(\frac{\Xi\left(\boldsymbol{w}, \hat{\boldsymbol{x}}\right)}{T}\right) . \tag{2.102}$$

Although recent empirical studies have shown that temperature scaling can be surpassed in performance by certain training-time approaches like label-smoothing [134], these are computationally intensive and also unfeasible to apply to black-box models [130]. This supports the continued use of temperature scaling and its successor (in terms of improved expressiveness) parameterized temperature scaling in ongoing research [167].

The phrasing of the parameter T as a temperature is not by accident. In statistical physics, the reciprocal temperature describes the derivative of the entropy $\mathcal{H}$ w.r.t. the internal energy E [62, 131]:

$$\frac{1}{T} := \frac{\partial \mathcal{H}}{\partial E} \, . \tag{2.103}$$

This implies that atoms move slowly and occupy minimal-energy states for low temperatures. As temperature rises, atoms gain energy, which makes higher-energy states accessible and the state variability increases. This concept can be transferred to the AI world by substituting the thermodynamical entropy with the Shannon entropy [152] from information theory. As an example, in AI language models, temperature controls the variability in word predictions [197]. The model predicts a likelihood for each word within the vocabulary and the subsequent word is chosen randomly according to the predicted pmf. Calibrating the model with a temperature of zero leads to deterministic behavior and complete repeatability. Increasing the temperature allows words with lower likelihoods to still have a chance, creating more diverse results. This process mirrors the Boltzmann distribution in physics, where states are sampled based on energy levels. In AI, the Softmax function performs a similar task, treating model logits as negative energies. Higher temperatures broaden the probability distribution, making all outcomes more equal, while lower temperatures sharpen the probability distribution.

2.4.6 Explainable AI

Deep convolutional neural networks are inherently highly non-linear, wherefore it is generally difficult to assess the global sensitivity of the model predictions on single input features. One way to tackle this problem is by considering the outcome of the model with and without a particular feature. If all input feature subsets $\mathcal{S}$ are considered regarding the marginal contribution of feature i to the sub-coalition performance then the correlations between different features are inherently incorporated. Averaging the weighted marginal contribution of feature i over all possible input feature coalitions of different cardinality results in a sensitivity metric, which fulfills all fairness properties in game theory namely the efficiency-, symmetry-, linearity- and the null player condition [150]. This sensitivity metric was initially introduced by Shapley [153] in the field of economics and has been widely adopted in the explainable Artificial Intelligence (AI) world since an approximative evaluation method was found by Lundberg & Lee [116]. In general, the Shapley value φ for feature i and objective function $\hat{\Xi}$ is determined under a particular input feature set $\mathcal{M}_f$. Hence, the Shapley value is a local explanation method [127], which describes the feature effect by quantifying the direction and magnitude of the local gradient in the feature space. As a consequence, if the entire feature space is sampled equidistantly, then the Shapley values will generate a distribution for each input feature i. The shape of the

Shapley distribution for feature i in contrast to feature j might indicate differences in the global feature importance for the neural network inference. The Shapley values:

$$\varphi_i(\hat{\Xi}) := \sum_{\mathcal{S}\subseteq\mathcal{M}_f\setminus\{i\}} \binom{|\mathcal{M}_f|-1}{|\mathcal{S}|}^{-1} \frac{\left[\hat{\Xi}\left(\mathcal{S}\cup\{i\}\right) - \hat{\Xi}\left(\mathcal{S}\right)\right]}{|\mathcal{M}_f|}\,, \tag{2.104}$$

are determined by weighting the individual coalition merit with the inverse of the binomial coefficient, which quantifies the number of sub-coalitions with cardinality $|\mathcal{S}|$.

2.5 Chatterjee's rank correlation measure

Within this thesis we want to study how optical quality measures of the windshield correlate to the performance drop of convolutional neural networks, which they experience due to the implicit data set shift generated by optical aberrations. The non-linear correlation between those KPIs can be quantified by the Chatterjee's rank correlation measure [26, 155]. We will shortly address the theoretical foundations of the Chatterjee's rank correlation measure as it will be required in order to select the most suitable optical metric.

In order to quantify the correlation between two input signals, suppose x and y, the Pearson correlation coefficient ρ is typically employed. Unfortunately, the Pearson correlation coefficient is restricted to linear relationships. In order to evaluate the non-linear correlation between the two signals – in our case 'optical quality' and 'AI performance' – alternative metrics are required. A very fundamental definition for non-linear correlation measures is given by the Dette-Siburg-Stoimenov's rank correlation metric $\langle\xi_n\rangle$ [155] defined as:

$$\begin{aligned}\langle\xi_n\rangle &:= \frac{\left\langle \mathrm{VAR}_x\left[\mathrm{E}_y\left[\mathbf{1}_{\{y\geq t\}}(y)\mid x\right]\right] \mid \mathrm{pdf}_y(t)\right\rangle}{\left\langle \mathrm{VAR}_y\left[\mathbf{1}_{\{y\geq t\}}(y)\right] \mid \mathrm{pdf}_y(t)\right\rangle} \\ \Leftrightarrow \langle\xi_n\rangle &\overset{(2.93)}{=} 1 - \frac{\left\langle \mathrm{E}_x\left[\mathrm{VAR}_y\left[\mathbf{1}_{\{y\geq t\}}(y)\mid x\right]\right] \mid \mathrm{pdf}_y(t)\right\rangle}{\left\langle \mathrm{VAR}_y\left[\mathbf{1}_{\{y\geq t\}}(y)\right] \mid \mathrm{pdf}_y(t)\right\rangle}\,.\end{aligned} \tag{2.105}$$

The quotient of the second term is given by the expected unexplained variance over the expected total variance. According to the law of variance decomposition [183], the total variance $\mathrm{VAR}[y]$ is given as the sum of the explained variance $\mathrm{VAR}[\mathrm{E}[y\mid x]]$ and the unexplained variance $\mathrm{E}[\mathrm{VAR}[y\mid x]]$. If there is a functional relationship between x and y (without any noise) then the unexplained variance vanishes. The expectation value of the unexplained variance of the indicator function $\mathbf{1}_{\{y\geq t\}}(y)$ over the distribution of y is required in order to scan through all possible ranking thresholds t according to their likelihood. Hence, the Dette-Siburg-Stoimenov's correlation coefficient is a rank correlation metric.

Equation (2.105) is hard to evaluate numerically given a discrete sample. The Chatterjee's rank correlation coefficient ξ [26] presents an approximation of $\langle \xi_n \rangle$ that converges to the expectation value as the sample size $n \mapsto \infty$. The Chatterjee's rank correlation coefficient is given by:

$$\xi_n := 1 - \frac{n}{2} \cdot \frac{\sum_{i=1}^{n-1} |r_{i+1} - r_i|}{\sum_{i=1}^{n} l_i \, (n - l_i)} \quad \text{with:} \quad \begin{cases} r_i := \sum_{j=1}^{n} \mathbf{1}_{\{y_j \leq y_i\}} \, , \\ l_i := \sum_{j=1}^{n} \mathbf{1}_{\{y_j \geq y_i\}} \, . \end{cases} \tag{2.106}$$

2.5.1 From theory to practice: Chatterjee vs. Pearson correlation

In order to demonstrate the powerfulness of the Chatterjee's rank correlation measure, a toy example case study is presented. Suppose the following test function:

$$y := \begin{cases} 2 - \cos(10x) & , \, x \in (-\infty, \, -2\pi) \cup (2\pi, \, \infty) \\ 12 + \sum_{n=1}^{10} \sin(nx) & , \, x \in [-2\pi, \, 0) \\ 12 - \sum_{n=1}^{10} \sin(nx) & , \, x \in [0, \, 2\pi] \, . \end{cases} \tag{2.107}$$

Additionally, the function is disturbed by random noise sampled from a Gaussian with zero mean and a standard deviation of $\sigma_\varepsilon = 0.3$. The corresponding graph is visualized in Figure 2.11 for $x \in [-10, \, 10]$ sampled uniformly and the corresponding Chatterjee's rank correlation measure amounts to $\xi_{1001} = 0.824$ considering $n = 1001$ samples.

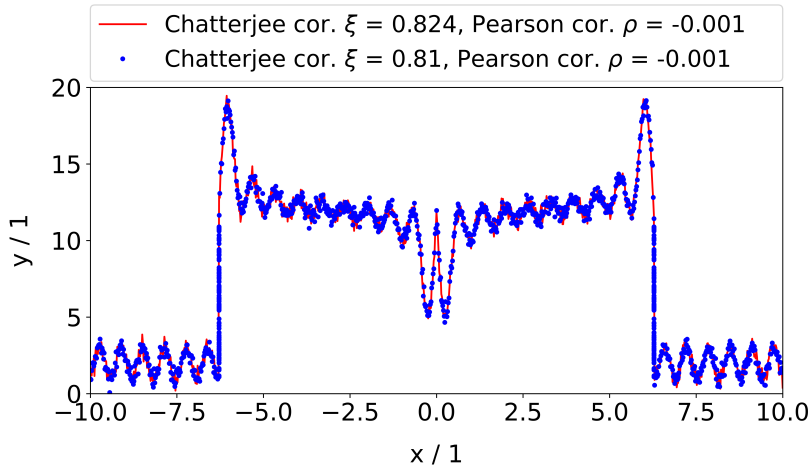


Figure 2.11: The Chatterjee's rank correlation measure ξ is compared to the Pearson correlation coefficient ρ for the test function presented in Equation (2.107). The red curve indicates the functional relationship with Gaussian noise applied to it. Furthermore, subsamples are added randomly at the discontinuity ($x = \pm 2\pi$) to demonstrate the sensitivity of ξ on the unexplained variance contribution, depicted by the blue dots.

If the Pearson correlation coefficient is evaluated in comparison, it can be noticed in Figure 2.11 that ρ almost vanishes. This indicates the insufficiency of ρ for non-linear relationships.

Since the Chatterjee's rank correlation measure quantifies the amount of unexplained variance within the sample, it is expected that ξ_n reduces if multiple samples are drawn within the discontinuity at $x = \pm 2\pi$. This is also illustrated in Figure 2.11, where $n_{\text{sub}} = 100$ subsamples were randomly added within each discontinuity increasing the unexplained variance contribution. The decay of ξ_n with increasing cardinality of the inserted subsample at each discontinuity is studied more systematically in Figure 2.12.

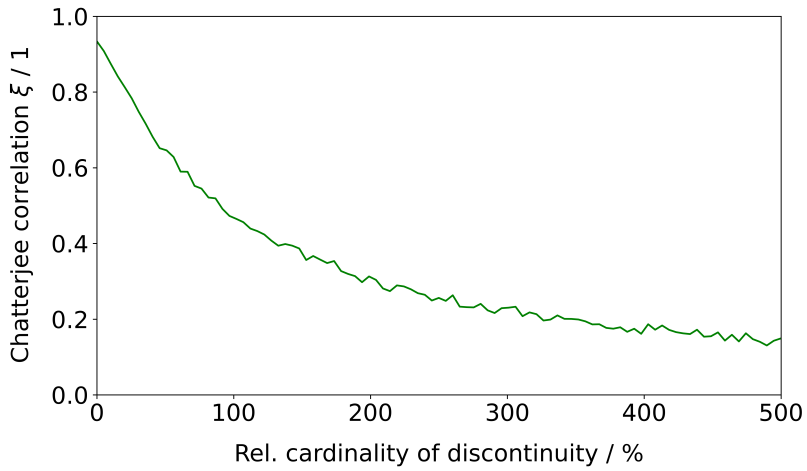


Figure 2.12: The Chatterjee's rank correlation measure ξ is shown as a function of the relative cardinality of the subsample inserted at the location of the test function discontinuity at $x = \pm 2\pi$.

3 BOS imaging for tilt correction

This chapter includes elements with editorial changes from

[192] Dominik Werner Wolf, Markus Ulrich, and Alexander Braun. "Novel developments of refractive power measurement techniques in the automotive world". In: *Metrologia* 60 (6 Sept. 2023). ISSN: 0026-1394. DOI: 10.1088/1681-7575/acf1a4. URL: https://iopscience.iop.org/article/10.1088/1681-7575/acf1a4

marked with a purple line and includes elements with editorial changes from

[191] Dominik Werner Wolf, Boris Thielbeer, Markus Ulrich, and Alexander Braun. "Wavefront aberration measurements based on the Background Oriented Schlieren method". In: *Measurement: Sensors* (2024), p. 101509. ISSN: 2665-9174. DOI: https://doi.org/10.1016/j.measen.2024.101509. URL: https://www.sciencedirect.com/science/article/pii/S2665917424004859

marked with a green line.

In this section, we want to address the impact of the tilting operator $\hat{\mathcal{T}}$ as introduced in Section 1.2. In general, the perturbation operator $\hat{\mathcal{H}}$, which characterizes the imaging process, is influenced by the ADAS camera as well as the windshield. Hence, the tilting operator $\hat{\mathcal{T}}$ is composed of the tilt component induced by lens distortions of the ADAS camera and the tilt component generated by distortions of the windshield. The camera tilt component can be corrected for by considering non-linear intrinsic parameters during the camera calibration process at the supplier site. By applying standard, non-linear correction schemes [158] a rectilinear mapping can be ensured suppressing the lens distortion to insignificance. Consequently, we are left with the challenge to characterize the tilt component induced by the windshield, which is generally too severe to be disregarded a priori [173].

The aberration pattern of windshields typically demonstrates significant spatial variations. This also generates local tilt components – if we consider the Zernike coefficient framework presented in Section 2.1.1 – which drives optical distortions that lead to a non-conformal mapping, e.g. a curvilinear mapping. In detail, conformality is given if the Jacobian matrix of a coordinate transformation can be expressed as a positive scalar (magnification) times an orthonormal matrix (rotation). In order to restore conformality, we will propose a non-parametric, grid-based rectification of the optical distortions induced by the windshield.

This procedure requires high-resolution, point-wise deflection vector maps as an input. In the following sections, we will demonstrate that this information can be provided by utilizing Background-Oriented-Schlieren (BOS) imaging [172, 144, 58]. The novel BOS measurement procedure employed is based on a cross-correlation algorithm and a speckle pattern. In order to ensure the comparability and traceability of the obtained results, we will investigate the first spatial derivative of the deflection vector maps, which corresponds to the local refractive power of the windshield. For the refractive power, well established measurement procedures exist but they lack on capturing the aberration pattern as seen by the ADAS camera. As an example, Figure 3.1 visualizes the refractive power distribution across the field of view of a commercially available ADAS camera. It is evident that there is a non-zero refractive power gradient in the vertical as well as in the horizontal direction, which demonstrates the amplification of the refractive power with the varying angle of field according to the Kerkhof model, introduced in Section 2.2.2.

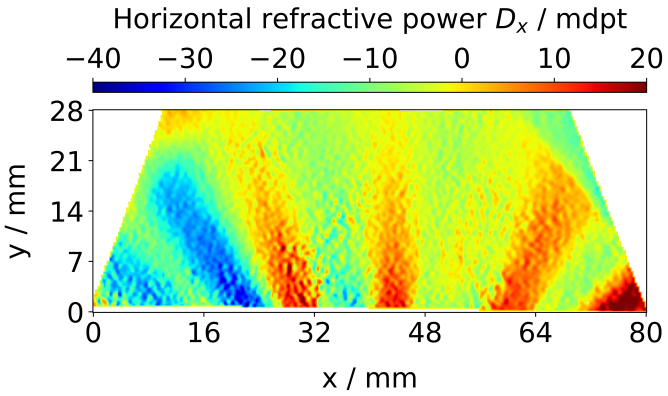


Figure 3.1: Horizontal refractive power measurement of the camera window area of a standard windshield. The fan-shaped aberration pattern, caused by the parallel float lines of the windshield, reflects the varying incident angles across the field of view of the ADAS camera.

3.1 Related work

Classical distortion correction methods [31, 19, 18, 20] strictly differentiate between rotationally symmetric and asymmetric distortions. The rotationally symmetric contribution can be further decomposed into radial and tangential distortions [115]. As the typical optical system in photogrammetry consists of a camera with a nominal rotational symmetric objective lens, the rotationally symmetric distortion is generally assumed to dominate over the asymmetric distortion [200, 115]. The rotationally symmetric contribution is modeled by distortion functions, which exclusively depend on the radial coordinate referenced to the location of the center of distortion. This center of symmetry does not need to be congruent with the optical axis, but it is highly questionable if such a point

even exists for distortions induced by the windshield (see Figure 3.1). Hence, for the particular optical setup we are concerned with, the radial symmetry is broken to such a high degree that rotationally symmetric correction schema are condemned to failure. Asymmetric distortions are addressed by different parametric models, e.g. Brown [19, 18] introduces two parameters to characterize distortions arising from the misalignment of the lenses within the camera objective lens [115]. In order to account for additional sources of asymmetric distortions (e.g. due to the lack of image-plane flatness), the complexity of the parametric model can be increased. Typically, this is done by introducing supplementary parameters that allow for a higher degree of freedom of the model [20, 174]. This implicitly allows the model to capture higher frequencies in the Fourier space of the distortion map. Nevertheless, those parametric, asymmetric distortion models are still highly limited in terms of the baseband bandwidth, which is critical for immensely spatial-varying aberration patterns as those imposed by the windshield. As a consequence, a non-parametric, grid-based rectification strategy for the windshield's distortions is recommended in order to account for the limitations of parametric distortion models in capturing local features of the distortion map.

Such a non-parametric, grid-based rectification strategy has already been employed successfully for Head-up-Displays (HUD) in the automotive industry. Here, distortions are introduced due to the reflection on the windshield inner-surface and they are actively compensated by inverting the distortions in the image on the display of the imaging unit, which is commonly known as warping [184, 186]. The required measurement information about the local deflection map is obtained by capturing an image of a reference pattern with a camera end-of-line. For the camera window area, an end-of-line measurement is not favorable as the hyperfocal distance of standard telephoto cameras in the automotive industry is in the order of 60 m. Hence, to measure the high-resolution local deflection maps of the windshield a priori, we will explore in this Chapter a novel measurement procedure based on BOS imaging [172, 144].

Standard refractive power measurement devices in the automotive industry are highly resolution limited [192], wherefore the measurement information is insufficient for compensating the effect of the windshield's tilting operator. BOS imaging possess the potential of resolving this limitation in resolution and implicitly accounts for the varying incident angles across the field of view of the ADAS camera as visualized in Figure 3.1. The contribution of this Chapter is to demonstrate that BOS imaging can be adapted for refractive power measurements, delivering results comparable to established standard refractive power measurement systems in the automotive industry, which align with the UN/ECE-R43 [148]. Furthermore, a calibration procedure, based on the Kerkhof model presented in Section 2.2.2, is experimentally tested for ensuring the traceability of the obtained local deflection maps to international metrological reference standards. These are key requirements for generating trustworthy measurement results, which is of utmost importance for safety-critical autonomous driving functionalities. To conclude,

the baseline distortion correction strategy as well as BOS imaging are widely established methods but the combination of both approaches to compensate the windshield's tilt component generates novelty.

Our proposed strategy to ensure a rectilinear mapping by utilizing high-resolution BOS imaging for a non-parametric, grid-based distortion correction has recently been explored by other fields as well [206]. In contrast to our comparability and traceability study regarding the local refractive power deduced from the distortion map, Zhan et al. [206] directly analyze the residual rectification error, which amounts to 0.02 px for their specific setup.

3.2 Refractive power measurements based on BOS imaging

The conclusions drawn from Figure 3.1 motivate the ambition to develop a new measurement scheme, where the ADAS camera is directly part of the measurement setup to retrieve unbiased deflection maps for the rectification of the windshield's tilt component. BOS imaging [172, 144, 44] serves those needs by capturing a speckle pattern with and without the windshield in the optical path, as visualized in Figure 3.2. By employing a cross-correlation algorithm, the perturbed reference pattern can be analyzed to extract high-resolution, point-wise deflection vector maps as required. The corresponding local deflection angles can then be utilized for determining local refractive power maps by recalling Equation (2.82), which states that the local refractive power is defined as the first derivative of the deflection angle.

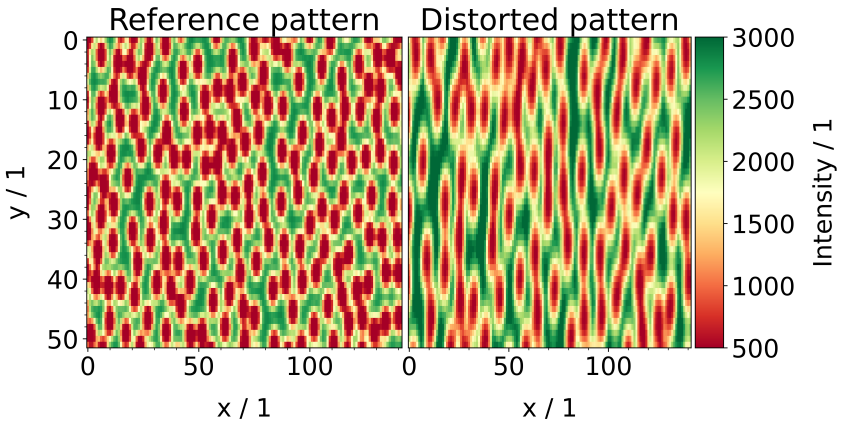


Figure 3.2: The reference pattern and the distorted pattern is visualized for the reference lens under a test angle of $\epsilon = 59.7°$. It is clearly noticeable that the circular reference pattern is distorted towards elliptical shapes due to the broken symmetry induced by the inclination angle, which amplifies the refractive power in this particular direction according to the Kerkhof model.

The local deflection angles are determined by the ratio of the corresponding deflection vector component related to the optical path length (as shown in Figure 3.3). Consequently, a local refractive power map can be calculated by numerical differentiation, which results in a uniform refractive power distribution in case of the reference lens since the slopes in Figure 3.3 are location-independent. The last step is a bit more complex than described because the optical parameters of the setup have to be considered for an accurate result. For this, the target vector is mapped by a matrix optics approach [76] onto the image plane. Consequently, this measurement setup implicitly assumes the validity of the small angle approximation, which breaks down for severe optical distortions. Nevertheless, for the scope of optical distortions in automotive applications, where D is much smaller than 1 m^{-1}, the assumption is valid. In addition, the mapping approximates the camera system as a pinhole camera and the distances between the optical elements are determined by a camera calibration procedure. Unfortunately, a more detailed description of those steps can not be given due to confidentiality obligations. The major advantage of this technique lies in the degree of freedom of choosing an arbitrary axis of interest for the refractive power calculation. Furthermore, the resolution of the system is only limited by the areal density of the reference pattern, which is subject to the magnitude of the distortion. Basically, if the distortions are severe, like in Figure 3.2, then the areal density has to be high enough, so that the cross-correlation algorithm can still identify enough reference points on the target for a given subset size. On the other hand, the areal density has to be low enough in order to guarantee for the discriminability of the reference points within the pattern.

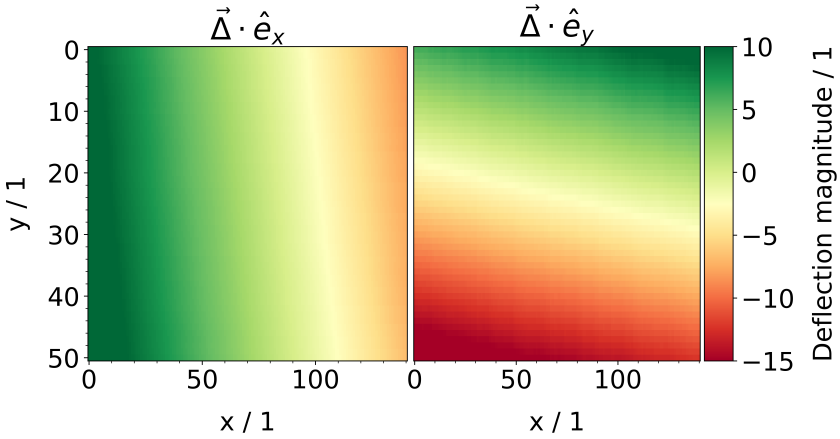


Figure 3.3: The horizontal and the vertical component of the deflection vector $\vec{\delta}$ is shown for the reference lens under a test angle of $\epsilon = 59.7°$. The homogeneity of the slopes indicates an uniform refractive power distribution. The mean refractive power value of the reference lens over the aperture surface amounts to $D_{\mathrm{meas}} = 577\ \mathrm{km}^{-1}$.

3.2.1 Cross-correlation

Cross-correlation lies at the heart of BOS imaging, wherefore we want to explicitly outline the mathematical fundamentals. In the continuous case, the cross-correlation function $\mathcal{C}[f,\ g](\delta_x,\ \delta_y)$ between the 2D, real-valued test functions f and g is given by:

$$\mathcal{C}[f,\ g](\delta_x,\ \delta_y) := \langle f(x,y) \mid g(x+\delta_x,\ y+\delta_y) \rangle \ . \tag{3.1}$$

Here, δ_x and δ_y denote the lag coordinates and x and y define the domain of the 2D test functions f and g. Consequently, the cross-correlation function represents the scalar product between the test function f and the lagged test function g. For our purpose, the test function f is given by the perturbed pattern I and the test function g is specified by the reference speckle pattern $\hat{I}$. Evaluating the scalar product in the discretized case leads to [203]:

$$\mathcal{C}[I_{\vec{r},\,\vec{\sigma}},\ \hat{I}_{\vec{r},\,\vec{\sigma}}](\delta_x,\ \delta_y \mid r_x,\ r_y,\ \sigma_x,\ \sigma_y) \approx \frac{\sum\limits_{i\leq\sigma_x}\sum\limits_{j\leq\sigma_y} I_{\vec{r},\,\vec{\sigma}}(i,\ j)\cdot \hat{I}_{\vec{r},\,\vec{\sigma}}(i+\delta_x,\ j+\delta_y)}{\sqrt{\sum\limits_{i\leq\sigma_x}\sum\limits_{j\leq\sigma_y} I^2_{\vec{r},\,\vec{\sigma}}(i,\ j)}\ \sqrt{\sum\limits_{i\leq\sigma_x}\sum\limits_{j\leq\sigma_y} \hat{I}^2_{\vec{r},\,\vec{\sigma}}(i+\delta_x,\ j+\delta_y)}} \ , \tag{3.2}$$

where the test functions are implicitly normalized to evaluate the cross-correlation independently of the mask-specific intensity level. The approximation of the cross-correlation formula introduces additional parameters, namely the position vector $\vec{r}$ and the width vector $\vec{\sigma}$ of a rectangular evaluation mask. Consequently, $I_{\vec{r},\,\vec{\sigma}}$ denotes the subset of intensity values within the perturbed pattern that fall into the mask with width $\vec{\sigma}$ at position $\vec{r}$. Hence, the summations across the pixel coordinates $(i,\ j)$ in Equation (3.2) are restricted by the mask widths σ_x and σ_y. For each position $\vec{r}$ within the image domain, the correlation function is computed and the lag coordinates $(\delta_x,\ \delta_y)$ of the maximum value are extracted. In order to account for local magnifications, the correlation function has to be marginalized across the width parameters σ_x and σ_y. Hence, for each position $\vec{r}$ a set of correlation functions is computed with different widths. From this set, the cross-correlation maximum and the corresponding lag coordinates of the maximum value are recorded element-wise. The maximum cross-correlation across the set is then taken as the result for the correlation function at position $\vec{r}$ and the corresponding lag coordinates are extracted. As a result, a 2D map of point-wise deflection vectors $\vec{\delta}$ across the image domain is obtained.

3.2.2 Comparability study

For the sake of comparability, we will subsequently investigate the refractive power dependency of a reference lens – made and calibrated by ZEISS – on the inclination angle ϵ. The Kerkhof model introduced in Section 2.2.2 provides an expectation value for the refractive power amplification factor according to Equation (2.85). In addition

to the novel BOS imaging setup, employed for measuring unbiased deflection vectors $\vec{\delta}$ in the first place, two standard measurement systems for the local refractive power are used for the comparability study. In detail, the most established refractive power measurement technique in the automotive industry is studied, which is based on Moiré interference patterns [106, 64, 82]. In a nutshell, a fine-line grid is coherently illuminated, propagated through the windshield and finally mapped onto a secondary line grid with a rotational offset. Without any optical perturbations of the windshield, the superposition of the line grid patterns creates a Moiré interference pattern with constant differentiation length l_0. Due to surface imperfections, optical distortions will perturb the transmitted line grid pattern and the superposition of the perturbed and unperturbed pattern will show local deviations from the constant differentiation length l_0. The deviation from the differentiation length Δl_0 gives rise to the local deflection angle. The refractive power map is finally obtained by differentiating the deflection angle profile along the transversal axes, similarly to the BOS imaging approach. The second refractive power measurement setup for the comparability study takes up on the proposed methodology by §9.2.1.1.3. of the UN/ECE-R43 regulation on safety glasses [148]. Very briefly, two parallel light beams (typically monochromatic laser beams) are propagated through the windshield and the transversal distance in between is monitored. By employing the matrix optics formalism [76], the transversal distance can be mapped to the refractive power. We will refer to the UN/ECE-R43 inspired measurement scheme as the laser-based method.

The deviation from the Kerkhof model is visualized in Figure 3.4 for all three measurement techniques. The local refractive power measurements obtained from BOS imaging (aka. cross-correlation method) follow the expected Kerkhof model to the first order.

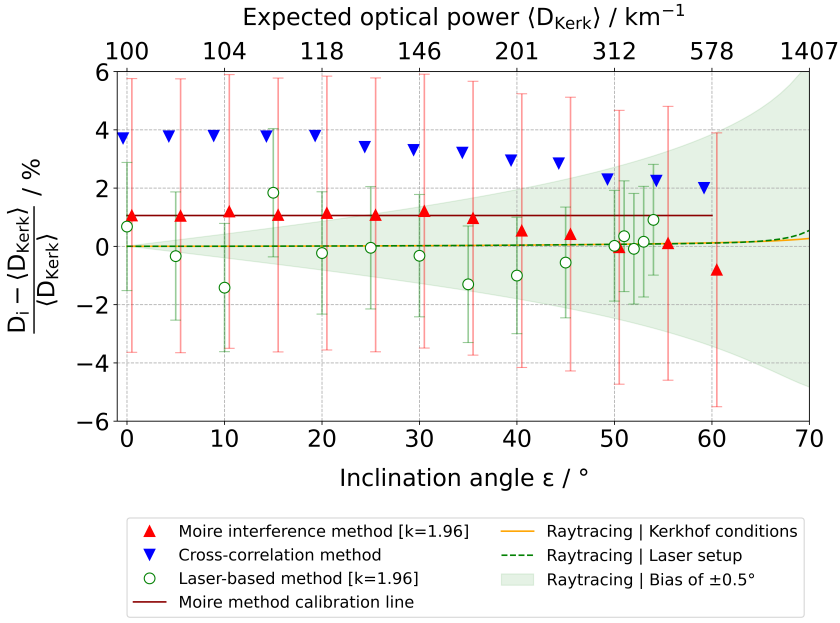


Figure 3.4: The relative refractive power deviation with respect to the Kerkhof model is plotted versus the inclination angle ϵ for different measurement systems at a 95% confidence level ($k = 1.96$).

Nevertheless, there is still a significant bias, which is subject to current investigations. Most likely the estimators for the distances of the measurement setup, which enter the refractive power calculation in the matrix optics mapping, are not sufficiently accurate. Hence, the setup alignment procedure has to be refined. In contrast, the measurement results from the laser setup are in agreement with the Kerkhof model under consideration of the measurement uncertainties and do not indicate a bias. Finally, the Moiré interference method also follows the expectation value given by the Kerkhof model. Nevertheless, the clustering appears to be slightly biased towards higher values, although this bias is not significant considering the uncertainty margins specified by the system manufacturer. If this bias was significant, then the current calibration strategy in place would need a refinement. Currently, the measurement device using a Moire interference pattern is calibrated by two reference lenses with ± 100 km^{-1}, which are traced back to international standards to ensure a valid calibration chain. The resulting calibration line is assumed to be valid over the entire measurement range. But the results in Figure 3.4 potentially contradict this assumption and indicate that the current calibration procedure should be extended to a non-linear model.

3.3 Interim conclusion

A non-parametric, grid-based approach has been proposed to correct optical distortions induced by the windshield, thereby restoring conformality. As this procedure requires high-resolution, point-wise deflection vector maps as an input, BOS imaging was studied. It has been shown that the deflection maps measured by BOS imaging implicitly account for the varying incident angle across the field of view of the camera. The comparability and traceability of the obtained results has been investigated by deducing the local refractive power from the deflection measurements via the first-order spatial derivative. The BOS imaging approach aligns with the expected behavior described by the Kerkhof model, except of a bias in the order of 4%. In summary, BOS imaging is a suitable measuring technique for capturing the windshield's optical distortions as seen by the ADAS camera. The corresponding deflection maps fulfill the requirements for the non-parametric, grid-based rectification of the windshield's tilt component. This suppresses the tilting operator $\hat{\mathcal{T}}$ to the identity mapping $\hat{\mathbb{1}}$. As a result, the perturbation operator $\hat{\mathcal{H}}$ simplifies to the blurring operator $\hat{\mathcal{B}}$, which will be the subject of the subsequent Chapters.

4 Limitations of current measures for windshield inspection

This chapter includes elements with editorial changes from

[193] Dominik Werner Wolf, Markus Ulrich, and Alexander Braun. "Windscreen Optical Quality for AI Algorithms: Refractive Power and MTF not Sufficient". In: *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)* (Sept. 2023), pp. 5190–5197. DOI: 10.1109/ITSC57777.2023.10421970. URL: https://ieeexplore.ieee.org/document/10421970/

marked with a blue line.

The methodology presented in Chapter 3 allows the reduction of the perturbation operator $\hat{\mathcal{H}}$ to the blurring operator $\hat{\mathcal{B}}$. Consequently, we are left with the challenge to safeguard the perception system against the impact of the blurring operator $\hat{\mathcal{B}}$. As a deconvolution approach generates an unacceptable system latency [101] given the limited on-board compute resources in the car, the blurring operator $\hat{\mathcal{B}}$ is supposed to be suppressed by adequate optical requirements. Given this ambition, a natural question arises: are the current optical quality requirements for the windshield sufficient for achieving this goal?

4.1 Related work

The optical requirements for the homologation of windshields is specified within the European Union by the UN/ECE-R43 regulation on safety glasses [148]. The latest version dates back to 2014 and does only account for the tilting operator $\hat{\mathcal{T}}$. As described in Section 1.1, the blurring operator $\hat{\mathcal{B}}$ gains importance due to the recent devolvement of utilizing telephoto cameras for ADAS/AD applications. As of today, the blurring effect is not part of any legal homologation requirement but there exist a few recommendations (e.g. by the IEEE P2020 working group [65] or Tier 1 suppliers), on what could be suitable measures for quantifying the effect of the blurring operator $\hat{\mathcal{B}}$ on the imaging process. For example, the MTF ratio evaluated at half-Nyquist frequency is one famous but controversial proposal in the industry.

In this chapter, we want to shine some light into this discussion and alleviate the research gap with our contributions. We will first address the research question, whether the information content of local refractive power measurements is sufficient to characterize the optical quality of windshields w.r.t. the blurring operator $\hat{\mathcal{B}}$. Subsequently, we will answer the research question: Is the MTF a valid alternative to the local refractive power for independent windshield quality assessment?

The contributions presented in this chapter are twofold. First, we will demonstrate that the refractive power evaluates the local curvature (local second order derivatives of the wavefront) but fails to capture the global aberration pattern across the aperture, which parameterizes the PSF of the blurring operator $\hat{\mathcal{B}}$. Consequently, we will conclude that the refractive power is not a suitable measure for quantifying the impact of the blurring operator $\hat{\mathcal{B}}$. This statement will be underpinned by a mathematical proof and by experimental evidence based on a Shack-Hartmann wavefront measurement.

Secondly, it will be mathematically proven that the MTF is invariant under global tilt aberrations and experimental evidence will be provided for the non-linear interaction between the windshield and the objective lens of the camera. As a consequence, we will demonstrate that the MTF ratio is invalid by definition as it inherently applies linear system theory to a non-linear system. This observation will also lead to the conclusion that the MTF can exclusively serve as a system-level measure, as the optical components couple non-linearly, prohibiting the specification of independent part-level requirements.

4.2 Refractive power

The refractive power is currently the most established KPI of the windshield w.r.t. optical distortions described by the tilting operator $\hat{\mathcal{T}}$. At this point it has to be emphasized that the UN/ECE-R43 regulation on safety glasses [148] does not impose a requirement on the refractive power. Instead a maximum local deflection angle across the windshield surface is specified according to §9.2.6. of the UN/ECE-R43 [148]. In detail, if the distance between two parallel light beams is monitored as described in Section 3.2.2, then the local deflection angle after transmitting through the refractive element has to be smaller than $2'$ [148]. This requirement is conditioned on an initial beam displacement of 4 mm according to §9.2.1.6. [148]. As the local refractive power is defined as the first spatial derivative of the local deflection angle according to Equation (2.82), the requirement on the local deflection angle can be translated into a local refractive power requirement by evaluating the difference quotient. As a consequence, a windshield in the European Union is not allowed to exceed a maximum local refractive power of 145.4 $\mathrm{km}^{-1} \mathrel{\widehat{=}} 145.4$ mdpt within the drivers main field of view. Typically, it is industrial practice to adopt this requirement for the camera window area as well, even though it is not legally binding. Furthermore, the requirements imposed by the UN/ECE-R43 apply in all possible spatial directions

but the current standard procedure in the automotive industry is to measure exclusively along the horizontal and vertical axis. This simplification will generate an invariance of the local refractive power requirement under certain aberration patterns, which are typically described in terms of the Zernike polynomials, as introduced in Section 2.1.1.

Even though the local refractive power is utilized as a measure for optical distortions, it is tempting to investigate the relationship to Fourier optical properties specifying the blurring operator $\hat{\mathcal{B}}$. In order to do so, a very important distinction has to be made. The underlying theory of the blurring operator $\hat{\mathcal{B}}$ is described by Fourier optics and is based on the wave description of light. On the contrary, the tilting operator $\hat{\mathcal{T}}$ originates from ray optics. A direct consequence of this fundamental difference in the nature of the perturbation operators is that the blurring operator $\hat{\mathcal{B}}$ includes information about the entire aperture of the optical system and describes the global effect, whereas the tilting operator $\hat{\mathcal{T}}$ quantifies local properties of the optical setup by characterizing directional deviations of single rays.

The geometric interpretation of the PSF paves the way for connecting Fourier optical properties with ray optical measures like the refractive power. In Section 2.1 the PSF was introduced as the impulse response function or the Green's function of an optical system. In addition to the Fourier optical approach, where the PSF is obtained as the Fourier-transform of the aperture function, there is also a ray optics equivalent, which is referred to as the geometrical PSF. In contrast to the Fourier optical approach, the geometrical PSF characterizes the blurring ellipse that arises in the observation plane from local wavefront aberrations in the aperture surface. The area of the blurring ellipse that encloses the energy fraction $\mathcal{C}$ around the focusing spot – in relation to the total amount of energy entering the system through the aperture at position $\vec{x}_a$ – is proportional to the Gaussian curvature [23] of the wavefront aberration map W at this location [163, 164, 119]. Equivalently speaking, the determinant of the Hessian matrix of the wavefront aberration map W, at position $\vec{x}_a$, determines the area of the geometrical PSF:

$$\iint\limits_{\iint \mathrm{PSF}(\vec{x}_o)\Big|_{\hat{z}_o}\,\mathrm{d}^2x_o \,\stackrel{!}{=}\, \mathcal{C}} 1\;\mathrm{d}^2x_o \;\propto\; \left|\begin{pmatrix} \dfrac{\partial^2}{\partial x_1^2}W(\vec{x}_a) & \dfrac{\partial}{\partial x_1}\dfrac{\partial}{\partial x_2}W(\vec{x}_a) \\ \dfrac{\partial}{\partial x_1}\dfrac{\partial}{\partial x_2}W(\vec{x}_a) & \dfrac{\partial^2}{\partial x_2^2}W(\vec{x}_a) \end{pmatrix}\right| . \tag{4.1}$$

The local curvature of the optical path difference map W physically characterizes the rate of change of the local phase gradients, which induce local deflection angles that equal those arising from optical distortions if and only if the reference wavefront has been characterized by a plane wave. Hence, the local curvature of W, with respect to the axis of interest x_i, equals the local refractive power D_{x_i} [163]:

$$D_{x_i}(\vec{x}_a) = \frac{\partial^2}{\partial x_i^2}W(\vec{x}_a)\ . \tag{4.2}$$

By adopting this relationship in the Hessian matrix, the connection between the geometrical PSF and the refractive power becomes evident. In this nomenclature, the Hessian matrix is also referred to as the dioptric power matrix $\boldsymbol{\mathcal{D}}$ [74]. By utilizing basic principles of linear algebra, the determinant of the dioptric power matrix $\boldsymbol{\mathcal{D}}$ can be rewritten in terms of the traces:

$$\iint_{\iint \mathrm{PSF}(\vec{x}_o)\Big|_{\hat{z}_o} \mathrm{d}^2 x_o \stackrel{!}{=} \mathcal{C}} 1 \, \mathrm{d}^2 x_o \overset{(4.1)}{\propto} \frac{1}{2}\left[\left(\mathrm{tr}\left(\boldsymbol{\mathcal{D}}\right)\right)^2 - \mathrm{tr}\left(\boldsymbol{\mathcal{D}}^2\right)\right] . \tag{4.3}$$

Here, the second term depends on the mixed partial derivatives of W, which characterize the rate of change of the local phase gradients in the x-direction as one moves into the y-direction and vice versa. As the local refractive power is defined as the second spatial derivative of W, measurements of D_{x_i} do not capture information about the local phase gradients. Consequently, only the diagonal elements of $\boldsymbol{\mathcal{D}}$ are measured and off-diagonal elements of the Hessian matrix remain untracked. As changes in the trace of $\boldsymbol{\mathcal{D}}^2$ are not observable with the measurement information available, the area of the blurring ellipse – and the imaging quality respectively – can be different for two setups that show the same measurement results in terms of D_{x_i}. This demonstrates that there is a blind spot in the quality assurance chain at the moment.

This conclusion can be further underpinned by a mathematical argument. The trace of $\boldsymbol{\mathcal{D}}$ is given by:

$$\mathrm{tr}\left(\boldsymbol{\mathcal{D}}\right) = \sum_{i=1}^{d} D_{x_i}(\vec{x}_a) = \triangle W(\vec{x}_a) . \tag{4.4}$$

If we solely observe the refractive power field and assume for a moment that we do not or cannot measure the wavefront aberration field induced by the windshield, then W can be treated as a pseudo potential with an implicit gauge freedom. Of course, this is just a thought experiment because fundamentally a gauge freedom in field theory is characterized by not changing the physics, which is the case if we manipulate the wavefront aberration field W. Hence, this argument only holds if we assume that W is not physically observable or we consider the fact that W is simply unknown since current specifications for windshields only specify requirements in terms of the refractive power D. The gauge freedom is manifested by the gauge field $\Gamma(\vec{x}_a)$, which indicates that the trace of $\boldsymbol{\mathcal{D}}$ is unaffected by wavefront aberration fields fulfilling the Laplace equation:

$$\triangle \Gamma(\vec{x}_a) \stackrel{!}{=} 0 . \tag{4.5}$$

As a result, the trace of $\boldsymbol{\mathcal{D}}$ is gauge invariant under aberration fields $\Gamma(\vec{x}_a)$ that are composed of harmonic functions (2.73). Any mode n of the superposition in Equation (2.73) can be directly correlated to the corresponding Zernike polynomial in Table 2.1 and associated with it's physical aberration effect. Hence, the lack of information about the off-diagonal

elements of the dioptric power matrix $\mathcal{D}$ [74] leads to a blind spot in the quality assurance chain, which is characterized by local wavefront aberrations in the form of harmonic functions.

In a nutshell, imposing requirements on the local refractive power is insufficient for extracting more fundamental information about the optical system in terms of the geometrical PSF. As the incoherent Fourier optical PSF relates to the marginalized geometrical PSF, obtained by integrating over the aperture function, it becomes evident that the local refractive power is not an adequate measure for quantifying the effect of the blurring operator $\hat{\mathcal{B}}$ on the imaging process.

Furthermore, the simplification made in the tracking of the UN/ECE-R43 requirements [148] – by exclusively measuring the local refractive power along the horizontal and vertical direction – generates a local invariance under oblique astigmatism (Z_3), which induces a curvature modulation of W along the diagonal while maintaining zero curvature along the canonical axes.

From previous studies on the effect of oblique astigmatism (Z_3) on road sign classification [133, 132] it becomes evident that refractive power measurements are insufficient as a stand-alone quality requirement for windshields in order to ensure reliable computer vision for autonomous driving vehicles. In summary, the image quality can be deteriorated even though the refractive power measurement indicates a compliant windshield sample.

Our conclusions are based on the validity of the relationship between the local curvature of the wavefront aberration map and the local refractive power as presented in Equation (4.2). This relationship is not well established in the automotive industry, wherefore it is desirable to proof it's validity. This will be done in the subsequent sections by studying the wavefront modulation induced by a convex lens with a certified refractive power. First, the validity of Equation (4.2) will be mathematically proven for the special case of a thin spherical lens and secondly a Shack-Hartmann wavefront measurement of a calibrated reference lens will provide experimental evidence.

4.2.1 Theoretical verification

If it is assumed that the aperture surface aligns with the principal plane of the refractive element and a spherical lens is considered, then the wavefront modulation W induced by the lens – with respect to a plane reference wave – directly relates to the focal length f of the lens:

$$f_{x_1}^2 = x_1^2 + (f_{x_1} - W(x_1))^2 \quad , \quad \text{w.l.o.g.:} \quad x_2 \overset{!}{=} 0 \, . \tag{4.6}$$

Here, the problem was simplified to a 1D aperture without loss of generality. Equation (4.6) can be rearranged in order to isolate the wavefront aberration map W:

$$W(x_1) \overset{(4.6)}{=} f_{x_1}\left(1-\sqrt{1-\left(\frac{x_1}{f_{x_1}}\right)^2}\right) = f_{x_1}\left(1-\left(1-\frac{1}{2}\left(\frac{x_1}{f_{x_1}}\right)^2+\mathcal{O}\left\{\left(\frac{x_1}{f_{x_1}}\right)^4\right\}\right)\right) . \quad (4.7)$$

The aperture surface directly limits the domain of the transversal axis x_1. As the aperture stop radius of the optical setup is typically much smaller than the focal length associated with local wavefront curvature modulations imposed by windshield aberrations, it is valid to discard higher order terms $\mathcal{O}$:

$$W(x_1) \overset{(4.7)}{\approx} \frac{x_1^2}{2f_{x_1}} \overset{(2.82)}{=} \frac{D_{x_1}}{2}\cdot x_1^2 \quad \Rightarrow \quad D_{x_1} = \frac{\partial^2}{\partial x_1^2}W(x_1) . \quad \blacksquare \quad (4.8)$$

As a result, under the assumption of a weak spherical thin lens the validity of Equation (4.2) has been proven.

4.2.2 Experimental verification

The mathematical proof presented in Subsection 4.2.1 can be underpinned with experimental evidence. As a test setup, a weak plano-convex lens with a refractive power of $\langle D\rangle = (100.3 \pm 2)$ mdpt is tested by a Shack-Hartmann wavefront sensor. The plano-convex lens under test is traced back to national standards by an accredited calibration authority and was produced by Zeiss. A Shack-Hartmann wavefront sensor captures the wavefront modulation induced by a refractive optical element under collimated light. In detail, the Shack-Hartmann sensor consists of a microlens array, which resolves the local wavefront perturbations by focusing a wavefront snippet on a Charge-Coupled Device (CCD) or CMOS sensor. Without any aberrations, the wavefront sensor will capture the light in each subaperture center of the microlens array. If aberrations are present, then the focusing spot will be displaced locally by d_x and d_y, respectively. The resulting local gradient β_i of the optical path difference W is given by:

$$\begin{aligned} \vec{\beta} &:= \left[\, \beta_x\big|_{\vec{x}_1} \;\cdots\; \beta_x\big|_{\vec{x}_m} ,\; \beta_y\big|_{\vec{x}_1} \;\cdots\; \beta_y\big|_{\vec{x}_m} \,\right]^{\mathrm{T}} \\ \Leftrightarrow \vec{\beta} &= \left[\, \frac{d_{x_1}}{\sqrt{f_{\mathrm{sh}}^2+d_{x_1}^2}} \;\cdots\; \frac{d_{x_m}}{\sqrt{f_{\mathrm{sh}}^2+d_{x_m}^2}} ,\; \frac{d_{y_1}}{\sqrt{f_{\mathrm{sh}}^2+d_{y_1}^2}} \;\cdots\; \frac{d_{y_m}}{\sqrt{f_{\mathrm{sh}}^2+d_{y_m}^2}} \,\right]^{\mathrm{T}} . \end{aligned} \quad (4.9)$$

Here, f_{sh} denotes the focal length of the microlenses and m specifies the number of microlenses within the array. As the Shack-Hartmanm sensor captures the local gradients β_i of the wavefront aberration map W, the local refractive power can be numerically determined by a standard central difference scheme in accordance with Equation (4.2). If the measurement plane – determined by the position of the Shack-Hartmanm sensor

– does not coincide with the principal plane of the refractive element, then the local refractive power has to be rectified by considering a calibration factor ξ that accounts for the longitudinal displacement z_m between the measurement plane, with coordinate vector $\vec{x}_m$, and the principal plane, with coordinate vector $\vec{x}_a$:

$$D_{x_i}(\vec{x}_a) \overset{(4.2)}{=} \frac{\partial^2}{\partial x_i^2} W(\vec{x}_a) = \xi(z_m) \cdot \frac{\partial^2}{\partial x_i^2} W(\vec{x}_m) \overset{(4.9)}{=} \xi(z_m) \cdot \frac{\partial}{\partial x_i} \beta(\vec{x}_m) \,. \tag{4.10}$$

The calibration factor ξ can be derived from geometrical considerations. The starting point is given by noticing that the spherical wave, which exits the spherical lens, propagates through space by converging towards the real focal point with focal length f_r. As the Shack-Hartmanm sensor is longitudinally displaced by the distance z_m from the lens, the senor measures a converging wave with reduced focal length f_m. The sum of the measured focal length f_m and the longitudinal displacement z_m lead to the unbiased focal length f_r of the lens under test:

$$f_r = z_m + f_m \overset{(2.82)}{\Leftrightarrow} \frac{1}{D_r} = z_m + \frac{1}{D_m} \Leftrightarrow D_r = \frac{1}{z_m D_m + 1} \cdot D_m := \xi(z_m) \cdot D_m \,. \tag{4.11}$$

The calibration factor ξ can be approximated by replacing D_m with it's expectation value $\langle D_m \rangle$:

$$\xi(z_m) \overset{(4.11)}{\approx} \frac{1}{z_m \langle D_m \rangle + 1} = 1 - z_m \langle D_r \rangle \quad , \quad \text{with:} \quad \langle f_m \rangle = \langle f_r \rangle - z_m \,. \tag{4.12}$$

Here, $\langle D_r \rangle$ denotes the expectation value of the refractive power of the calibrated reference lens under test. If the calibration factor ξ is accounted for in the Shack-Hartmanm wavefront measurement, then the results presented in Figure 4.1 are obtained for the local refractive power over the lens aperture.

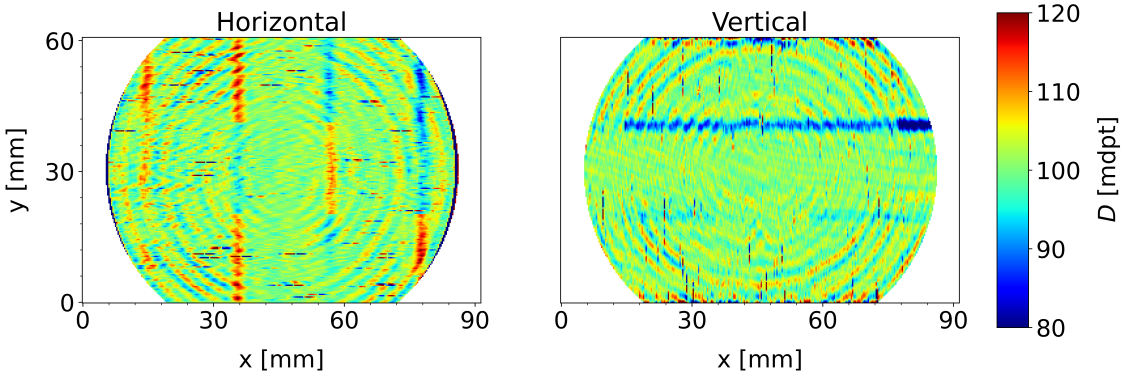


Figure 4.1: Wavefront measurement performed on a $\langle D \rangle = (100.3 \pm 2)$ mdpt reference lens. In order to cover the entire aperture of the lens, several Shack-Hartmann measurements have been stitched together. This procedure has introduced artifacts, which are visible in the measurement data by strongly pronounced vertical and horizontal lines. In total, 15 measurements have been performed over the calibration lens aperture of $d = 10$ cm.

The frequency distribution of the local refractive power across the entire aperture surface gives intel about the global refractive power of the lens as illustrated in Figure 4.2.

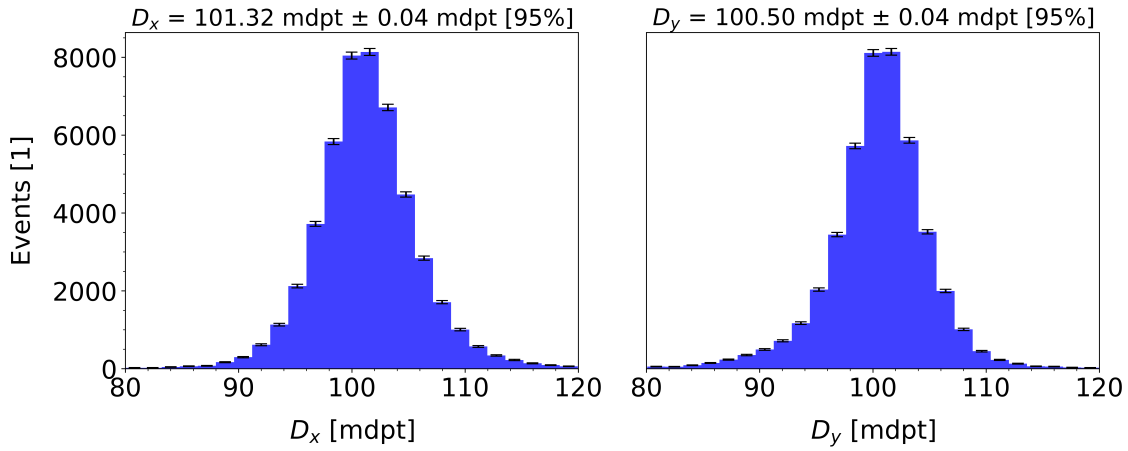


Figure 4.2: Frequency distribution of the local refractive power over the entire calibration lens aperture, deduced from a Shack-Hartmann wavefront measurement. The domain is limited to the $\pm 3\sigma$ interval, which deliberately discards outliers arising from the stitching process of several single-shot Shack-Hartmann measurements. This became necessary as the spanned sensitive area of the lenslet array was much smaller than the calibration lens aperture.

The expectation values for the global refractive power of the lens in the x- and y-plane amount to $D_x = (101.32 \pm 0.04)$ mdpt and $D_y = (100.50 \pm 0.04)$ mdpt on a confidence level of 95%. This is in agreement with the certified global refractive power of the calibration lens if the uncertainty intervals are considered. As a result, the validity of Equation (4.2) for weak spherical lenses has also been experimentally confirmed.

4.3 MTF

The MTF describes the real part of the Fourier-transform of the optical impulse response function (aka. PSF), as defined by Equation (2.53). The MTF and it's non-harmonic counterpart, the Spatial Frequency Response (SFR), are currently discussed for defining quality requirements of the windshield for telephoto-based perception systems in autonomous driving [16, 65]. In the next two subsections it will be shown that the MTF is invariant under global wavefront tilts and that the MTF ratio is an inadequate measure for defining quality requirements on the windshield.

4.3.1 The effect of tilt on the MTF

The global wavefront aberration pattern of the windshield can be decomposed into Zernike polynomials, as presented in Equation (2.72). If the Zernike polynomials are

considered in Cartesian coordinates (see Table 2.1), then it becomes obvious that the autocorrelation function in Equation (2.61) evaluates to a constant phasor (in terms of the integration variables ξ_x and ξ_y) times the autocorrelation function of the unperturbed aperture function P in case of pure y-tilt (ω_1) or x-tilt (ω_2) aberrations:

$$\widetilde{\mathrm{MTF}}(k_{x_o},\, k_{y_o}) \overset{(2.61)}{=} \left| \mathrm{e}^{-\frac{4\pi\mathrm{i}}{\lambda}\left(\omega_1 k_{y_o} + \omega_2 k_{x_o}\right)} \cdot \frac{P(k_{x_o},\, k_{y_o}) \star P(k_{x_o},\, k_{y_o})}{\iint\limits_{-\infty}^{\infty} \left|P(\xi_x,\, \xi_y)\right|^2 \,\mathrm{d}\xi_x \mathrm{d}\xi_y} \right| . \tag{4.13}$$

As the complex-valued factor has unit amplitude, the aberrated $\widetilde{\mathrm{MTF}}$ equals the diffraction-limited MTF if and only if the windshield's aberrations are entirely characterized by piston (ω_0) and tilt (ω_1 and ω_2):

$$\widetilde{\mathrm{MTF}}(k_{x_o},\, k_{y_o}) \overset{(4.13)}{=} \left| \mathrm{e}^{-\frac{4\pi\mathrm{i}}{\lambda}\left(\omega_1 k_{y_o} + \omega_2 k_{x_o}\right)} \right| \cdot \mathrm{MTF}(k_{x_o},\, k_{y_o}) = \mathrm{MTF}(k_{x_o},\, k_{y_o}) \,. \tag{4.14}$$

As a consequence, the diffraction-limited MTF is invariant under aberrations described by the Zernike coefficients of zeroth and first radial order. For that reason, piston and tilt do not represent optical aberrations in the classical sense. As the wavefront curvature remains unaffected, the blurring operator reduces to the convolution operator with the diffraction-limited PSF as the kernel function. Therefore, the second order Zernike coefficients are those of main interest for studying the effect of optical aberrations in terms of sharpness degradation on convolutional neural networks for autonomous driving.

4.3.2 Limitations in part-level measurements

The most common verification strategy in the automotive industry is given by the Vee-model [175]. Hence, there is always the objective to split system requirements (e.g., in terms of the mIoU for semantic segmentation) into part-level requirements (e.g., optical requirements on the objective lens, the windshield etc.). This V-shaped process implies that the requirements on the joint optical system has to be decomposable into two independent part-level requirements. The current praxis in the automotive industry is to define a MTF ratio [83], which quantifies the system MTF with the windshield in relation to the system MTF without the windshield.

In order to demonstrate that the MTF ratio is not sufficient for ensuring the fulfillment of system-level requirements, the interplay of the local refractive power of the windshield with the focal length of the camera system has to be considered. In general, a spherical lens maps the scene onto a curved surface, referred to as the Petzval surface [76]. As a consequence, the flat CMOS chip can never map a sharp image across the entire field of view, which gives rise to field curvature aberrations. These aberrations are

minimized by tuning the objective lens layout, but as the optimization rather results in a compromise depending on the target KPI than in a perfect compensation, there remains residual field curvature aberrations. This is a systematic aberration, which is not caused by manufacturing imprecisions as it is the case for optical aberrations induced by the windshield. Hence, the field curvature aberration is a property of the lens, characterized by the longitudinal offset Δz_{fc} over field, typically on the micrometer range. A symbolic field curvature is visualized in Figure 4.3.

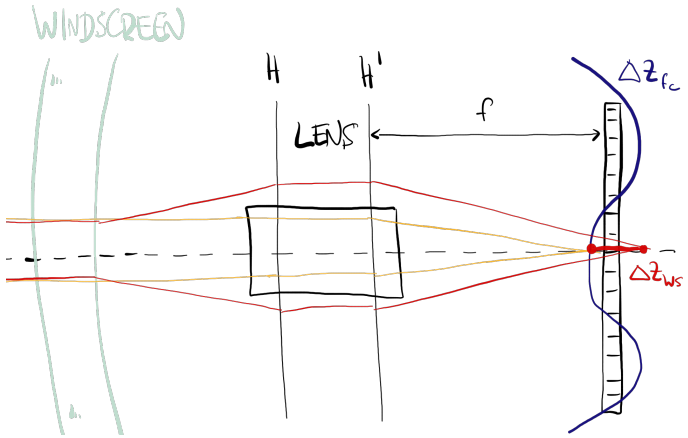


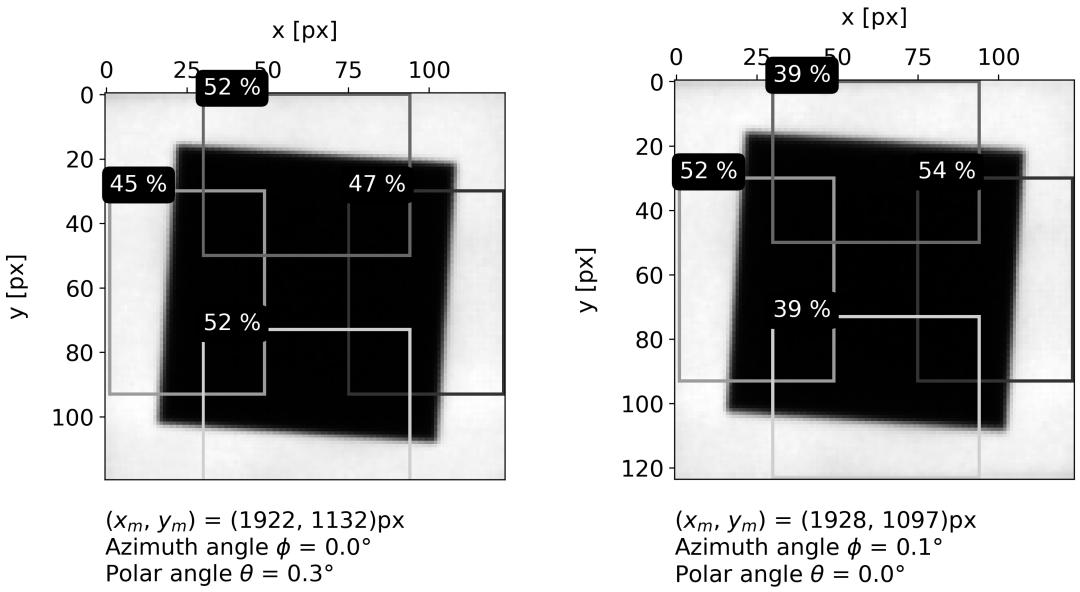
Figure 4.3: Windshield and lens form a joint optical system. H and H' are the principle planes of the lens, f is the nominal focal length. The blue line visualizes the field curvature (not to scale). Normally, parallel rays are focused onto the field curvature (yellow line). Windshield refractive power shortens or prolongs the effective focal length of the lens (red line). There are two different focus offsets Δz_{fc} and Δz_{ws} which may add or even cancel at different fields of view.

The local curvature modulation of the wavefront by the local refractive power of the windshield will also induce a longitudinal focus offset Δz_{ws} across the field of view. The two longitudinal offsets of the camera and the windshield add up to a system offset Δz_{sys}, as depicted in Figure 4.3, such that:

$$\Delta z_{\text{sys}} = \Delta z_{\text{ws}} + \Delta z_{\text{fc}} \, . \tag{4.15}$$

Importantly, both Δz_{ws} and Δz_{fc} can have positive or negative values, and thus the system offset Δz_{sys} may vanish when these terms cancel. A vanishing offset w.r.t. Δz_{sys} implies a sharpening of the system. Here, an MTF measurement of the camera alone would yield a certain number, while putting a windshield in front of the camera would act like glasses and the image would become sharper.

This hypothesis is experimentally validated by a slanted edge measurement of a standard windshield. Figure 4.4 depicts two slanted edge measurements, one without a windshield (4.4a), and one with a windshield placed in front of the camera system (4.4b).



(a) ADAS camera. **(b)** ADAS camera + windshield.

Figure 4.4: MTF measurement for an ADAS system based on the slanted edge method according to ISO 12233 [86].

The insets indicate the MTF values at half-Nyquist frequency derived from the numerical evaluation for all four edges, using an ISO12233-compliant algorithm [86]. There are two horizontal and two vertical values. The two vertical values (top and bottom) distinctly decrease from $52 \pm 1.5\,\%$ [95 %] to $39 \pm 1.7\,\%$ [95 %] when a windshield is placed in front of the camera. However, for the horizontal direction (left and right) the MTF values both significantly increase from $45/47 \pm 1.5\,\%$ [95 %] to $52/54 \pm 1.6\,\%$ [95 %] when the windshield is placed in front of the camera. The results experimentally confirm that the defocus Δz_{ws} and Δz_{fc} may cancel to a certain degree, increasing the sharpness like glasses would do for a myopic person. This conclusion is well established in physics [76] for decades but the implications for the quality assurance testing procedure of ADAS systems in the automotive industry are not well prevalent.

It has to be emphasized that the decrease in the vertical MTF is plausible, even though the horizontal MTF increases. The aberration pattern of windshields is mostly dominated by vertical astigmatism (Z_5) as the inclination angle amplifies the local refractive power according to the Kerkhof model (2.85). Consequently, the vertical field curvature aberration of the camera is typically overlaid by the vertical astigmatism (Z_5) of the windshield, such that the vertical MTF drops.

4.4 Interim conclusion

In summary, the local refractive power assesses the impact of local wavefront curvature modulations, whereas the MTF depends on the global optical properties across the entire

aperture surface. This statement is underpinned by experimental evidence and was proven for the special case of a thin spherical lens.

Additionally, it was shown that local refractive power measurements, which are exclusively taken along the horizontal and vertical direction, are locally invariant under oblique astigmatism. Furthermore, it has been demonstrated that the information content provided by local refractive power measurements is insufficient for extracting more fundamental information about the optical system in terms of the geometrical PSF, wherefore the local refractive power is an inadequate measure for assessing the effect of the blurring operator $\hat{\mathcal{B}}$ on the imaging process.

The concept of the MTF is based on the linearity of the Helmholtz Equation (2.3) and is therefore part of linear system theory. Hence, the MTF of independent optical systems can be multiplied in the frequency domain to retrieve the system MTF, which is the case for the lens and the imager for example. In contrast, the windshield and the camera are not independent components as their individual wavefront modulation pattern interferes in the complex-valued domain of the electromagnetic field ρ. This argument was supported by mathematical and experimental evidence on the interaction of the field curvature of the camera with the refractive power distribution of the windshield. In detail, the horizontal sharpening of the slanted edge in Figure 4.4 demonstrates the non-linear process, which prohibits the use of linear system theory. Consequently, MTF part-level measurements are not possible if both systems entail optical aberrations. As a result, any solution using MTF would have to measure the MTF on the combined system of produced windshield and camera system with their individual production tolerances. This could be either at the production site of the Tier 1 or the OEM. But there are still several important open questions that make this an unattractive proposal: if an assembly is non-compliant, is it worth finding a compliant combination, does it make economically sense? How big are the assembly tolerances fitting the windshield into the car body? If the OEM wants the measurement system at the site of the Tier 1, one should be aware that the assembly of the windshield into the car produces distinct mechanical tolerances, changing the shape and internal tension of the windshield. As we are looking for subtle differences in optical quality, this may affect the pre-assembled camera system as well. These open challenges give rise to the idea of specifying part-level requirements in terms of maximum wavefront aberrations, realized by imposing limits on the Zernike coefficients.

To conclude, it was shown that local refractive power measurements and the MTF are not capable of effectively characterizing part-level requirements for the windshield, such that the effect of the blurring operator $\hat{\mathcal{B}}$ on the imaging process could be neglected by imposing corresponding quality requirements. This leaves us with the puzzle of what could be a way forward? Conducting a large-scale experimental study with thousands of cameras and windshields is impractical and very expensive. Instead, the windshield optical quality needs to be directly linked to the AI performance by simulation, utilizing a physics-based optical threat model. This will be the subject of the subsequent chapter.

5 Sensitivity analysis of AI-based semantic segmentation

This chapter includes elements with editorial changes from

[194] Dominik Werner Wolf, Markus Ulrich, and Nikhil Kapoor. "Sensitivity analysis of AI-based algorithms for autonomous driving on optical wavefront aberrations induced by the windshield". In: *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)* (Oct. 2023), pp. 4102–4111. DOI: 10.1109/ICCVW60793.2023.00443. URL: https://ieeexplore.ieee.org/document/10350923/

marked with an orange line and includes elements with editorial changes from

[190] Dominik Werner Wolf, Alexander Braun, and Markus Ulrich. *Optical aberrations in autonomous driving: Physics-informed parameterized temperature scaling for neural network uncertainty calibration*. (Under review). 2024. arXiv: 2412.13695 [cs.CV]. URL: https://arxiv.org/abs/2412.13695

marked with a cyan line.

As demonstrated in Chapter 4, current optical measures are insufficient for imposing requirements on the windshield quality, such that the impact of the blurring operator $\hat{\mathcal{B}}$ on the imaging process could be effectively suppressed. This gives rise to studying the correlation between different optical measures and the neural network performance in more detail.

5.1 Related work

The interplay of the laws of nature with AI systems is a topic of high relevance. Different scientific fields evolved in this area, e.g. solving nonlinear partial differential equations [145] by introducing a physics-based regularization term or constraining the network architecture to inherently satisfy physical invariants [91]. This trend also spread into the computer vision domain. In particular, Banerjee et al. [6] pointed out that there are three different possibilities to account for governing physical laws by modifying the input data with an observational bias, by adjusting the network architecture with an inductive

bias, or by tailoring the loss function with physics-informed regularization terms, typically denoted as a learning bias. This holistic approach of developing AI systems in line with fundamental, physical prior knowledge also gave rise for the scientific branch of deep optics [185]. In deep optics the camera is interpreted as an encoder [185, 79], which embeds the information about the world into a feature representation, typically a 2D intensity array. This conception triggers the efforts to optimize the camera design end-to-end with the target application, where the decoder typically consists of a deep neural network. The optical encoder and neural network-based decoder architecture is an outstanding example of bionics as similar structures can be observed in nature. For instance, the optical system of living organisms – namely, their eyes – are inherently optimized to best suit the specific conditions of their natural habitats. This adaption involves the layout of the eyes as well as the rearrangement of the brain connections, e.g. the eyes of mantis shrimps have twelve different spectral bands [120, 165] to cope with the diversity of color in coral reefs [185].

This raises the question of whether the hypothesis that higher optical quality in the perception chain automatically leads to superior results in computer vision tasks is really true. Yang et al. [198] addressed this question and pointed out that image quality is not all that matters. In detail, they parameterized the phase modulation induced by a lens in terms of a polynomial expansion and the corresponding coefficients were assigned as trainable parameters for an end-to-end training, targeting on image classification. They find that the learned lens layout achieves higher accuracy – with fewer lens elements – than conventional lenses even though the PSF of the optimal lens layout exhibits long-tails, which is not preferable from an image quality point of view [198]. As this result is contradictory at first glance, they investigated the consistency of the optimal PSF shape across various network models [198].

This finding underpins the validity of the conception of a camera as an optical encoder and triggered research efforts of well renowned industrial leaders. Yow et al. [199] propose a reinforcement learning approach to tailor the lens design, but they identify the definition of a suitable reward mechanism for the agent as an unsolved challenge. Most recently, Dai et al. [36] contributed to the applicability of the deep optics approach in the real world – governed by manufacturing limitations in precision – by introducing data augmentation into the deep optics training pipeline. Their setup is similar to variational Bayesian neural networks, where the weights are sampled from a Gaussian distribution and the corresponding mean and variance are learned during training (see Section 2.4.4). In contrast to traditional variational Bayesian neural networks, the variance is determined by production tolerances a priori and the mean of the distribution regarding the optical hyperparameter of interest is learned exclusively. In summary, deep optics approaches are gaining momentum in the devolvement of novel perception systems.

In a nutshell, image-based deep learning architectures should be optimized holistically as it is done in nature [185]. It has been demonstrated by previous work that the PSF can be tailored to enhance the embedding of information for downstream AI-driven tasks [198].

This violates the assumption of better optical quality equals higher performance and poses the question, which optical quality measures are even suitable to define system-level requirements for machine learning-based autonomous driving functionalities.

The challenge of characterizing an optical merit function that aligns with the objectives of complex AI-based tasks was also highlighted by previous work [129, 16, 199]. Solving this puzzle is a prerequisite for issuing statistical safety guarantees on a predefined confidence level, as it will be required to homologate autonomous systems [16]. Current optical quality measures in the automotive industry are fundamentally limited as elaborated in Chapter 4 and experimental evidence has underpinned this statement, as the AI performance for binary classification can remain constant even if the optical quality in terms of the system MTF differs significantly [129].

This observation gives rise to a set of research questions, which will be systematically addressed in this chapter. First of all, can the decoupling of the system MTF from the AI performance also be verified for semantic segmentation? Are there superior optical measures that impose a bijective relationship between AI performance and optical quality? Furthermore, from a windshield manufacturer point of view, the question might be raised, which Zernike coefficients are the most critical ones for CNN-based predictions? This might indicate where an optimization strategy for the windshield design or manufacturing process should focus on first. Fundamentally, a model prediction without a confidence/uncertainty estimate is worthless. Hence, the research question of how optical measures correlate with the AI performance should also comprise the calibration quality, as it is expected that temperature scaling is negatively affected by dataset shifts [39].

With our work, we will provide evidence that system optical measures, which entail more information about the PSF, exhibit significantly higher correlations in terms of the Chatterjee's rank correlation measure. This conclusion will be demonstrated to hold true across different neural network architectures and datasets. As a second contribution, we will evaluate the local sensitivities of the mIoU and mECE on changes in the Zernike parameterization of the wavefront aberration map by utilizing Shapley values [153] from the domain of explainable AI [116]. The global sensitivities deduced from the Shapley value distributions will reveal valuable inside into the selection process of potential optimization strategies for the windshield. With this knowledge, the windshield could be tuned towards the most optically informative PSF [98], which might differ from the PSF with the highest imaging fidelity. Finally, it will be demonstrated experimentally that optical aberrations induced by the windshield significantly degrade the network calibration. The severity of this issue will motivate further research in this direction in Chapter 6.

5.2 Evaluation setup

Three different CNN-based neural networks will be studied in this chapter (HRNet, MTL, and UNET), which were trained on three distinct datasets (KITTI [55], CARIAD, and A2D2 [56]). As our main focus lies on investigating the relationship between optical system quality and neural network performance, the models employed are considered as black-box models. In detail, we will utilize a pre-trained HRNet (see Section 2.4.1) from Google [63, 124, 2] with a test image batch size of 40. The HRNet was trained on the KITTI [55] dataset and a random instance of the test image batch is visualized in Figure 5.1 for different perturbation scenarios. In addition, an in-house MTL model (see Section 2.4.1) from CARIAD will be considered with a test image batch size of 20. The MTL model was trained on an in-house dataset from CARIAD. The test images are drawn from the corresponding test dataset, which can not be visualized due to confidentiality obligations. Finally, a customized UNET architecture (see Section 2.4.1) is evaluated, which will be introduced comprehensively in Section 6.2, but for the moment we will also consider this network as a black-box model. The UNET will be trained on the A2D2 [56] dataset from AUDI and a random instance from the test dataset is illustrated in Figure 5.2 for distinct perturbation magnitudes. The evaluation of the impact of different Zernike vector parameterizations on the segmentation performance of the UNET will be based on a test image batch size of 196. In general, the test image batch will be degraded by the optical threat model presented in Section 2.1.2 for distinct windshield configurations. For each perturbation scenario, the mean performance and the corresponding standard deviation of the mean will be recorded. The baseline scenario is given by the diffraction-limited case, where it is assumed that the windshield does not modulate the wavefront curvature. Hence, all Zernike coefficients with radial order two or higher are zero. In this thesis, we will restrict the sensitivity analysis to the Zernike coefficients of the second radial order, as these are expected to be the most significant contributors when considering aberrations of windshields [191]. As a consequence, the perturbation scenarios are parameterized by sampling the Zernike coefficients of second radial order equidistantly in the range of $\omega_n \in [-\lambda,\ \lambda]\ \forall\ n \in \{3,\ 4,\ 5\}$.

As an aside, the alert reader might wonder why the test image batch size between the evaluations varies but such a huge margin. The reason behind this lies in the compute resource limitations. Generally, the evaluation should account for all images in the test dataset in order to deduce the most reliable conclusions possible. Unfortunately, this was not realizable within justifiable runtime given the limited computational resources. Consequently, the maximum allowed batch size on each subsystem was used.

Furthermore, it should be noted that, although we parameterize the optical threat model with the same Zernike coefficient vector, this does not imply that the optical quality of different camera systems is affected equally. This dependency on the camera system is imposed by Equation (2.50) for the incoherent PSF. The PSF explicitly depends on the

focal length (if z is approximated accordingly) and the aperture diameter respectively. Since all three datasets under consideration were captured with different cameras and objective lenses (see Section 2.4.2), these setups will demonstrate varying sensitivities to the impact of the wavefront modulation imposed by the windshield. Nevertheless, as the scope of this chapter is to identify generalizable correlations between neural network KPIs and optical quality measures, rather than to derive quantitative requirements, this dependency is not a matter of concern for this study. However, if the objective is to derive system-level requirements for ADAS/AD setups with telephoto cameras, the experiments presented in this chapter must be repeated.

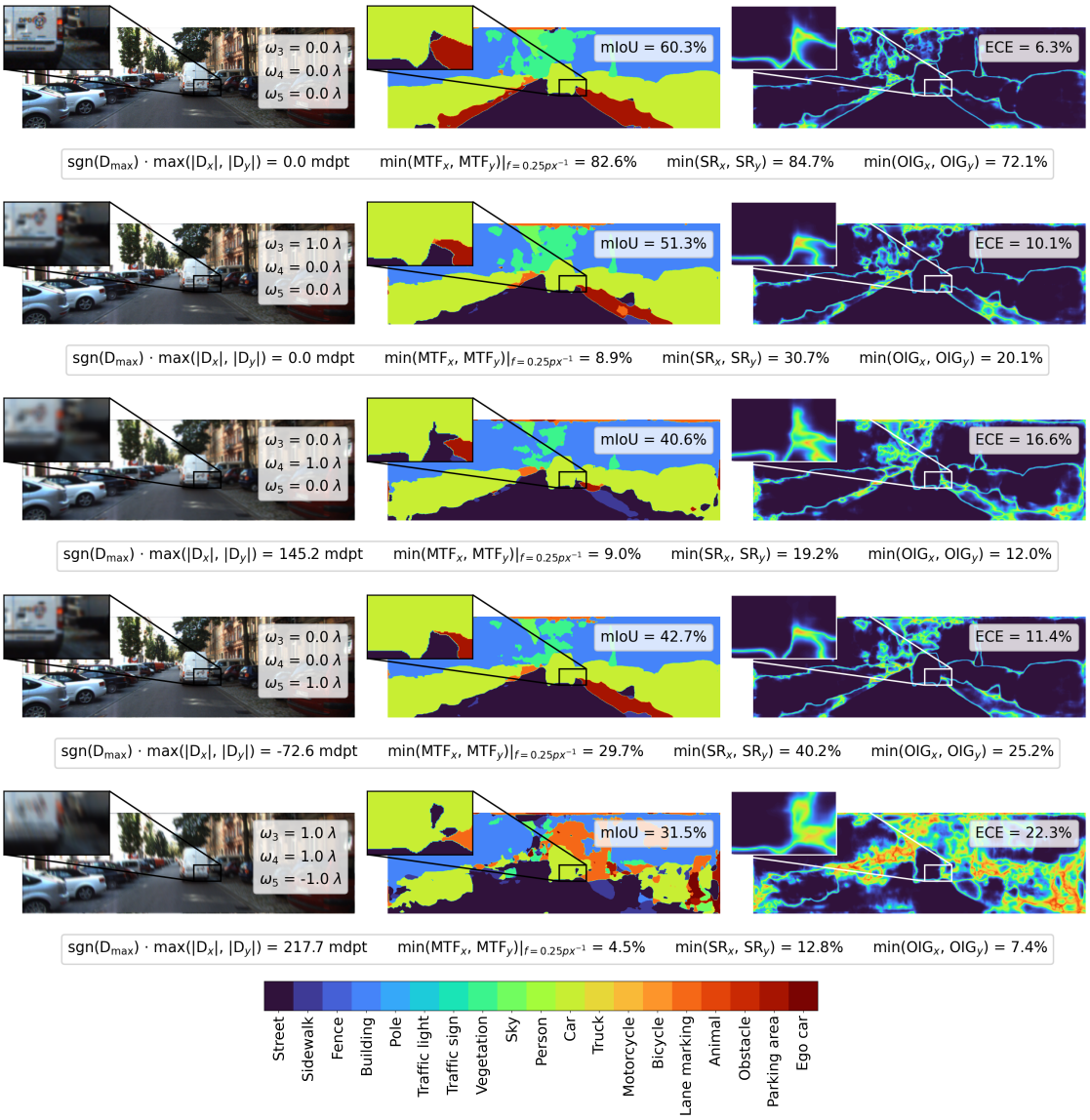


Figure 5.1: On the left, the degradation induced by the optical threat model (see Section 2.1.2) is visualized for a random instance of the KITTI dataset for different Zernike vector parameterizations. In the middle, the semantic segmentation maps obtained by employing a pre-trained HRNet from Google [63, 124, 2] are shown. On the right, the predictive uncertainty is visualized for every pixel based on the variation ratio. For each perturbation scenario, the optical system measures are indicated in the legend.

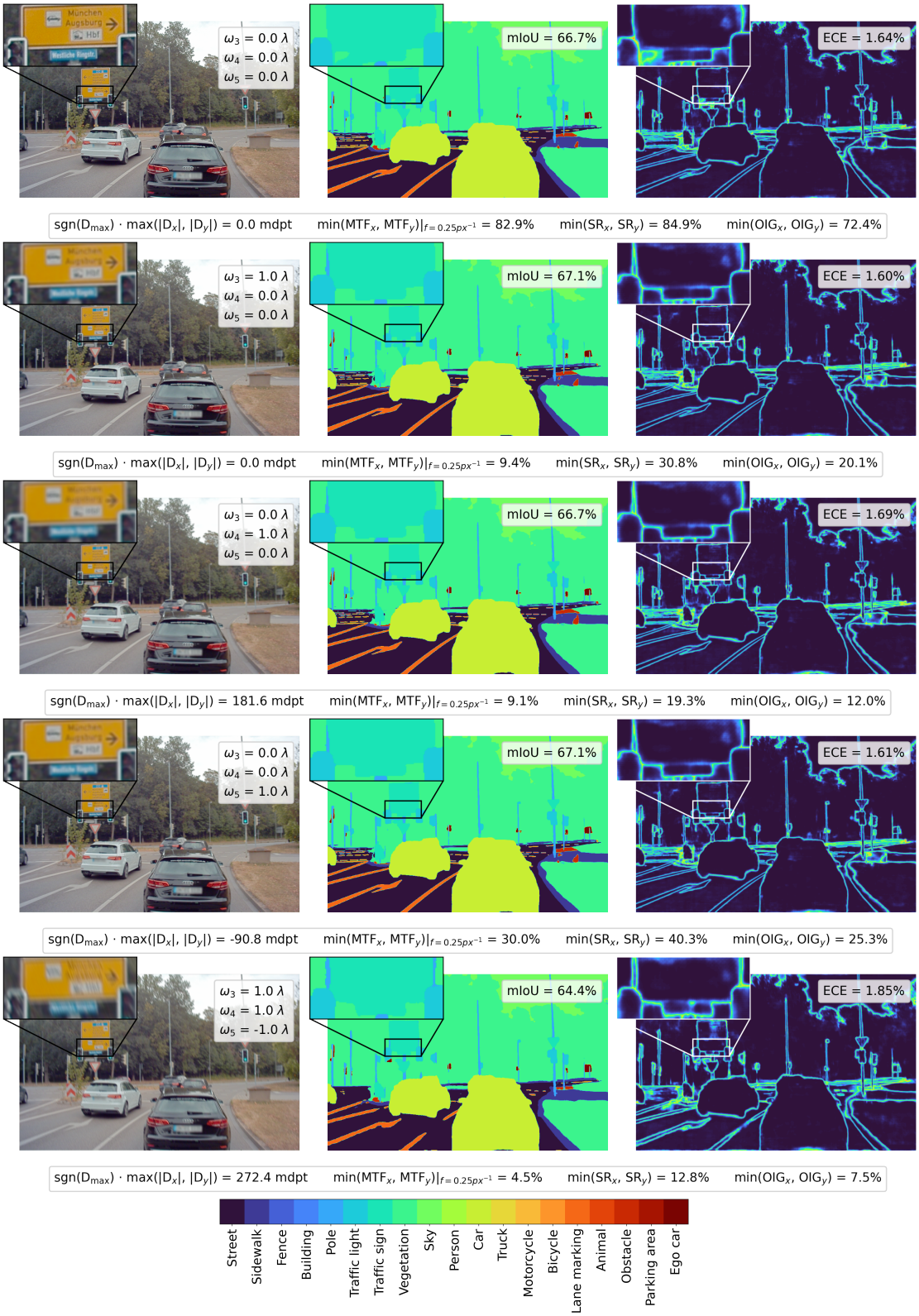


Figure 5.2: On the left, the degradation induced by the optical threat model (see Section 2.1.2) is visualized for a random instance of the A2D2 dataset for different Zernike vector parameterizations. In the middle, the semantic segmentation maps obtained by employing a customized UNET (see Section 6.2) are shown. On the right, the predictive uncertainty is visualized for every pixel based on the variation ratio. For each perturbation scenario, the optical system measures are indicated in the legend.

5.3 Correlation study

First, the results obtained by utilizing the HRNet [63, 124, 2] introduced in Section 2.4.1, trained on the KITTI dataset [55], are presented in Figure 5.3. The results clearly indicate that the correlation to the mIoU is most significant for the Strehl ratio (SR) and the Optical Informative Gain (OIG). In contrast, the current optical measures used for the qualification of windshields, the refractive power (D) and the Modulation Transfer Function (MTF), do not demonstrate a bijective relationship with the mIoU drop. The bijectivity criterion is required to ensure the invertibility of the observed dependency of the mIoU on the optical merit function of interest. This requirement arises from the objective to derive unique optical quality limits for the windshield given a maximum tolerable neural network performance drop.

Furthermore, Figure 5.3 indicates that the calibration of the confidence estimates, assigned to the prediction, is getting biased as the aberration magnitude increases. This trend occurs to be almost symmetrical around the baseline performance if the mIoU drop is considered in relation to the mECE rise. This observation motivates a more detailed analysis of the calibration degradation under optical aberrations induced by the windshield in Section 5.5.

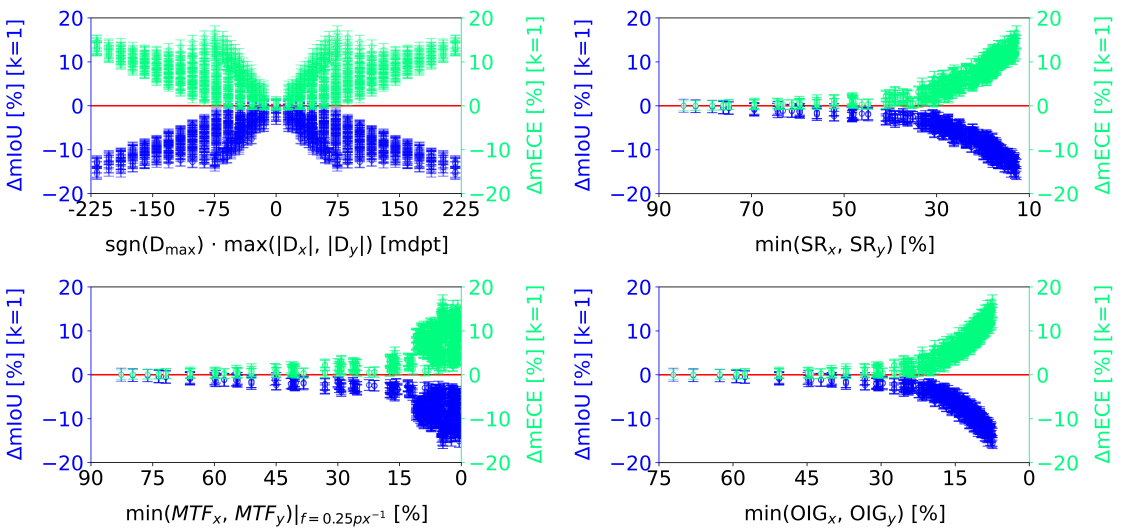


Figure 5.3: The dependency of the mIoU and the mECE on different optical merit functions – namely the maximum local refractive power D_{max} across the horizontal and vertical axes (upper left), the minimum Strehl ratio SR_{min} (upper right), the minimum MTF at half-Nyquist frequency (lower left), and the minimum OIG (lower right) – is shown for the HRNet trained on the KITTI dataset. The uncertainty bars are given by the standard deviation of the mean of the mIoU and the mECE regarding the test image batch of size 40.

Figure 5.4 illustrates the results obtained by utilizing an in-house MTL model from CARIAD for the correlation study. As for the HRNet, each data point corresponds to a different optical attack representing the impact of distinct windshield realizations

on a test image batch of size 20. For each task, the envelope function is computed by binning the data and extracting the corresponding KPI minimum value for each bin. Summing over all head-specific KPIs by considering the task-wise predictive uncertainty as an inverse weighting factor, leads to the combined performance metric. The bin-wise performance spread is illustrated by supplementary box-plots with a special focus on the large aberration regime, which is highlighted on the right-hand side in Figure 5.4. The box-plots clearly indicate that the variance and kurtosis of the bin-wise distributions are significantly reduced if the Strehl ratio and the OIG is utilized as an optical quality measure of the windshield instead of the MTF at half-Nyquist frequency. In summary, the MTL model provides further evidence for the superiority of the Strehl ratio and the OIG over the MTF at half-Nyquist frequency, which is in agreement with the results obtained by the HRNet.

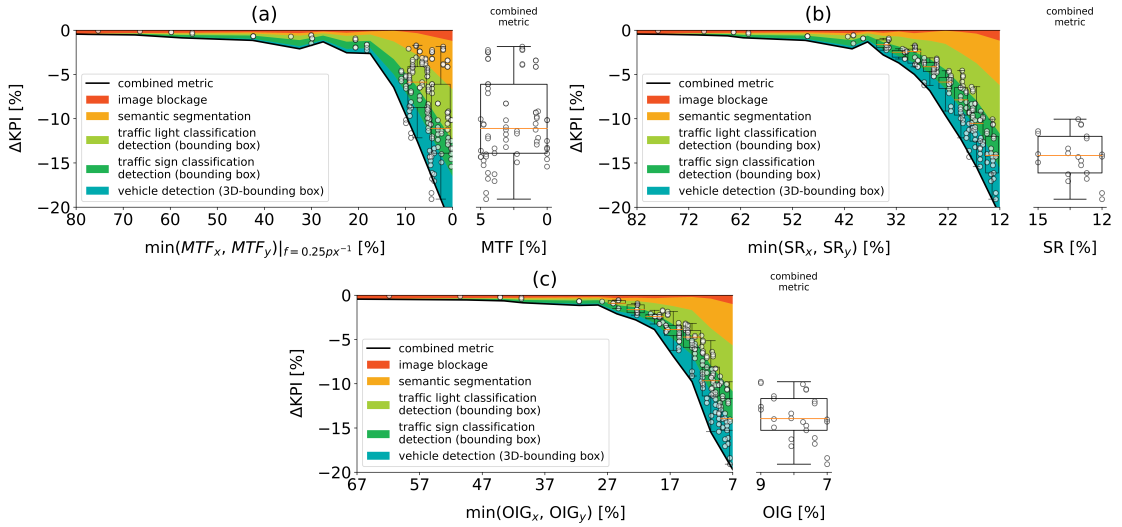


Figure 5.4: The multi-task performance of the MTL model, trained on an in-house dataset from CARIAD, is plotted versus (a) the MTF at half-Nyquist frequency, (b) the Strehl ratio and versus (c) the OIG.

As a third architecture, a customized UNET trained on the A2D2 dataset [56] was used to investigate the correlation between the semantic segmentation performance and different optical measures as depicted in Figure 5.5. In contrast to the previous results, the peak performance is not given in the absence of optical aberrations (diffraction-limited case), which might be counterintuitive at first glance. The mIoU is maximized for instances that reflect the mean-level of optical aberrations within the training dataset. As a consequence, it is of paramount importance to incorporate the optical aberrations within the perception chain proportionally to their occurrence in part-level measurements. By doing so, the augmented training dataset will be centered at the expected optical quality of the produced perception chain, and the mIoU as well as the mECE will be implicitly tuned for this aberration scenario.

This conclusion aligns well with previous work in the field of deep optics [25, 168, 198], where the perception chain is holistically optimized alongside the neural network training. This is done by constructing a differentiable, physics-based optics model and by assigning the corresponding optical parameters as trainable hyperparameters. The contributions in this field [25, 168, 198] strongly indicate that optical quality is not all what you need. To the best of our knowledge, this result has only been shown with respect to the target application performance, e.g., image classification, depth estimation, 3D object detection etc. With our work, we demonstrate that this effect also holds true in terms of the neural network calibration performance.

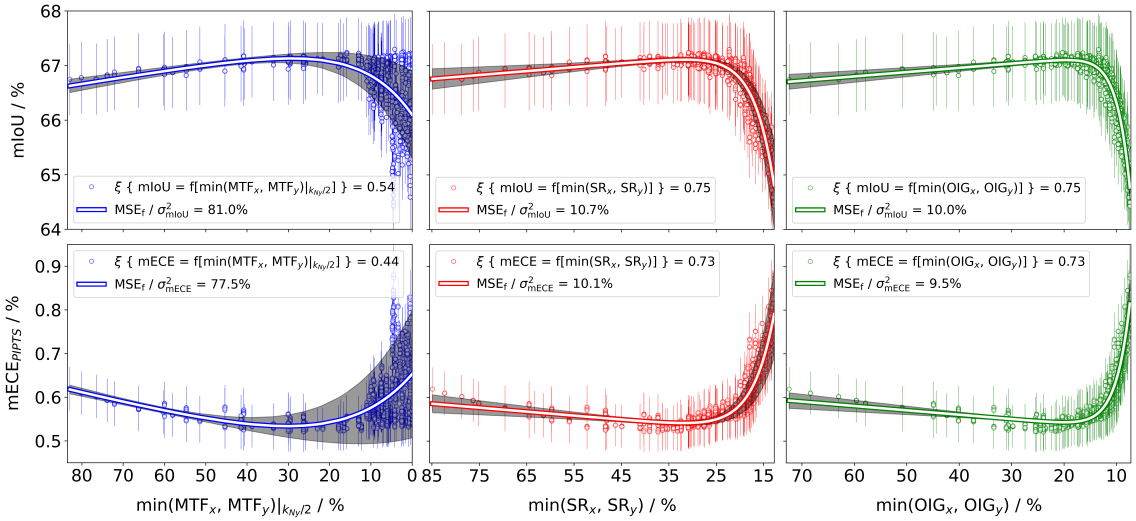


Figure 5.5: The dependency of the mIoU (upper row) and the mECE (lower row) on the MTF at half-Nyquist frequency (left column), the Strehl ratio (middle column) and the OIG (right column) is plotted for the customized UNET architecture trained on the A2D2 dataset. The Strehl ratio and the OIG demonstrate a superior correlation to the mIoU and the mECE in terms of the Chatterjee rank correlation measure than the MTF at half-Nyquist frequency. As a consequence, the regression function from Equation (5.1) also fails to capture the non-existing relationship in the large-aberration regime but it performs well for the Strehl ratio and the OIG, which is quantitatively measured by the ratio of the Mean Squared Error (MSE) over the variance (σ^2), referred to as the unexplained variance component.

In order to quantitatively assess this effect we postulate a regression function f:

$$f\left(x\,;\,\vec{\beta}\right) := \beta_1 \exp\left(\beta_2\left(x-\beta_3\right)\right) + \beta_4 x + \beta_5\,, \tag{5.1}$$

which is supposed to capture the mIoU and mECE performance as a function of the optical quality. The first term accounts for the exponential decay of the mIoU in the large-aberration regime and the exponential increase of the mECE, respectively. Furthermore, the last term denotes an ordinate offset. Finally, the performance gain from the low-aberration regime to the mean-aberration regime is captured by a linear term and the corresponding slope measures the magnitude of the aforementioned effect. In addition, the global extremum of the regression function quantifies the mean optical quality of the

training dataset. The regression function is fit to the data in Figure 5.5 and demonstrates high explanatory power for the PSF-based optical measures.

The regression function f is parameterized by the coefficient vector $\vec{\beta}$ and the combined uncertainty σ_c will be determined by applying the multivariate law of uncertainty propagation [114]:

$$\sigma_c(x_i) = k_{\nu_{\text{eff}}} \sqrt{\left(\vec{\nabla}_{\vec{\beta}} f\right)\Big|_{x_i}^{T} \begin{bmatrix} \sigma_{\beta_1}^2 & \cdots & \rho_{\beta_1,\beta_d}\sigma_{\beta_1}\sigma_{\beta_d} \\ \vdots & \ddots & \vdots \\ \rho_{\beta_1,\beta_d}\sigma_{\beta_1}\sigma_{\beta_d} & \cdots & \sigma_{\beta_d}^2 \end{bmatrix} \vec{\nabla}_{\vec{\beta}} f\Big|_{x_i}} \,. \tag{5.2}$$

The covariance matrix for the parameters β_i is calculated by a Monte-Carlo study [147] considering the batch-wise standard deviation of the mIoU and the mECE respectively. In total $N = 1000$ regression curves were calculated, wherefore the extension factor $k_{\nu_{\text{eff}}}$ is given by $k_{995} = 1.96$ for a confidence level of 95% [68, 140]. Finally, the symmetrical interval spanned by the combined uncertainty σ_c determines the confidence bands around the regression function, as illustrated in Figure 5.5 in gray.

The explanatory power of the postulated regression function (5.1) is further quantified in terms of the unexplainable variance, which is given by the ratio of the mean squared error (MSE) [169] over the variance of the dataset itself. Statistically, the MSE measures the ensemble spread around the regression line. Benchmarking regression models according to the unexplainable variance is favorable because it effectively measures how well the regression model outperforms the naive estimate given by the arithmetic mean [192]. In addition, if the complexity of the regression model is increased the degrees of freedom in the MSE computation decreases, which acts as a penalty for more complex regression models like in Ridge regression [159, 162, 47]. The quantitative values for the unexplainable variance are presented in Figure 5.5 and indicate that the model function does not significantly outperform the arithmetic mean in case of the MTF. On the other hand, the model function is superior for the Strehl ratio and the OIG. Consequently, the coefficient vector $\vec{\beta}$ can be interpreted as a sensitivity vector in case of the Strehl ratio and the OIG. Of special importance are β_4 for the linear performance drift in the low-aberration regime and the coefficients β_1 and β_2 for the exponential decay in the large-aberration regime. These sensitivity coefficients can be employed for determining valid operational domains for the perception chain.

The main reason why the proposed regression function does not sufficiently capture the variability of the mIoU and the mECE as a function of the MTF lies in the lack of correlation. For all three optical measures of interest – the MTF at half-Nyquist frequency, the Strehl ratio as well as the OIG – the Chatterjee rank correlation measure ξ, introduced in Section 2.5, is computed. The correlation is most evident for the Strehl ratio and the OIG ($\xi_{1331} = 0.75$), whereas the MTF at half-Nyquist frequency is a significantly worse indicator for the AI performance ($\xi_{1331} = 0.54$).

Nevertheless, it is evident that the MTF at half-Nyquist frequency fits the regression function well in the low-aberration regime but the huge spread in the codomain for the large-aberration regime leads to a wide confidence band around the global trend. The correlation of the MTF breaks down for severe optical aberrations because of the monotonicity violation in the large-aberration regime. As long as the MTF function is monotonically decreasing, the value of the MTF function at half-Nyquist frequency will correlate with the area enclosed by the MTF curve. Since the area under the MTF curve, which approximates the Strehl ratio if normalized by the diffraction-limited case, shows a robust correlation to the mIoU across the entire spatial frequency domain, the MTF at half-Nyquist frequency is a valid optical performance indicator as long as the MTF curve is monotonically decreasing. As the optical aberrations in the automotive industry are typically too severe in magnitude to satisfy the monotonicity constraint, the MTF at half-Nyquist frequency is not a suitable measure for safeguarding AI-based autonomous driving algorithms against optical perturbations. Consequently, system MTF requirements evaluated at half-Nyquist frequency should be considered as invalid for the large-aberration regime.

5.4 Sensitivity study

The regression function introduced in this work by Equation (5.1) captures the dependency of the mIoU and mECE under dataset shifts induced by optical aberrations of the windshield. Hence, the local sensitivity of the mIoU on changes in the optical quality, e.g. due to windshield aging, can be evaluated by the first derivative. From an engineering point of view, it is much more interesting to investigate, which aberration pattern has the most significant impact on the mIoU, such that an optimization of the windshield production process could be initiated. A suitable local sensitivity measure is given by the Shapley value introduced in Section 2.4.6. If the Zernike coefficients, which parameterize the wavefront modulation induced by the windshield, are treated as features, then the local sensitivity of the mIoU on changes in the windshield aberration pattern can be studied. Figure 5.6 depicts the Shapley value distribution, given a merit function of interest, across different aberration scenarios. The Shapley values are determined by utilizing the pre-trained HRNet from Google [63, 124, 2] trained on the KITTI dataset [55]. The shape of the Shapley value distribution for two distinct features may suggest differences in the global feature importance.

In general, the Shapley values were normalized for comparability reasons to the effect of ω_4, which physically represents defocus. Each point in Figure 5.6 quantifies the local derivative of the neural network prediction w.r.t. the input features, here parameterized in terms of the Zernike coefficients of second radial order. This holds true for the mIoU and the mECE as a merit function. In contrast, the Shapley values obtained by utilizing

an optical quality measure as a merit function are independent from the neural network and just reflect the local derivative that is propagated through the optical threat model. The local derivatives are computed for every Zernike polynomial Z_n and the magnitude of the corresponding Zernike coefficient ω_n is visualized by the color bar. The Shapley value density correlates linearly with the vertical extent of the depicted points, similar to how bar height reflects frequency in a histogram.

It is evident that the global feature importance significantly varies between the optical merit functions and the neural network measures indicating a non-linear mapping as described by the regression function in Equation (5.1). Furthermore, the results for the refractive power reflect the inherent constraint on how refractive power is measured in the automotive industry as elaborated in detail in Chapter 4. Hence, the sensitivity regarding ω_3 vanishes. An observation common to all studied merit functions is that defocus (ω_4) has the greatest impact on the performance among the Zernike coefficients of second radial order. As a consequence, the manufacturing process of the windshield should first be optimized such that the defocus is minimized. This conclusion also holds true for the MTL model from CARIAD as well as the customized UNET architecture, however, the graphs are omitted in the interest of compactness.

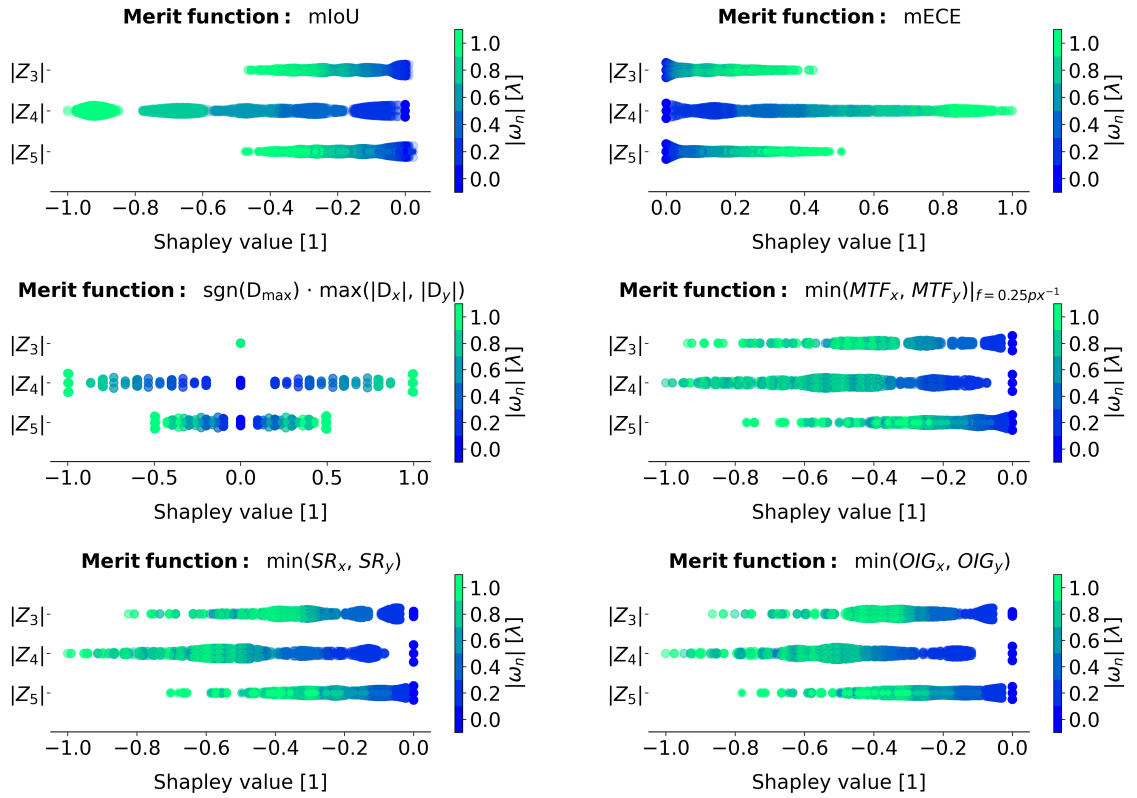


Figure 5.6: The sensitivities of the mIoU and the mECE as well as the sensitivities of several optical KPIs on wavefront aberrations, parameterized by Zernike coefficients ω_i, are quantified and visualized in terms of Shapley values. The sensitivities are retrieved by utilizing the pre-trained HRNet from Google [63, 124, 2] trained on the KITTI dataset [55]. The impact of an induced defocus (Z_4) surpasses the effect of oblique- (Z_3) and vertical astigmatism (Z_5) for all merit functions considered.

5.5 Calibration degradation

The observed mirror symmetry of the mIoU and the mECE with respect to the abscissa in Figure 5.3 and with respect to the ordinate in Figure 5.6 motivates further studies on the calibration degradation under optical aberrations induced by the windshield. The mirror symmetry originates from the fact that the model becomes more and more overconfident as the aberration amplitude increases. For the HRNet, this effect is so pronounced, such that the accuracy degradation is roughly a magnitude larger than the confidence drop, as illustrated in Figure 5.7. This ultimately imposes the observed symmetry between the mIoU and the mECE. In contrast to the HRNet, the customized UNET architecture does not show this level of susceptibility due to the implicit tuning of the network architecture, which will be elaborated in detail in Section 6.4. Furthermore, for the MTL model from CARIAD the task-wise predictive uncertainty estimates are not explicitly available, wherefore the calibration degradation can not be evaluated.

The reliability diagram for the HRNet is depicted in Figure 5.7 and demonstrates a significant overconfidence of mECE $= 15.6\%$ in the diffraction-limited case $(\omega_n = 0 \; \forall \; n \in \mathbb{N}_0)$. Hence, the publically available HRNet is not well calibrated even if no dataset shift is considered. As the accuracy degradation and the confidence drop is incoherent, the calibration measure identifies growing bin-wise confidence biases if different windshields are considered, as indicated by the red triangles. In the case of low prediction confidences and low prediction accuracies, the binned network accuracy seems to slightly increase if aberrated test data is used. This behavior is counterintuitive and represents an artifact of the visualization method. Essentially, optical aberrations drive the probability flow of the confidence distribution towards lower values shifting the predictions into low-confidence bins. Since the domain is limited and quantized, the bin composition varies affecting the binned accuracy. Physically, low-confidence predictions in semantic segmentation mostly correspond to class area borders. If the image is perturbed, then the contours are getting blurred as illustrated by Figure 2.5, which results in lower prediction confidences.

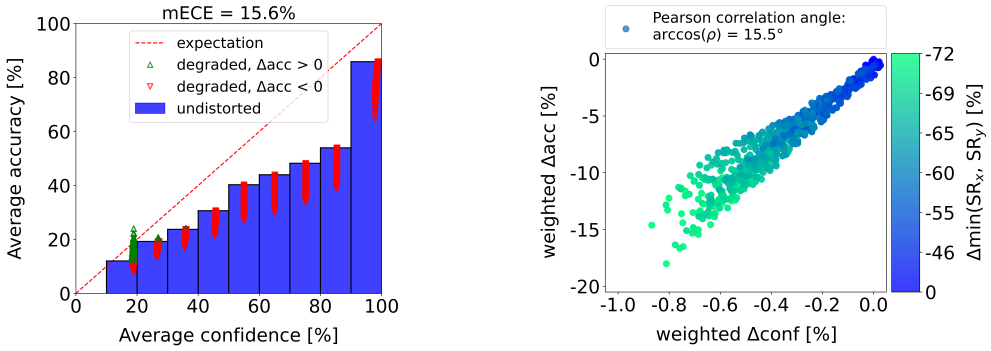


Figure 5.7: On the left, the reliability diagram for the HRNet is shown for different perturbation scenarios, where each triangle per bin corresponds to a distinct windshield configuration. The mECE increases as the aberrations are amplified, which is corroborated by the weighted accuracy vs. confidence correlation plot shown on the right-hand side.

5.6 Interim conclusion

The results of this Chapter clearly indicate that optical measures, which incorporate additional information about the PSF (esp. Strehl ratio and OIG), demonstrate a superior correlation to neural network KPIs, esp. mIoU and mECE. This conclusion was quantitatively demonstrated by evaluating the Chatterjee rank correlation measure and a regression function was postulated in Equation (5.1) that adequately captures the relationship between neural network KPIs and PSF-based optical measures. By employing those system-level optical measures, suitable ADAS working requirements can be derived to suppress the effect of the blurring operator. On the contrary, the MTF at half-Nyquist frequency is insufficient as a quality criterion for windshields, as the MTF is non-monotonic for the large aberration regime.

In addition, the sensitivity study utilizing Shapley values indicated that the optimization of the manufacturing process of the windshield should first target on defocus (ω_4) aberrations as they induce the highest performance degradation among the Zernike polynomials of second radial order.

Finally, the confidence calibration is heavily affected by dataset shifts induced by optical aberrations of the windshield. This has been observed for the uncalibrated HRNet and the customized UNET, which was calibrated for the diffraction-limited case by employing temperature scaling [70]. This conclusion aligns well with previous studies on the robustness of temperature scaling under dataset shifts [39], which is the driving force for performance drops in terms of the mECE. In order to safeguard autonomous driving systems, it is of paramount importance to have well calibrated confidence/uncertainty estimates for the predictions in order to make informed decisions about the system reliability. If the neural network predictions, which e.g. steer an autonomous driving car, are made with a confidence falling below a predefined safety threshold, then the system is supposed to automatically switch into a safe state mode. As optical aberrations might bias the estimation of the predictive confidence/uncertainty – leading to a violation of the calibration condition defined in Equations (2.97-2.98) – the system might continue to operate despite exceeding the safety threshold, increasing the risk for passengers and ultimately compromising trustworthiness. In order to prevent this scenario from happening, the next Chapter will propose a new calibration strategy, which implicitly accounts for dataset shifts by incorporating a physical inductive bias into the neural network architecture.

As a final remark, the results presented in this Chapter were consistent across three different CNN-based architectures (HRNet, MTL and UNET) as well as three different datasets (KITTI, CARIAD and A2D2). Consequently, it can be hypothesized that the conclusions are network architecture independent as long as CNN-based models are considered.

6

Enhancing calibration robustness by physical inductive biases

This chapter includes elements with editorial changes from

[190] Dominik Werner Wolf, Alexander Braun, and Markus Ulrich. *Optical aberrations in autonomous driving: Physics-informed parameterized temperature scaling for neural network uncertainty calibration*. (Under review). 2024. arXiv: 2412.13695 [cs.CV]. URL: https://arxiv.org/abs/2412.13695

marked with a cyan line.

Dataset shifts are of major concern for the homologation of safety-critical autonomous driving functions. Dataset shifts are generally given if the network infers information from an instance, which does not share the same underlying probability density function as the training dataset distribution. Consequently, the neural network utilizes the learned functional relationship between input and output for extrapolating into a different domain. This gives rise to a performance drop of the model. The performance is not only affected in terms of the target KPI but also the corresponding uncertainty estimation might become biased [194] as observed in the previous Chapter.

Calibrated uncertainties are an essential requirement for a physically sound sensor fusion process and for system monitoring. On the one hand, fusing feature attributes should incorporate the associated embedding uncertainties in order to achieve the most reliable latent space representation. On the other hand, in order to safeguard autonomous systems, the prediction uncertainties need to be tracked. If the uncertainty is low, and hence the situation is identified with sufficient confidence, a reliable decision can be made. If the confidence is insufficiently low then an independent secondary system must contribute additional information for the decision-making process or the system has to fall back into a safe state mode automatically. Consequently, the trustworthiness of the uncertainty estimates, quantified by the ECE, is decisive for the reliability of autonomous driving systems. However, if a dataset shift is induced, e.g. by optical aberrations of the windshield as done in the previous Chapter, then the calibration of the network confidences - and uncertainties vice versa - breaks down and the network becomes increasingly overconfident [194].

6.1 Related work

Finding the optimal temperature is the key for establishing model trustworthiness. The straightforward way to determine the temperature is given by minimizing the negative log-likelihood [70]. Unfortunately, this standard Temperature Scaling (TS) methodology is highly limited in terms of the model information capacity (one degree of freedom).

As an extension, Ensemble Temperature Scaling (ETS) [207] computes a weighted average over three different calibration maps, the TS calibrator with adjustable temperature T, TS with $T = 1$ (identity mapping) and TS with $T = \infty$ (uniform mapping). Hence, ETS has four degrees of freedom.

In order to further increase the information capacity of the calibration method, Parameterized Temperature Scaling (PTS) [167] was proposed. PTS leverages a neural network to predict an instance-wise temperature based on the corresponding logits tensor, while preserving model accuracy.

A similar methodology is Sample-Dependent Adaptive Temperature Scaling [89], which also predicts an instance-wise temperature. In contrast to PTS, the approach leverages the latent space representation of a Variational Autoencoder (VAE) [96] as the input for a post-hoc Multi-Layer Perceptron (MLP) for predicting an instance-wise temperature. The benefit of using the VAE's latent space embeddings instead of the logits tensor for the post-hoc MLP lies in the effectiveness of the VAE to cluster the predictions based on their calibration quality, which improves the calibration performance of the MLP under distribution shifts [89].

Most recently, Adaptive Temperature Scaling (ATS) [102] has been proposed as a calibration technique, which enhances the reliability against out-of-distribution samples without the need of training a post-hoc MLP calibrator. The core idea of ATS lies in computing an instance-wise temperature based on the intermediate layer activations of the baseline neural network. After training, the Cumulative Distribution Function (CDF) of the mean activation for each layer is computed across the entire training dataset. During inference, the layer-wise mean activations are compared to the precomputed CDFs to calculate layer-wise p-values [151, 169], which are mapped to an instance-wise temperature. In addition to an enhanced calibration performance, low temperatures indicate in-distribution samples, while high temperatures suggest out-of-distribution inputs [102].

For all accuracy preserving calibration methods mentioned so far, it is required to assume an uncertainty estimator. In mathematical terms, calibration quality metrics measure the bias of a predictive uncertainty estimator. Consequently, neural network calibration and predictive uncertainty estimation are two distinct concepts. A perfect predictive uncertainty estimator is unbiased such that the calibration temperature is equal to one. There are several different methods to estimate the predictive uncertainty for semantic segmentation [53] as elaborated in detail in Section 2.4.4. For this study, the variation

ratio [117] is selected as a measure for the predictive uncertainty because it is invariant under differences in the probability mass allocation over the wrong classes. Hence, the variation ratio inherently accounts for an essential constraint of the negative log-likelihood loss, aka. cross-entropy loss. The same does not hold true for the Shannon entropy [152] for example, which is the standard measure for the total predictive uncertainty according to information theory [80], as introduced in Section 2.4.4. This discrepancy gives rise to a degree of freedom in the uncertainty evaluation as entropy-based uncertainties for two independent instances might differ even though the model confidence predictions are equivalent [189].

For the scope of this thesis, the distribution shift is imposed by optical aberrations within the perception chain, e.g., the windshield or diverse weather phenomena. As a consequence, it seems natural to incorporate this prior knowledge into the calibration process. In order to do so, the optical aberrations need to be estimated online alongside the target application. The most fundamental way to characterize optical aberrations is given by quantifying the optical path difference map in terms of the Zernike coefficients [204, 205, 10]. Utilizing a neural network for predicting the Zernike coefficient vector is a well established approach in astronomy [121, 4] as a way to replace the need for on the fly Shack-Hartmann measurements. The Very Large Telescope (VLT) of the European Southern Observatory (ESO) uses the information about the Zernike coefficients to perform an online correction of the wavefront aberrations induced by the atmosphere [125]. This is realized by adjusting deformable mirrors, which is a technique from adaptive optics [73].

The work of Jaiswal et al. on physics-driven turbulence image restoration with stochastic refinement [88] utilizes the Zernike coefficients to parameterize a physics-based turbulence simulator. By coupling the vision-transformer-based [41] image restoration network with the Fourier-optical aberration model during training, they are able to effectively disentangle the stochastic degradation caused by atmospheric turbulence from the underlying image. This enhances the generalizability of the image restoration network across real-world datasets with varying turbulence strength [88]. As a consequence, by incorporating a physical inductive bias to the transformer architecture they effectively reduce the sensitivity of their target application on dataset shifts induced by turbulence-driven optical aberrations.

Those promising results in the field of image restoration gives rise to the question: Can the calibration performance of post-hoc methods such as PTS also be enhanced by utilizing physical prior knowledge? How to incorporate information about optical aberrations within the perception system into the calibration pipeline? Is it feasible to directly utilize the ECE as a loss function for training the post-hoc calibration network? Is it possible to implement a real-time calibration mechanism that accounts not only for static optical aberrations but also for dynamic perturbations arising from factors such as windshield aging, weather conditions, and similar factors?

We will seize the idea of Jaiswal et al. [88] and transform their concept to the domain of neural network calibration. We will accomplish this by coupling a post-hoc calibration network with physical prior knowledge. As the thesis deals with dataset shifts induced by optical aberrations of the windshield, the physical prior consists of the Zernike coefficients of second radial order, which are the most dominant components for windshields, and the target application is semantic segmentation instead of image restoration. In detail, we will build upon the state-of-the-art PTS [167] approach by incorporating a physical inductive bias [6] in order to increase the calibration robustness under dataset shifts induced by optical aberrations. This novel layout, which adds physical prior knowledge to PTS, will be termed as Physics Informed Parameterized Temperature Scaling (PIPTS). As a baseline model for semantic segmentation, a tailored UNET will be employed utilizing a negative log-likelihood loss, which makes it favorable to rely on point-wise predictive uncertainty calibration estimators because it matches the nature of the ground truth label distribution, which allocates all statistical mass to the ground truth class and analytically resembles the Kronecker delta function. For this reason, the ECE as a point-wise calibration measure is employed.

In order to train the PIPTS calibrator, we will employ the ECE as a loss function. Generally, any loss function must be continuously differentiable (C^1) to enable backpropagation. This constraint is not fulfilled by the binned ECE measure. As a consequence, we propose a smoothed and modulated variant of the ECE measure, enabling its use as a loss function for neural network training.

The novel PIPTS layout requires the prior knowledge about the Zernike coefficient vector that parameterizes the optical system. Consequently, we will extend the UNET architecture such that the Zernike coefficients of second radial order are estimated alongside the segmentation map. This adjustment of the baseline layout will not only serve the goal of ensuring valid confidence calibration under dataset shifts induced by optical aberrations, but it will also enable us to monitor the optical quality of the perception pipeline providing a mechanism for proactive maintenance and fault detection – a crucial capability if long-term effects such as windshield aging are considered.

In general, the benefits of PIPTS will be demonstrated for semantic segmentation because of the assumption that a dataset shift in terms of sharpness will affect a pixel-wise prediction the most. At this point, it is important to underscore that, irrespective of the specific AI task selected, this methodology exhibits encouraging potential for effective generalization across a broad spectrum of tasks.

6.2 Baseline multi-task network

In order to solve the target task and to provide the physical prior for the PIPTS network, a CNN-based architecture with two coupled decoder heads and an additional residual

encoder is utilized for identifying the Zernike coefficients of the second radial order (oblique astigmatism ω_3, defocus ω_4, and orthogonal astigmatism ω_5). The layout of the multi-task network is based on the UNET [149] architecture and the predicted Zernike coefficients shall estimate the effective wavefront aberration map of the overall optical system consisting of windshield and ADAS camera lens.

The network will be trained on the A2D2 [56] dataset because it provides pixel-wise labels for high resolution images with 1920×1208 pixels that were captured without a windshield. This is an essential requirement for the degradation model introduced in Section 2.1.2, which is applied to enrich the data heterogeneity. As telephoto cameras are the most sensitive ones regarding optical aberrations induced by the windshield, only the narrow-view, front-center camera (Sekonix SF3325-100 [56]) of the A2D2 segmentation dataset is utilized. Consequently, the training dataset comprises 5400 images and the test dataset consists of additional 1350 images, which corresponds to a test ratio of 20%. Furthermore, the labeling taxonomy of the Cityscapes dataset [32] was adopted to ensure the comparability of the presented results. For that reason, the ground truth annotations for the 38 A2D2 classes were mapped to the 19 classes Cityscapes taxonomy. The training of the neural network was performed on two Nvidia RTX A6000 with 48GB. The distributed training was terminated after the validation loss reached its minimum. To accomplish this, a learning rate schedule was applied that reduced the learning rate by a factor of 10 if the validation loss did not improve during the last 40 epochs and the training was finally terminated if the validation loss did not improve at all within the last 100 epochs.

In order to enrich the A2D2 dataset with the optical aberrations induced by different windshield configurations, the Fourier optical degradation model presented in Section 2.1.2 is employed for data augmentation. The range of the Zernike coefficients must be sufficiently sampled, such that the statistical complexity of the perception chain - individual part tolerances and installation tolerances - is accounted for. In this work, a uniform grid sampling of the Zernike coefficients of the second radial order in the range of $\omega_n \in [-\lambda,\ \lambda]$ is applied. This does not guarantee that the complexity of the production process is sufficiently reflected, but it serves as a baseline for a proof of concept study.

The layout of the baseline multi-task network is visualized in Figure 6.1 and further specified in Table 6.1. In the subsequent paragraphs, the architecture will be discussed in detail. In general, the model adopts the encoder-decoder structure presented in Section 2.4.1. The shared encoder consists of five encoder blocks with the number of filters doubling in each subsequent block. Each encoder block consists of two convolution layers followed by a batch normalization layer and a max pooling layer for downsampling. After the batch normalization layer, the computational graph is split into two branches to bypass information to the decoder for alleviating the vanishing gradient problem [75, 171, 202].

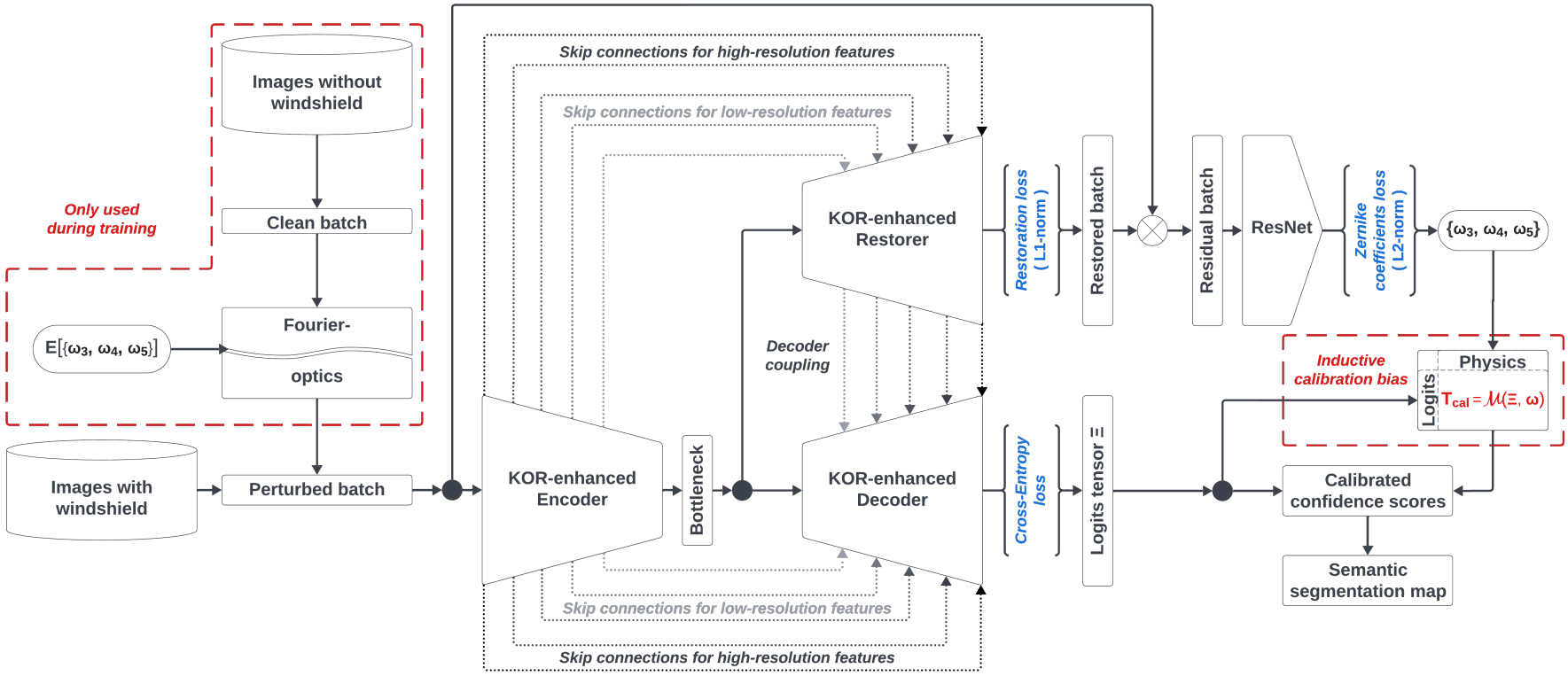


Figure 6.1: The layout of the multi-task network for semantic segmentation and for predicting the effective Zernike coefficients of the optical system is shown. The multi-task network builds upon the UNET architecture with two coupled decoder heads and a downstream ResNet encoder for retrieving the Zernike coefficients of the second radial order. Additionally, the Fourier optical degradation model for the data augmentation process and the post-hoc PIPTS calibration network are indicated. The PIPTS calibrator extends the PTS approach by incorporating a physical inductive bias for ensuring the trustworthiness of the baseline multi-task network predictions under optical aberrations.

Multi-Task Network (batch size, v := 14)					
Layer			**Output Shape**		
Input			(v, 604, 960, 3)		
Encoder Block 0			(v, 302, 480, 19)		
Encoder Block 1			(v, 151, 240, 38)		
Encoder Block 2			(v, 76, 120, 76)		
Encoder Block 3			(v, 38, 60, 152)		
Encoder Block 4			(v, 19, 30, 304)		
Bottleneck			(v, 19, 30, 304)		
Segmentation (SEG) Decoder		**Restoration (RES) Decoder**		**Zernike Regression (REG)**	
Layer	**Output Shape**	**Layer**	**Output Shape**	**Layer**	**Output Shape**
Decoder Block 4	(v, 38, 60, 304)	Decoder Block 4	(v, 38, 60, 304)	(Input, RES)	(v, 604, 960, 6)
Decoder Block 3	(v, 76, 120, 152)	Decoder Block 3	(v, 76, 120, 152)	ResNet Cell 0	(v, 302, 480, 19)
Decoder Block 2	(v, 151, 240, 76)	Decoder Block 2	(v, 151, 240, 76)	ResNet Cell 1	(v, 151, 240, 38)
Decoder Block 1	(v, 302, 480, 38)	Decoder Block 1	(v, 302, 480, 38)	ResNet Cell 2	(v, 76, 120, 76)
Decoder Block 0	(v, 604, 960, 19)	Decoder Block 0	(v, 604, 960, 19)	ResNet Cell 3	(v, 38, 60, 152)
Output SEG	(v, 604, 960, 19)	Output RES	(v, 604, 960, 3)	ResNet Cell 4	(v, 19, 30, 304)
–	–	–	–	Flattening	(v, 173280)
–	–	–	–	Dense Unit 0	(v, 19)
–	–	–	–	Dense Unit 1	(v, 19)
–	–	–	–	Dense Unit 2	(v, 19)
–	–	–	–	Dense Unit 3	(v, 19)
–	–	–	–	Dense Unit 4	(v, 3)
–	–	–	–	Output REG	(v, 604, 960, 3)
Concatenated Output (SEG, RES, REG)			(v, 604, 960, 25)		

Table 6.1: Layer composition of the multi-task network. The neural network comprises 16,380,894 parameters, of which 16,366,568 parameters are trainable and 14,326 parameters are non-trainable. Assuming a batch size of $v = 14$, the model consumes approximately 9.27 GB of GPU memory under a mixed precision policy (float16).

The bottleneck of the UNET, distinguished by the lowest spatial feature resolution, consists of two convolution layers and subsequent batch normalization layers. After the bottleneck, the latent space representation is supposed to provide an embedding of the input information that optimally reflects the degrees of freedom of the underlying problem. Subsequently, the embedding is fed into two decoder heads. The decoder head for restoration aims to eliminate the optical aberrations induced by the windshield and the decoder head for semantic segmentation targets on the pixel-wise classification.

The decoder head for restoration consists of five transposed convolution blocks with the number of filters halving in each subsequent block. Each transposed convolution block consists of a transposed convolution layer followed by a batch normalization layer and a merging node in order to incorporate the high-fidelity information provided by the skip connection of the corresponding encoder block. The concatenated tensor is then smoothed by two subsequent convolution layers with stride equal to one, to preserve the dimensions.

The decoder head for semantic segmentation is similar in structure to the decoder head for restoration, but the high-fidelity information from the encoder is merged with the corresponding restoration layer before entering the concatenation layer in the decoder. This decoder coupling is supposed to enhance the segmentation performance against optical aberrations.

The downstream residual encoder for the Zernike coefficients regression consists of five ResNet [75] cells to predict the aberrations in terms of the Zernike coefficients of second radial order. Each ResNet cell consists of two convolution layers and two batch normalization layers, as well as a concatenation layer to merge the convolved signal with the input signal to alleviate the vanishing gradient problem [75, 171, 202]. Finally, a max pooling layer is employed for downsampling. After the ResNet cells, the signal is flattened and fed into five dense layers with batch normalization and dropout for regularization.

The multi-task network is trained by utilizing a customized loss function. The loss function consists of three components. The first component is the negative log-likelihood for semantic segmentation, aka. cross-entropy. The log-likelihood term is focused by an additional factor of $(1-\hat{p}_i)^\gamma$, which enforces a focus of the learning process on classes that are hard to learn [112]. Furthermore, class balancing is applied to equalize the representation of different classes within the dataset [42]. The weighting factors τ_i for each class i are calculated based on their occurrences in the dataset:

$$\tau_i = \frac{\log\left(1.1 + \frac{c_i}{N}\right)^{-1}}{\sum_{i=1}^{N_c} \log\left(1.1 + \frac{c_i}{N}\right)^{-1}}, \tag{6.1}$$

where c_i denotes the number of instances for class i and N characterizes the total number of pixels within the A2D2 dataset. The second term quantifies the mismatch between the unperturbed image and the restored image by utilizing the L1-norm [210]. The third term quantifies the discrepancy between the predicted Zernike coefficient vector and the ground truth vector by applying the L2-norm. All components are summed up by considering individual weighting factors, which are determined during hypertuning.

In addition to the weighting factor for the restoration loss and the Zernike loss, there are four other hyperparameters to tune:

- The weighting factor of the Kernel-Orthonormality-Regularizer (KOR) term (see below), which is added linearly to the loss function as a penalty.
- The learning rate, which determines the increment in the optimization process.
- The focal loss exponent γ.
- The batch size, which relates to the number of images processed before the trainable variables are updated.

When employing a large batch size, the model's quality often degrades, particularly in terms of its generalization capabilities. Models with large batch sizes are prone to reaching sharp minima in the loss landscape, which are generally associated with reduced generalization performance. Conversely, small batch sizes tend to converge to flatter minima due to the inherent noise in the gradient estimation [93].

For the multi-task network, as for the PIPTS model, the input and output is normalized to zero mean and unit variance in order to equalize the dissimilarity of feature units.

The Gaussian Error Linear Unit (GELU) [78] is used as an activation function for the shared encoder, the segmentation decoder, and the restoration decoder. GELU is defined as $x\phi(x)$, where $\phi(x)$ is the standard Gaussian cumulative distribution function. The GELU activation function returns a likelihood-based output, unlike the Rectified Linear Unit (ReLU) [136], which simply gates the input according to the sign. Generalizing from monotonic (e.g. ReLU) to non-monotonic (e.g. GELU) activations can increase a neuron's discriminative capacity as has been demonstrated for the XOR problem [9]. For the regression task, the Exponential Linear Unit (ELU) [29] is employed in the ResNet cells and the dense layers. Moreover, the activation functions used in the output layers are specifically tailored for their respective tasks. The final layer of the segmentation head applies a softmax activation function. The restoration decoder implements the ReLU as an activation function for the output layer. Furthermore, the regression output is obtained by a linear activation function.

In order to enhance the generalization capabilities of the model, Kernel Orthonormality Regularization (KOR) [95] is utilized. KOR penalizes orthogonality violations of the convolutional kernel matrices leading to reduced feature redundancy, which enriches the information capacity of the latent space embedding and boosts the model generalizability. To implement this, the convolutional kernel tensor is reshaped to a 2D-kernel matrix maintaining the innermost dimension (number of output channels). Afterwards, the Gramian matrix is computed from the kernel matrix and the Frobenius norm is used for quantifying the residuals w.r.t. the identity matrix. The Frobenius norm corresponds to the Euclidean norm of the vector of eigenvalues of the matrix.

6.3 PIPTS calibration network

The PIPTS approach is implemented by a secondary, downstream CNN model, which utilizes the predicted logits tensor $\boldsymbol{\Xi} \in \mathbb{R}^{(v,\,604,\,960,\,19)}$ from the semantic segmentation head of the baseline model and the estimated Zernike coefficient vector $\vec{\omega} \in \mathbb{R}^{(v,\,3)}$ from the restoration head of the baseline model to predict an instance-wise temperature $\mathrm{T}_{\mathrm{cal}} \in \mathbb{R}_{+}^{(v,\,1)}$ for online calibration. Using the flattened logits tensor $\boldsymbol{\Xi}$ and the Zernike coefficient vector $\vec{\omega}$ to directly determine the temperature did not achieve satisfying results. A

superior performance was achieved by encoding the logits tensor Ξ with another CNN before concatenating it with the Zernike coefficient vector $\vec{\omega}$. From this experiment, it can be concluded that the spatial distribution of logits in the image plays an important role for the calibration quality, as the spatial distribution of objects of different semantic classes (e.g., sky, persons etc.) also exhibits spatial features (e.g. the sky is up). The best compromise between calibration quality and efficiency was established, when seven encoder blocks were employed for the logits tensor feature extraction.

For the PIPTS model, the ECE is directly utilized as a loss function. The ECE metric is not differentiable because it implicitly relies on a counting operation for the confidence binning. Consequently, the ECE can not be used as a loss function a priori. Hence, a mathematical trick is required to smooth the ECE metric such that it becomes continuous (differentiability class C^0). The trick consists of employing the continuous softmax function with a large exponential scaling factor ($\beta_s = 1000$) in places where discontinuous operations are used, e.g., replacing the argmax operation. This will result in a differentiable function but it is not guaranteed that the derivative is continuous as well. In order to establish continuous differentiability (C^1), the smoothed ECE function is modulated.

Loss modulation is applied in order to raise the discontinuity in the ECE gradient around zero. For that reason, the gradient is modulated by the sigmoidal function $f'(x;\eta) := \tanh^2(\eta x)$. As a result, the loss function is modulated by the function $f(x;\eta) := x - \eta^{-1}\tanh(\eta x)$, which imposes an inflection point around $x = 0$. The hyperparameter η has been determined experimentally by hypertuning to $\eta = 50$.

Finally, temperature regularization is applied in order to penalize predictions close to the temperature scaling pole at $T = 0$. The regularization term for the loss gradient is chosen as $g'(x;\kappa) := \tanh^2(\kappa x) - 1$. Considering the constraint that the regularization term for the loss function has to be positive-definite, the regularization term for the loss function is given by $g(x;\kappa) := -\kappa^{-1}(\tanh(\kappa x) - 1)$. The hyperparameter κ has been quantified empirically by hypertuning to $\kappa = 8$.

In a nutshell, the loss function for the PIPTS training is given by:

$$\begin{aligned} \mathcal{L}(\mathrm{ECE},\ \mathrm{T}) &:= f(\mathrm{ECE}) + g(\mathrm{T}) \\ \Rightarrow \mathcal{L}(\mathrm{ECE},\ \mathrm{T}) &= \mathrm{ECE} - 0.02\tanh(50\ \mathrm{ECE}) - 0.125\left(\tanh(8\ \mathrm{T}) - 1\right)\ , \end{aligned} \tag{6.2}$$

and the gradient w.r.t. the weight θ_k is modulated by:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \theta_k} &\overset{(6.2)}{=} \left(\frac{\partial f}{\partial \mathrm{ECE}} \cdot \frac{\partial \mathrm{ECE}}{\partial \mathrm{T}} + \frac{\partial g}{\partial \mathrm{T}}\right) \frac{\partial \mathrm{T}}{\partial \theta_k} \\ \Rightarrow \frac{\partial \mathcal{L}}{\partial \theta_k} &= \left(\tanh^2(50\ \mathrm{ECE}) \cdot \frac{\partial \mathrm{ECE}}{\partial \mathrm{T}} + \left(\tanh^2(8\ \mathrm{T}) - 1\right)\right) \frac{\partial \mathrm{T}}{\partial \theta_k}\ . \end{aligned} \tag{6.3}$$

Figure 6.2 visualizes the development process of constructing the loss modulation function f and the temperature regularizer g presented in Equation (6.2). In addition to the

ECE, the AUREC calibration measure introduced in Section 2.4.4 was studied. It is evident in Figure 6.2a that the ECE and the AUREC are not coherently minimized, indicating a decoupling of the calibration measures, which will be the subject of Chapter 8.

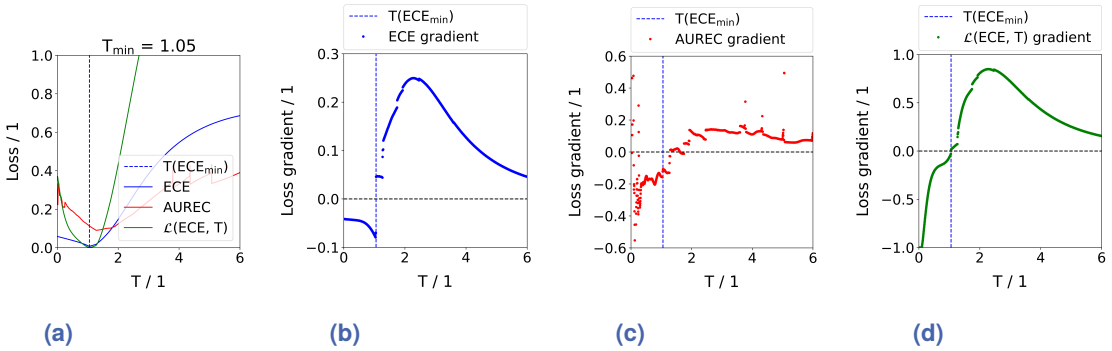


Figure 6.2: Loss function study for the PIPTS calibration network. The loss is indicated for a random instance as a function of the calibration temperature in (a) with $\beta_s = 1000$ and $N_c = 10$. The smoothed ECE measure is plotted as a blue line and the corresponding gradient is visualized in (b). The discontinuity at the optimal temperature T_{min} indicates the need for an additional modulation function. The gradient of the total loss function, containing the modulation function and the temperature regularization term, is visualized in (d). It can be concluded that the total loss $\mathcal{L}$ is sufficiently continuous differentiable (C^1) for backpropagation. Furthermore, the gradient of the AUREC is plotted in (c) as a function of the calibration temperature. The number of peaks indicates that the smoothing of the AUREC loss function by the softmax function was insufficient to ensure continuity. Hence, the AUREC loss function is inadequate for backpropagation and for neural network training respectively.

Furthermore, the AUREC could not be employed as a loss function for the PIPTS training as the smoothing by the softmax function was insufficient for establishing continuity.

As an aside, the hypertuning of the hyperparameters η and κ is supposed to remain valid for different datasets and neural network architectures, as long as the normalization of the input and output is maintained and the characteristics of the ECE metric are not modified. In particular, the number of bins will heavily affect the value of the hyperparameters because the number of bins determines the number and magnitude of the discontinuities in the ECE metric.

6.4 PIPTS calibration quality

Before the results regarding the calibration robustness against dataset shifts driven by optical aberrations of the windshield are discussed in detail for PIPTS, the baseline performance of the multi-task network in terms of the semantic segmentation for the A2D2 dataset is evaluated by the mIoU. The segmentation of a random instance of the test dataset is visualized in Figure 5.2. Across the entire test dataset, the average mIoU evaluates to $\text{mIoU}_{\text{test}} = (67.3 \pm 0.7)\%$.

Regarding the calibration performance, Figure 6.3 depicts the reliability diagram for the baseline model calibrated with TS. As the multi-task network is inherently well calibrated for the diffraction-limited case by utilizing class-balancing and loss-focusing during training, the optimal temperature according to TS is given by $T_{cal} = 1.0$, which corresponds to an identity mapping of the confidences. In addition, Figure 6.4 visualizes the reliability diagrams for the multi-task network after applying PTS and PIPTS, respectively. It is evident that the instance-wise calibrators significantly outperform TS by over 1%.

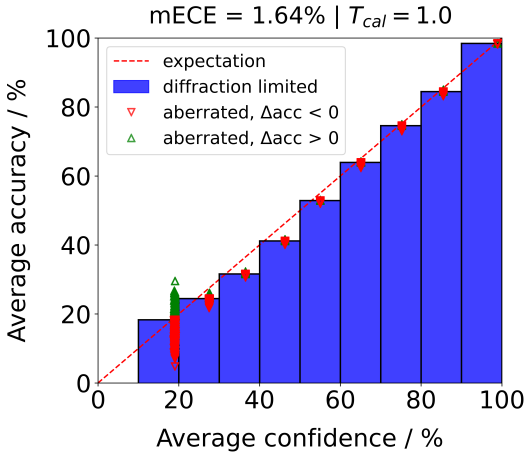


Figure 6.3: Reliability diagram for the multi-task network after applying TS. If different optical perturbation scenarios are considered, the individual bins are affected as indicated by the red (average bin accuracy is decreased) and green (average bin accuracy is increased) triangles.

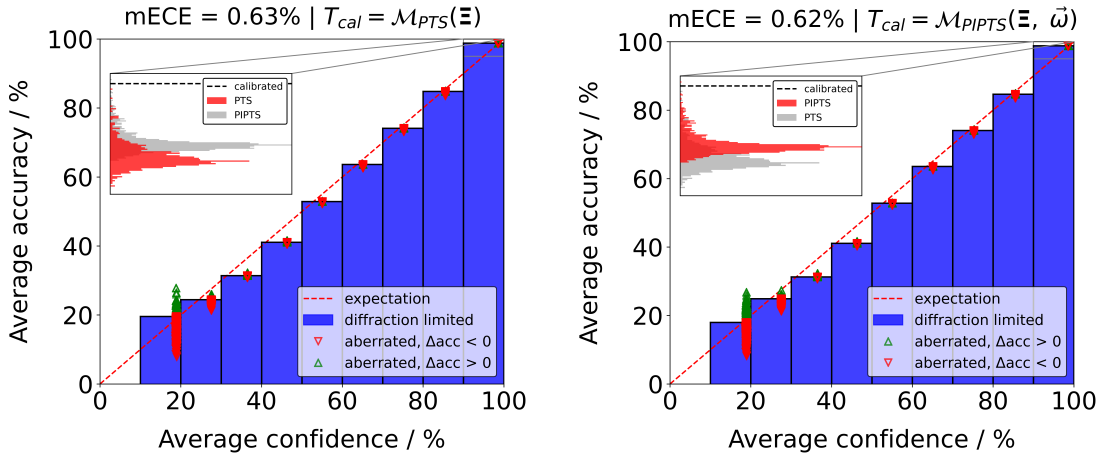


Figure 6.4: The reliability diagrams for the multi-task network are presented if calibrated by: (left) PTS or (right) PIPTS. If TS is directly contrasted against PTS, then the performance gain in terms of the mECE amounts to over 1%, which is tremendous. The calibration performance can be further boosted if a physical inductive bias is included into the PTS architecture. This benefit becomes significant if dataset shifts due to optical aberrations in the perception pipeline are induced. The impact of different perturbation scenarios is visualized by red (average bin accuracy is decreased) and green (average bin accuracy is increased) triangles. For the bin with the highest cardinality w.r.t. the subset of predictions, the distribution of the calibration error in terms of the difference between the average confidence and average accuracy is highlighted by an auxiliary plot. If PTS is contrasted against PIPTS, it is noticeable that the calibration bias is significantly reduced by adding additional information about the magnitude of the dataset shift. Furthermore, a slight reduction of the variance of the calibration error distribution can be observed.

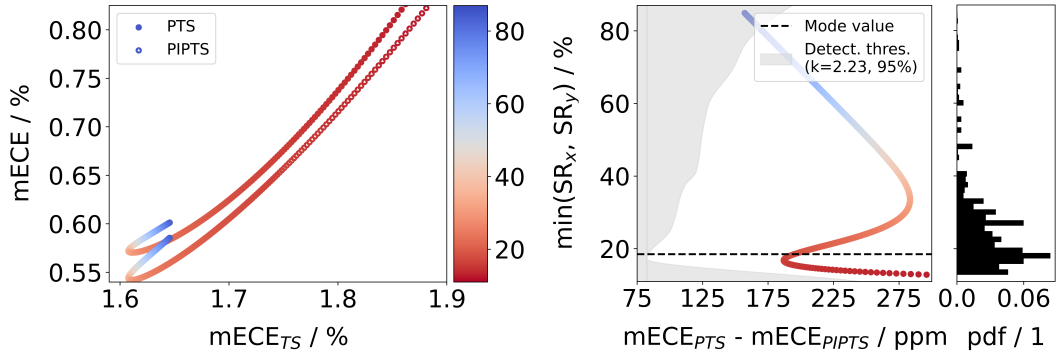
This highlights the gain in expressive power provided by the superior information capacity of the post-hoc CNN-based calibrators [167]. The physical inductive bias of PIPTS results in an additional mECE boost of roughly 100 ppm for the diffraction-limited case. If aberrations are considered, then the knowledge about the dataset shift magnitude successfully counteracts the calibration bias. In addition to accuracy improvements, PIPTS also slightly improves the calibration precision, which is notable by narrower bin-wise calibration error distributions.

A valid question might be raised about how significant this performance boost is. In order to tackle this question, the Deep Ensemble approach [107] presented in Section 2.4.4 was utilized. An ensemble of 11 PIPTS models has been trained with the same hyperparameters (congruent loss function landscape in the parameter space) but random and hence different weights initialization. Each ensemble member is used to calibrate the baseline model and the robustness analysis presented in Figure 5.5 is repeated for the well-correlating optical metrics, the Strehl ratio and the OIG. Consequently, the violation of the calibration condition (see Equation (2.97)) under optical aberrations can be quantified by extracting the regression function parameters β_q. Subsequently, the mean performance boost across all ensemble members and the corresponding standard deviation for the mean are evaluated and visualized in Figure 6.5a and 6.5b. Utilizing the standard deviation of the mean calibration performance in terms of the mECE as a predictive uncertainty estimator for the PIPTS calibrator implicitly neglects the aleatoric uncertainty component. If it is intended to include this component into the uncertainty estimation regarding the mECE, second order distributions (see Section 2.4.4) have to be considered. This could be done by extending the output space of the PIPTS calibration network such that the conditional uncertainty of the temperature is predicted alongside the instance-wise calibration temperature. The conditional uncertainty can then be propagated to retrieve an uncertainty estimate for the mECE. As we restrict our study to first order distributions, the PIPTS calibrator exclusively predicts the temperature. The significance level of the PIPTS performance gain has been set to 95% utilizing the Student's t-distribution and a corresponding extension factor of $k_{10} = 2.23$ for an ensemble with $\nu_{\text{eff}} = 10$ degrees of freedom [68, 140]. Here, the Student's t-distribution has to be considered as the standard deviation is estimated based on a very limited number of degrees of freedom [68, 140].

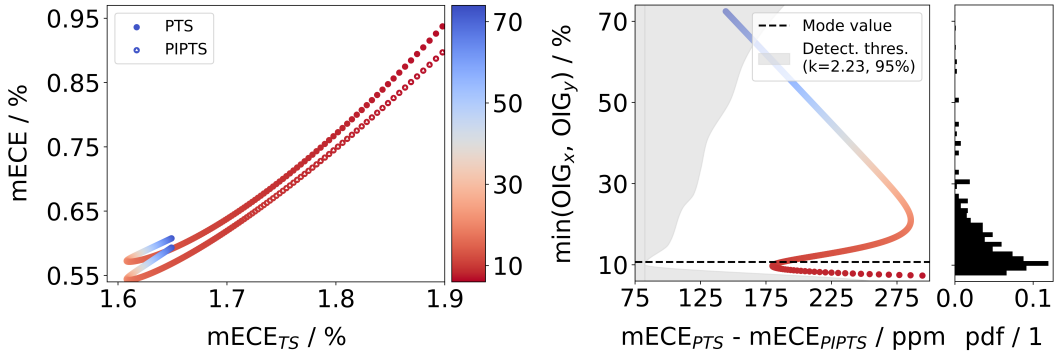
Figure 6.5a depicts the mECE curves for the PTS and PIPTS calibrator against the TS calibration results for the Strehl ratio. Those curves do not represent functional relationships because the calibration performance for all three post-hoc techniques is maximized for the mean-aberration regime. The PIPTS performance boost is not significant for the diffraction-limited case if a confidence level of 95% is considered in the view of the Student's t-distribution. The performance boost increases almost linearly with the aberration magnitude until the mean-aberration regime is approached. As the optical degradations approach the level of the mean optical quality of the training dataset, given by the dataset centroid, the benefit of employing PIPTS decreases. Hence, the supplementary

information about the Zernike coefficients is not as beneficial as for long-tail samples with fewer aberrations in terms of the Strehl ratio. This might be the case if a neural network trained on augmented data is used to infer information from synthetic data, which typically does not comprise optical aberrations of the windshield. On the other hand, in the large-aberration regime, the performance boost of PIPTS increases again after passing the mode value of the augmented dataset distribution. This scenario might occur if aging effects of the windshield are considered.

A similar outcome is observed if the OIG is considered as the optical target metric, as illustrated in Figure 6.5b. As a result, the PIPTS performance boost is significant on a 95% confidence level for the mean- and large-aberration regime but not for the low-aberration regime close to the diffraction limit, as the physical prior does not add significant information within this domain.



(a) The net performance boost of the PIPTS calibrator is indicated as a function of the Strehl ratio.



(b) The net performance boost of the PIPTS calibrator is indicated as a function of the OIG.

Figure 6.5: On the left-hand side, the mECE curves for the PTS and PIPTS calibrator are plotted versus the TS calibration performance. The colorbar indicates the aberration magnitude in terms of the (a) Strehl ratio and (b) OIG. The curves are obtained by averaging over 11 post-hoc models in a Deep Ensemble fashion. The graph in the middle shows the performance boost of PIPTS in comparison to PTS over the aberration magnitude as well as the corresponding detection threshold on a 95% confidence level in gray. On the right-hand side, the histogram of the augmented dataset distribution in terms of the (a) Strehl ratio and (b) OIG is visualized. It is evident that the mode value of the (a) Strehl ratio distribution and (b) OIG distribution correlates with the local minimum in the performance boost curve. In summary, the performance boost induced by the physics prior in PIPTS is significant for the mean- and large-aberration regime if PSF-based optical measures – in particular the Strehl ratio or the OIG – are considered.

The significant calibration performance boost of PIPTS in the order of 250 ppm needs to be considered in the light of the number of instances that occur in the lifetime of an autonomous driving car fleet. As a thought experiment, if a fleet of 10 million cars is taken into account, which corresponds to the annual production volume of large car manufacturers, and a mileage of 200,000 km is considered per car, then a calibration error reduction of 250 ppm corresponds to an increase in the safety margin of 500,000,000 km. This perspective underpins the benefits of PIPTS for autonomous driving.

The performance boost of PIPTS – generated by the physical inductive bias – manifests itself as a slight distribution shift in the predicted temperature over the entire test dataset. As a toy-example, Figure 6.6 depicts the histograms of the predicted temperature deviation for the aberration scenario: $\omega_3 = -0.2\ \lambda$, $\omega_4 = 1.0\ \lambda$, and $\omega_5 = -1.0\ \lambda$. The temperature deviation is given as the difference between the instance-wise temperature predicted by the corresponding calibrator (PTS or PIPTS) and the instance-wise optimal temperature, which is obtained by minimizing the mECE as a function of the temperature for each instance separately. The distributions show a non-zero bias, which is subject to the dataset shift induced by the optical aberrations. The bias is significantly reduced if the PIPTS calibrator is employed instead of the PTS calibrator. The parameters of the underlying Gaussian distribution (bias μ and standard deviation σ) were determined by a negative log-likelihood fit and the corresponding parameter uncertainty is estimated by the inverse of the local curvature of the negative log-likelihood curve according to the Cramer-Rao-Frechet bound [35, 146, 50].

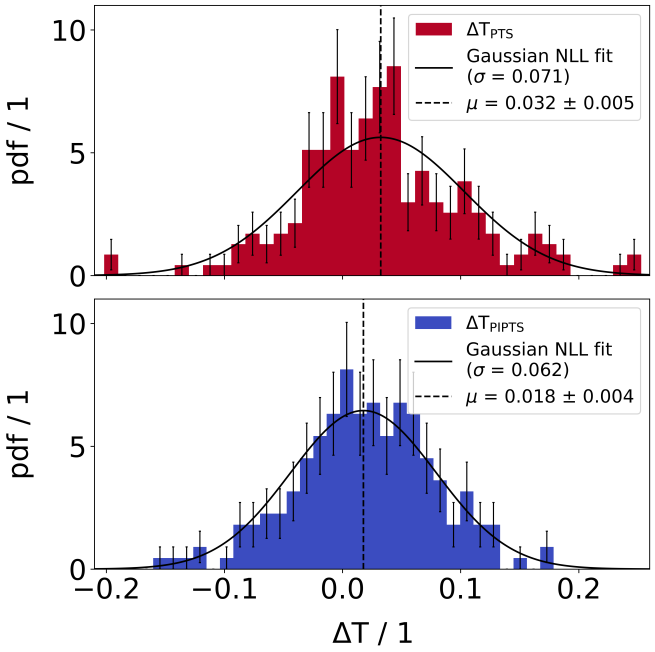


Figure 6.6: The temperature deviation ΔT between the predicted temperature of the (top) PTS, (bottom) PIPTS calibrator and the optimal temperature is plotted as a histogram. Both distributions indicate a significant bias, which reflects the influence of the dataset shift on the post-hoc calibrators. The bias μ is significantly reduced if the PIPTS calibrator is employed instead of the PTS model.

6.5 Interim conclusion

In this Chapter, it was experimentally demonstrated that enriching the PTS calibrator by a physical inductive bias leads to superior robustness against dataset shifts. By employing PIPTS, the trustworthiness of the baseline model's confidences can be enhanced by mitigating the dataset shift related bias induced by optical aberrations.

Furthermore, the implicit online monitoring of the effective optical system aberrations, in terms of the Zernike coefficient vector, allows for a system health analysis. For that, the relationship between PSF-based optical measures and neural network KPIs, presented in Section 5.3, can be used to derive safety margins in terms of the Zernike coefficients. If the predicted Zernike coefficients violate the safety threshold regarding a minimum optical quality, then the perception chain is compromised and car maintenance might be necessary.

Those system-level optical safety margins can be mapped to part-specific requirements by inverting the deterministic Fourier optical model. Part-specific requirements are essential for the verification process in the automotive industry, which typically relies on the validity of the Vee-model [175] that implicitly assumes linear system theory. From Section 4.3.2 it is evident that this mapping needs to be performed in the complex-valued domain of the electromagnetic field instead of the intensity-based world.

As a requirement for training the PIPTS calibrator, the Zernike coefficients for the combined optical system has to be known as well as the part-level wavefront aberrations. The system-level Zernike coefficients can be retrieved from the Fourier optical model, which relies on the information about the part-level wavefront modulations. For the ADAS camera, this information can be retrieved by a Shack-Hartmann measurement, where the local gradients from Equation (4.9) need to be integrated in order to obtain the optical path difference map. This leaves the challenge of determining the wavefront modulation induced by the windshield conditioned on the reference wavefront given by the non-diffraction-limited ADAS camera. This can not be realized by a standard Shack-Hartmann sensor as the measurement procedure implicitly assumes a plane reference wavefront. The next Chapter will propose a BOS imaging-based measurement procedure to reconstruct the wavefront modulation induced by the windshield as seen by the ADAS camera.

In addition, the results of Figure 6.2a, on the dependency of the ECE and AUREC on the calibration temperature, indicate a decoupling of neural network calibration measures. This hints on a safety-critical issue, as it is implicitly assumed that the calibration quality measures itself are unbiased. Chapter 8 will address this puzzle in detail.

In a nutshell, PIPTS paves the way for establishing trustworthiness and robustness in AI-based autonomous driving functionalities by ensuring superior confidence calibration under optical aberrations in the perception chain and by providing a physical sound toolchain for deriving part-specific optical requirements.

7 Measurement of wavefront aberrations induced by the windshield

This chapter includes elements with editorial changes from

[191] Dominik Werner Wolf, Boris Thielbeer, Markus Ulrich, and Alexander Braun. "Wavefront aberration measurements based on the Background Oriented Schlieren method". In: *Measurement: Sensors* (2024), p. 101509. ISSN: 2665-9174. DOI: https://doi.org/10.1016/j.measen.2024.101509. URL: https://www.sciencedirect.com/science/article/pii/S2665917424004859

marked with a green line.

The PIPTS calibrator requires the knowledge about the wavefront modulation induced by the windshield as a ground truth information for training. In this Chapter, a novel measurement procedure based on BOS imaging and image cross-correlation is proposed to address this challenge. In detail, the wavefront aberrations of the windshield are reconstructed by employing high-resolution refractive power measurements. In order to do so, we will exploit the fact that the local refractive power characterizes the local curvature of the optical path difference map, as formulated in Equation (4.2).

7.1 Related work

The impact of wavefront aberrations is of major concern for many highly sophisticated technical systems [125, 90, 57, 88]. As an example, airborne communication systems [90] and airborne laser-based weapon systems [57] are considerably affected by wavefront aberrations from turbulences. A similar issue can be observed in autonomous driving as addressed by Jaiswal et al. [88]. Those complications are phenomenologically identical to our challenge of handling the optical aberrations induced by the windshield, as the underlying physics is equivalent. In contrast to turbulence-driven wavefront aberrations, the windshield's optical aberrations are generally assumed to be time independent, which only holds true as long as ageing effects of the windshield are neglected.

In Section 6.5 we summarized that the specification of optical part requirements has to be made in the complex domain of the electromagnetic field to account for phase modulations. We arrived at this conclusion, as the governing Vee-model [175] in the automotive industry requires the decomposition of optical system requirements into part tolerances. If we apply this top-down verification approach to the perception chain, it is crucial to pay attention to the codomain shift. The system requirements might be deduced from a regression function that sufficiently captures the dependency of the performance of an underlying AI algorithm on the input image quality, as demonstrated by Equation (5.1). This system-level performance measure will be intensity-based because a CMOS sensor captures only the average Poynting vector [76] in the direction normal to the sensor surface. Hence, the intensity – i.e. the squared electric field amplitude – determines the number of electron-hole pairs created in the semiconductor. Nevertheless, part-level requirements have to incorporate the information about the modulation of the complex-valued electromagnetic wave by spatial variations of the refractive index within the propagation path. Otherwise, interference effects are untraceable, which would prevent the definition of unique part-level requirements [193] as the superposition of two aberration fields can lead to constructive or destructive interference [76].

In order to measure the windshield's wavefront aberrations, we will propose to utilize the BOS imaging methodology presented in Chapter 3 based on image cross-correlation.

The equivalence of the physical principles underlying wavefront aberration measurements based on Shack-Hartmann sensors and image distortion measurements based on BOS imaging was discovered over a decade ago [11, 12, 166]. In general, both approaches perform a relative measurement between the perturbed wavefront and the reference wavefront and target on observing the spatial distribution of local deflection angles, which relates to the local wavefront gradients w.r.t. the transversal axes, as introduced in Equation (4.9). The wavefront aberration map can be reconstructed by numerical integration and the local refractive power is obtained by differentiating the measured wavefront gradients [192, 193]. As a result, the complex-valued phase modulation induced by the windshield can be measured independently of the camera objective lens.

The advantage of BOS imaging over a Shack-Hartmann sensor manifests itself in three keypoints. On the one hand, the aperture size and the resolution of a Shack-Hartmann sensor is limited by the spanned sensitive area of the lenslet array and the number of CMOS pixels per microlens, respectively. Secondly, the reference wavefront in BOS imaging is implicitly given by the objective lens. Therefore, the expected nominal wavefront is characterized by a sphere with a radius equal to the nominal focal length f of the objective lens, which is naturally subject to deviations due to manufacturing tolerances. Consequently, the BOS imaging setup assesses an optical path more representative of the final application in comparison to a collimated measurement configuration, which current regulations demand for, in particular the UN/ECE-R43 [148]. Finally, Shack-

Hartmann sensors are more expensive than BOS setups if a significant area under test is considered [11].

Bichal et al. [11] laid the foundation of BOS imaging-based wavefront measurements. They experimentally verified the measurement principle by reconstructing the local optical path difference – induced by a plano-convex lens with 1 m focal length – across the lens aperture [11]. By comparing the experimental results with theoretical expectations they conclude that the absolute error for their specific setup is less than 5% [11]. Nevertheless, they particularly highlight that this measurement uncertainty estimate strongly depends on the setup alignment and a more detailed analysis is needed for arbitrary objects under test [11].

This leaves us with a bunch of open research questions: Is the accuracy of wavefront measurements based on BOS imaging sufficient to study the aberration pattern of windshields and how to establish a sanity check that easily assesses the validity of the measurement setup? How can detailed knowledge of the wavefront modulation induced by the windshield be utilized to improve the robustness of AI-based perception systems? Is BOS imaging the most suitable measurement technique for this purpose?

We will address these research questions in this chapter. First of all, the accuracy of BOS imaging is evaluated, based on the global wavefront aberration map, characterized by the Zernike coefficients defined in Section 2.1.1. In detail, we will investigate if the Zernike coefficient vector obeys the expected rotation transformation property for a counterclockwise rotation of 90°. To do so, we will start by deriving the governing equations for reconstructing the Zernike coefficients of the global wavefront aberration map based on high-resolution refractive power measurements. Subsequently, we will formulate the expected transformation properties of the Zernike polynomials by analyzing the periodicity. In a nutshell, we contribute to the industrialization of BOS imaging-based wavefront measurements by extending the accuracy assessment, targeting on the fulfillment of the essential pillars of metrology regarding a valid measurement, as described in Section 2.3.

Secondly, we will highlight the necessity of BOS imaging-based wavefront measurements for enhancing the robustness of AI-based perception systems in the automotive industry. As a starting point, we need to acknowledge that the ability to measure the complex-valued phase modulation induced by the windshield relative to the reference wavefront determined by the ADAS camera is a prerequisite for the PIPTS calibration strategy presented in Chapter 6. We will propose to further utilize the measurement information about the wavefront modulation induced by the windshield to enrich the training dataset through physics-informed data augmentation [156, 194], thereby enhancing data heterogeneity without the need to gather additional data on the road.

Thirdly, the independence of the windshield's wavefront aberration measurement from the objective lens will enable us to account for the codomain shift in part-level testing

and pave the way for utilizing innovative glazing concepts through a holistic deep optics optimization of the perception pipeline. This poses the unique selling point of BOS imaging-based wavefront measurements in contrast to conventional wavefront measurement techniques [141, 22].

7.2 BOS imaging for wavefront aberration measurements

In contrast to previous work, our setup utilizes high-resolution measurements of the local refractive power to determine the Zernike coefficients. This is possible because the unit vectors within the function space – spanned by the second-order derivatives of the Zernike polynomials – are orthogonal, which implies that the reconstructed coefficients are unique. We introduce this change in the information processing pipeline as the local refractive power is the most established measure in the automotive industry. Consequently, the comparability to existing standard measurement devices, which comply with the UN/ECE-R43 [148], can more easily be evaluated. Furthermore, the refractive power can be directly traced back to international standards by utilizing a calibrated reference lens, as described in Chapter 3.

The Zernike coefficients ω_n for parameterizing the wavefront aberration map can be retrieved in theory from the measurement of the local refractive power distribution by exploiting Equation (4.2). Hence, the refractive power vector $\vec{D}$, which embeds the spatial distribution of local refractive power measurements into a flattened vector:

$$\vec{D} := \left[\left. D_x \right|_{\vec{\eta}_1} \quad \dots \quad \left. D_x \right|_{\vec{\eta}_m} , \quad \left. D_y \right|_{\vec{\eta}_1} \quad \dots \quad \left. D_y \right|_{\vec{\eta}_m} \right]^T , \tag{7.1}$$

can be related to the Zernike coefficient vector $\vec{\omega}$ as:

$$\vec{D} = \frac{1}{\rho_a^2} \cdot \boldsymbol{\mathcal{M}} \cdot \vec{\omega} , \tag{7.2}$$

with the matrix $\boldsymbol{\mathcal{M}}$ quantifying the local curvature of the corresponding Zernike polynomial Z_n at location $\vec{\eta}_m$:

$$\boldsymbol{\mathcal{M}} := \begin{bmatrix} \left. \frac{\partial^2 Z_4}{\partial^2 \tilde{x}} \right|_{\vec{\eta}_1} & \cdots & \left. \frac{\partial^2 Z_4}{\partial^2 \tilde{x}} \right|_{\vec{\eta}_m} , & \left. \frac{\partial^2 Z_4}{\partial^2 \tilde{y}} \right|_{\vec{\eta}_1} & \cdots & \left. \frac{\partial^2 Z_4}{\partial^2 \tilde{y}} \right|_{\vec{\eta}_m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \left. \frac{\partial^2 Z_n}{\partial^2 \tilde{x}} \right|_{\vec{\eta}_1} & \cdots & \left. \frac{\partial^2 Z_n}{\partial^2 \tilde{x}} \right|_{\vec{\eta}_m} , & \left. \frac{\partial^2 Z_n}{\partial^2 \tilde{y}} \right|_{\vec{\eta}_1} & \cdots & \left. \frac{\partial^2 Z_n}{\partial^2 \tilde{y}} \right|_{\vec{\eta}_m} \end{bmatrix}^T . \tag{7.3}$$

The Zernike decomposition coefficients ω_n are uniquely determined if $|\boldsymbol{\mathcal{M}}^T\boldsymbol{\mathcal{M}}| \neq 0$. In other words, the Gramian matrix $\boldsymbol{\mathcal{M}}^T\boldsymbol{\mathcal{M}}$ has to be invertible, wherefore $\boldsymbol{\mathcal{M}}^T\boldsymbol{\mathcal{M}}$ needs to have full rank. If this condition is fulfilled, then the Zernike coefficient vector $\vec{\omega}$ is obtained by:

$$\vec{\omega} \overset{(7.2)}{=} \rho_a^2 \cdot \left[\boldsymbol{\mathcal{M}}^T\boldsymbol{\mathcal{M}}\right]^{-1} \cdot \boldsymbol{\mathcal{M}}^T \cdot \vec{D} \,. \tag{7.4}$$

It is important to note here that the second order derivatives of the zeroth to third order Zernike polynomials vanish, and hence the reconstruction of the Zernike coefficients is limited to those with an ANSI index greater than three (see Table 2.1). Generally, the Zernike coefficients of zeroth and first radial order are irrelevant because they do not impact the imaging quality, as elaborated in detail in Section 4.3, wherefore they are typically discarded as an offset when performing standard wavefront measurements.

The missing relevant Zernike coefficient Z_3 for imaging applications, oblique astigmatism, can be retrieved if two consecutive measurements are taken with an angular displacement of 45°. The expected rotation invariance of all azimuth-angle independent Zernike polynomials can be utilized to verify the validity of the measurement pair and to check if the alignment accuracy is sufficient. In addition, the consistency of the measurement technique can be evaluated if an angular displacement of 90° is considered and the rotation transformation properties of the Zernike coefficient vector are exploited. For this parameterization, the expected Zernike coefficient vector after a counterclockwise rotation by 90° is given by:

$$\vec{\omega}_{\alpha+90^\circ} = \boldsymbol{\mathcal{R}}_{+90^\circ} \cdot \vec{\omega}_\alpha \,, \tag{7.5}$$

if the initial measurement has been performed at an arbitrary azimuth angle α. The corresponding rotation transformation matrix $\boldsymbol{\mathcal{R}}_{+90^\circ}$ is retrieved by analyzing the sinusoidal frequency for every Zernike polynomial and by exploiting the respective periodicity. Consequently, the counterclockwise rotation matrix $\boldsymbol{\mathcal{R}}_{+90^\circ}$ is determined by:

$$\boldsymbol{\mathcal{R}}_{+90^\circ} = \begin{bmatrix}
\mathbf{1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \mathbf{-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \mathbf{-1} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \mathbf{1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \mathbf{-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \mathbf{1} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{-1} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{1} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{-1} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{1}
\end{bmatrix} . \tag{7.6}$$

As a result, the experimental setup can be benchmarked by comparing the measured Zernike coefficient vector $\vec{\omega}_{\alpha+90^\circ}$ with the expected Zernike coefficients after rotating $\vec{\omega}_\alpha$

by the matrix $\boldsymbol{\mathcal{R}}_{+90^\circ}$. In the case of a rotation-symmetric optical setup, the vector pairs are expected to be congruent. Since a camera with a non-square CMOS imager is utilized for the BOS imaging setup, the measurements have to be restricted a posteriori to the minimum field of view across all axes.

7.3 Resolution requirements for BOS imaging

In order to enable the reconstruction of local wavefront aberrations based on local deflection maps, BOS imaging must be performed with a certain minimal resolution. This resolution requirement will be derived subsequently. As a starting point, the limitation in the imaging resolution imposed by the pixel pitch τ of the CMOS sensor of the ADAS camera needs to be considered. According to the Nyquist theorem, the discretization in the sampling schema limits the recoverable information content in the frequency domain to: $k_x,\ k_y \in \mathbb{R} \mid (-(2\tau)^{-1},\ (2\tau)^{-1})$. The boundaries of this domain correspond to the Nyquist frequency. The cardinalities of the sets $\{k_x\}$ and $\{k_y\}$ of measurement samples are determined by the number of pixels N_i of the CMOS sensor along the corresponding axis. Hence, the frequency domain is partitioned into segments with an incrementation of $\Delta k_i = (\tau N_i)^{-1}$. As a result, the spatial sampling rate of the PSF in the observer plane is given by the pixel pitch $\Delta x_i^{(o)} = \tau$. Consequently, the required spatial resolution in the aperture plane for the generalized aperture function $\tilde{P}(\vec{x}_a)$, introduced by Equation (2.60), is determined by $\Delta \tilde{x}_i^{(a)} = (\tau N_i)^{-1}$. The aperture coordinate $\tilde{x}_i^{(a)}$ still contains the normalization factor introduced in Equation (2.58). In order to obtain the spatial resolution in physical length units, the variable substitution performed in Equation (2.58) needs to be inverted. This ultimately leads to a minimal sampling rate requirement in the aperture plane of:

$$\Delta x_i^{(a)} \overset{!}{\leq} \frac{\lambda z}{\tau N_i} \,. \tag{7.7}$$

According to the thin lens Equation (2.39), the longitudinal distance z between the aperture surface and imaging plane can be approximated by the focal length of the system if and only if the longitudinal distance γ between the object plane and the aperture surface is by magnitudes larger than z. This condition especially applies to telephoto cameras, where the target objects typically occur in a distance of ≥ 60 m. Hence, Expression (7.7) can be rewritten as:

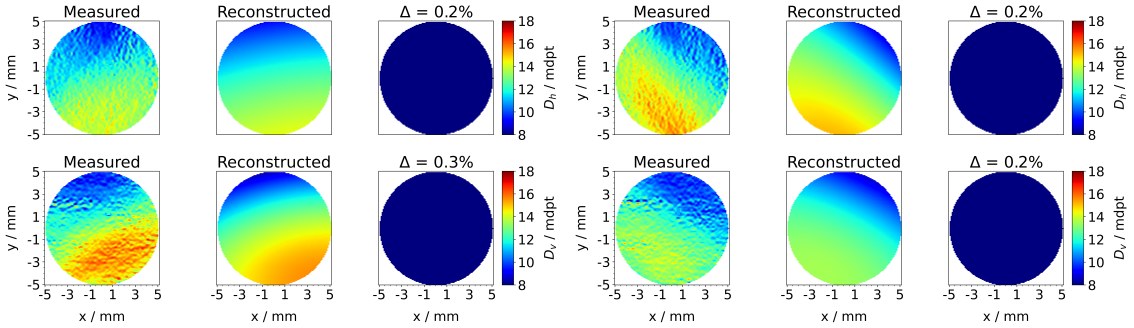
$$\Delta x_i^{(a)} \overset{!}{\leq} \frac{\lambda f}{\tau N_i} \,. \tag{7.8}$$

If the CARIAD dataset introduced in Section 2.4.2 is considered, then the required spatial resolution of BOS imaging in the aperture plane is given by $\Delta x_y^{(a)} \overset{!}{\leq} 1.3\ \mu$m for the vertical axis and it evaluates to $\Delta x_x^{(a)} \overset{!}{\leq} 0.65\ \mu$m for the horizontal direction. This resolution requirement is difficult to meet with a BOS imaging setup. Nevertheless, if the

global wavefront modulation is the main matter of concern rather than local wavefront aberrations, then BOS imaging might be capable of reconstructing an effective wavefront. This equals a local averaging of wavefront snippets. In the next Section, a proof of concept study will be performed to demonstrate that global wavefront modulations of the windshield can indeed be reconstructed by high-resolution BOS imaging if the hyperparameters of the BOS setup, described in Section 3.2, are tuned appropriately.

7.4 Sanity check for the proposed BOS measurement setup

The following study examines a flat laminated safety glass sample. Laminated safety glass consists of two flat glass layers, which encapsulate a tough-elastic, tear-proof foil made of polyvinyl butyral (PVB), as elaborated in detail in Section 2.2.1. Figure 7.1a presents the local refractive power measurement results for the safety glass sample across the aperture domain using BOS imaging, considering an initial alignment angle α. The set of Zernike basis polynomials has been limited to the 14^{th} order for the reconstruction of the wavefront aberration map. If a counterclockwise rotation of the glass sample by $90°$ is performed, then the measurements presented in Figure 7.1b are obtained.



(a) Measurement i_1 for an initial alignment angle α. **(b)** Measurement i_2 after rotation to $(\alpha + 90°)$.

Figure 7.1: Results of the local refractive power measurements of a flat laminated safety glass sample (a) for an initial alignment angle α and (b) after a counterclockwise rotation by $90°$ to $(\alpha + 90°)$. The upper row shows the results for the horizontal transversal axis and the lower row indicates the outcomes for the vertical transversal axis. Furthermore, the left column presents the raw measurement results and the middle column displays the reconstructed aberration field utilizing Equation (7.4). Finally, the right column visualizes the mismatch between the measurement and the global reconstruction over the aperture domain. The discrepancy is quantified in terms of the non-explainable variance, given by the ratio of the mean squared error (MSE) over the variance of the measurement set.

With the knowledge of both measurements, the validity of the expected rotation properties according to Equation (7.5) can be investigated. The initial measurement is directly compared to the transformed values from the second, rotated measurement in Figure 7.2. It can be concluded that up to the ninth order, the data points align in magnitude but

they are not congruent. The discrepancy between the results quantifies the impact of the symmetry break and the lack of alignment accuracy. The CMOS sensor of the BOS imaging camera is non-square, wherefore the field of view is not rotationally symmetric. The post-hoc procedure of restricting the area of interest within the recorded images to the minimum field of view across all axes is subject to numerical errors, which can lead to a weak symmetry violation. Additionally, the angular alignment underlies mechanical limitations regarding the alignment accuracy of 0.1°.

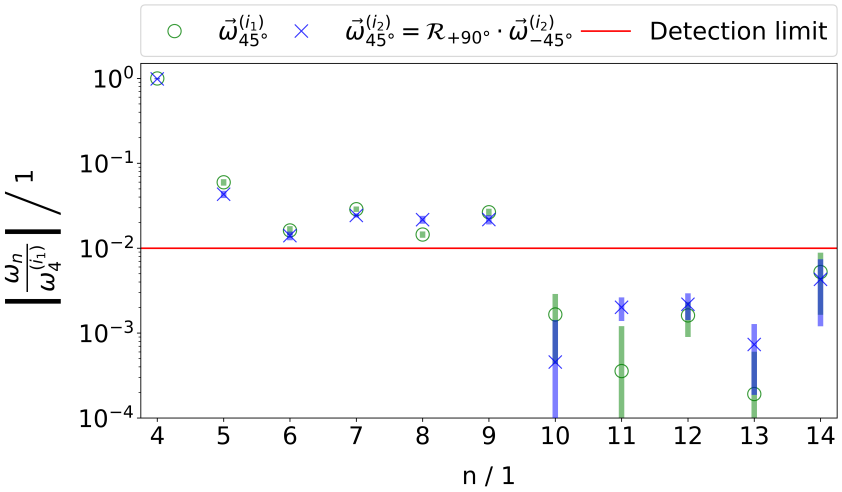


Figure 7.2: Transformation study for the reconstructed Zernike coefficients up to the 14th order. The uncertainty bars indicate the 95% confidence intervals. $\vec{\omega}_{45°}^{(i_1)}$ denotes the Zernike coefficient vector measured under an azimuthal alignment angle of $\alpha = 45°$. In addition, $\vec{\omega}_{45°}^{(i_2)}$ indicates the expected Zernike coefficient vector after transforming a secondary Zernike coefficient vector $\vec{\omega}_{-45°}^{(i_2)}$ obtained under an azimuthal alignment angle of $\alpha = -45°$. The transformation is mathematically described by the rotation matrix $\mathcal{R}_{+90°}$ presented in Equation (7.6). The detection threshold is set to 1% of the leading expansion coefficient.

The uncertainty intervals in Figure 7.2 were obtained by a Monte-Carlo study. The input uncertainty distributions are assumed to follow a Gaussian pdf. In total, the influence of three alignment parameters is considered, namely the uncertainty in the x and y coordinate of the rotation point and the uncertainty in the rotation angle itself. The corresponding standard deviations are specified with: $\sigma_x = 0.1$ mm, $\sigma_y = 0.1$ mm and $\sigma_\alpha = 0.1°$. Furthermore, the effect of Gaussian noise on the number of generated free electron-hole pairs in the CMOS sensor is investigated. The standard deviation of the pixel-wise intensity distribution is estimated to be ±5%. The ensemble set of Monte-Carlo samples has been initialized with a cardinality of $N = 1000$. From the output pdfs for every Zernike coefficient the median value and the corresponding 95% confidence interval have been computed. Figure 7.2 takes up on those statistical properties and further visualizes a significance threshold, which is simply set to be at 1% of the maximum Zernike coefficient (here defocus ω_4). Everything below this threshold has been assigned to noise.

7.5 BOS imaging for end-of-line testing

The measurement procedure based on BOS imaging is supposed to be utilized as a 100% end-of-line testing system. It will be required to analyze the BOS imaging measurements over the field of view of the ADAS camera by considering subapertures. For each subaperture a Zernike coefficient vector can be reconstructed, which will lead to a 2D-map of Zernike coefficient vectors over field. Each Zernike coefficient vector of the windshield can be used to parameterize a Fourier-optical degradation model (compare to Section 2.1.2). In addition, the Zernike coefficient vector for the ADAS camera must be known from a Shack-Hartmann wavefront measurement, in order to realistically simulate the MTF of the optical system. For illustration purposes, the camera optics will be assumed for the moment as diffraction-limited, wherefore the information obtained by BOS imaging is sufficient to parametrize the system MTF. As a result, a discretized 2D-map of the field-angle-dependent 2D system MTF curves can be determined as illustrated in Figure 7.3.

At this point, it has to be emphasized that those system MTF curves are simulated under strong assumptions. Most importantly, the assumption of a diffraction-limited ADAS camera is out of touch with reality as the production tolerances are too severe to be neglected. Furthermore, field-curvature aberrations – arising from the planar nature of the CMOS sensor in comparison to the spherical imaging plane as discussed in Section 4.3.2 – are neglected and the active alignment procedure of merging camera objective lens and CMOS sensor is not taken into account.

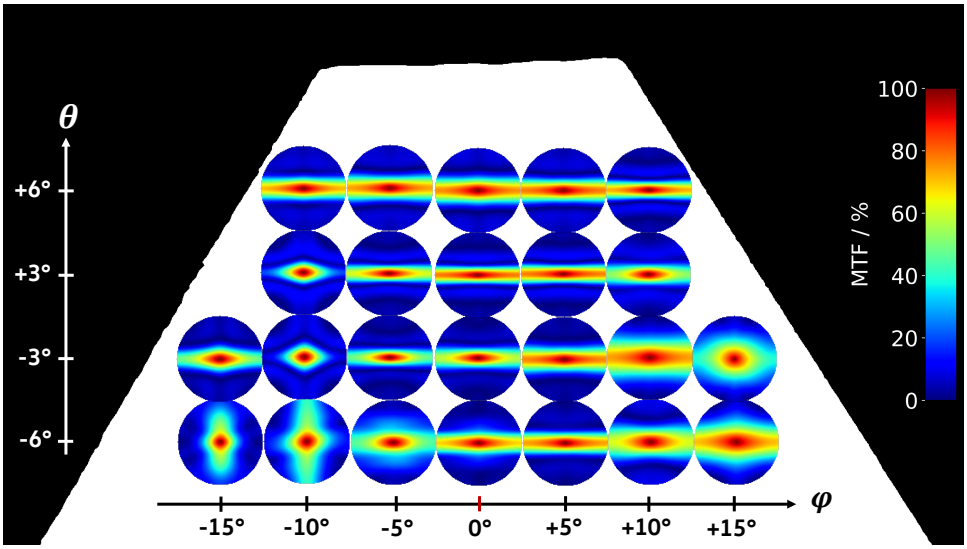


Figure 7.3: Distribution of the 2D system MTF over the field of view of the ADAS camera for a standard windshield. The field of view is parameterized by the azimuth angle φ and the polar angle θ. Furthermore, the domain of the system MTF is limited to the Nyquist frequency. For calculating the system MTF, a diffraction-limited camera with a focal length of $f = 16$ mm has been assumed. The system MTF of windshield and camera objective lens is illustrated with the underlying assumption of vanishing field curvature aberrations.

7.6 Interim conclusion

The capability of measuring the induced phase modulation by the windshield independently of the ADAS camera opens the solution space to handle aberration-based dataset shifts in the automotive industry. Generally, the robustness of CNN-based multi-task networks is increased by training a network with aberrated images taken behind different windshield configurations, as demonstrated in Chapter 5. By doing so, the data heterogeneity is amplified, which promotes the generalizability of any baseline neural network. With the presented BOS imaging measurement setup to reconstruct the Zernike coefficients of the windshield's wavefront modulation independently of the ADAS camera, a deterministic Fourier-optical model – as presented in Section 2.1.2 – can be parameterized to artificially degenerate images. This methodology increases the sample density in the data augmentation subspace without the need to experimentally take images behind different windshield configurations. This requires datasets, where the cameras are mounted on the car roof (e.g. A2D2 [56], KITTI [55]) but it drastically reduces the necessary resources for capturing a heterogeneous training dataset. In conclusion, by utilizing the proposed measurement strategy and a physics-realistic Fourier-optical model, the number of driven kilometers for capturing training data can be significantly reduced, which saves costs and reduces the CO_2 footprint in the development phase.

In addition, the knowledge about the discretized 2D-map of the field-angle-dependent 2D system MTF curves enables a more sophisticated optimization strategy w.r.t. the windshield design. As concluded in Chapter 5, such an optimization attempt in the windshield manufacturing process should target primarily on defocus aberrations but the degree of severity might vary across the field of view of the ADAS camera, which is resolvable by the proposed procedure.

Furthermore, the measurement information about the Zernike coefficient vector of the windshield is an essential requirement for the PIPTS network training, as elaborated in Chapter 6. In the current PIPTS architecture, the Zernike coefficients are assumed to be marginalized across the domain of the azimuth angle φ as well as the polar angle θ. This double integration can be discretized by utilizing the reconstructed field-angle-dependency of the Zernike coefficients, as illustrated in Figure 7.3 in the 2D MTF space.

Moreover, BOS imaging can be utilized for part-level testing of the windshield's Zernike coefficient vector. As elaborated in Section 4.4, the part testing strategy of the Vee-model [175] is subject to a codomain shift from the intensity-based system-level requirements on the perception chain to the part-level requirements in the complex-valued domain of the electromagnetic field ρ. This highlights the need for measurement systems such as the one presented in this chapter. On the contrary, if we do not further strengthen our devolvement effort to keep track of this codomain shift and continue to exclusively rely on intensity-based part measurements, then the verification strategy would need to utilize the maximum error estimation principle. Hence, the worst-case scenario is considered

to account for untraceable interference effects, which ultimately results in unrealistically high-quality requirements for the windshield. Those exaggerated optical quality requirements can not be met by standard production processes in place. Consequently, the rejection rate would dramatically increase wasting resources that deteriorate the goal of small CO_2 footprints in production [66].

On the other hand, if the codomain shift is adequately considered then the part-specific phase modulations can be incorporated into the AI architecture and a holistic optimization of the computer vision setup could be undertaken in a deep optics [36, 185] fashion. This opens the solution space for innovative glazing concepts. For example, thinner and hence lighter windshields are desirable because electric vehicles are range-limited due to the low energy density of batteries compared to fossil fuels [181]. In this context, a deep optics approach is inevitable since the windshield thickness negatively correlates with the aberration-proneness.

As a final remark, the presented methodology of employing high-resolution refractive power measurements for the reconstruction of the global wavefront modulation induced by the windshield is still subject to further development efforts until an end-of-line deployment is realizable. Notwithstanding this, the sanity check based on the flat laminated safety glass sample illustrated the feasibility of the proposed measurement technique in alignment with previous experimental results for a plano-convex lens [11].

8 Decoupling of neural network calibration measures

This chapter includes elements with editorial changes from

[189] Dominik Werner Wolf, Prasannavenkatesh Balaji, Alexander Braun, and Markus Ulrich. "Decoupling of neural network calibration measures". In: *German Conference on Pattern Recognition (GCPR)* (2024)

marked with a mint green line.

The training of the PIPTS calibrator in Section 6.3 revealed different optimal temperatures for the minimum of the ECE and the AUREC calibration quality measure. This decoherence hints on a possible decoupling between neural network calibration measures, which compromises the goal of ensuring trustworthiness in the model's predictions by calibrated uncertainty estimates obtained with PIPTS. From a metrological point of view, it is expected that the uncertainty measure coincides with the observed error rate in the network's predictions if the model is properly calibrated. As elaborated in Section 2.4.4, there exist point-wise predictive uncertainty estimators and entropy-based uncertainty measures for semantic segmentation. This distinction directly translates to point-wise predictive uncertainty calibration estimators (e.g., ECE, AUSE utilizing the variation ratio for sorting, etc.) and entropy-based calibration measures (e.g., UCE, AUSE utilizing the Shannon entropy for sorting, etc.), as introduced in Section 2.4.5. Ideally, the choice of point-wise or entropy-based calibration measure should not affect the safety statement w.r.t. proper calibration. Within the zoo of calibration techniques, temperature scaling [70] is one of the most prominent post-hoc techniques as introduced in Section 2.4.5. If all calibration measures were coupled, then one would expect that the temperature that minimizes the logits-based Negative Log-Likelihood (NLL) [70] simultaneously imposes a global minimum in the ECE curve over temperature. This should also hold true for alternative measures, presented in Section 2.4.5, attempting to estimate the calibration quality. To put into a more general perspective, it is expected that all calibration measures are coherently minimized and approach a minimum at the same temperature. Nevertheless, the observed decoupling between the ECE and AUREC gives rise to study the coupling mechanism among different calibration measures more thoroughly.

For this evaluation, the Uncertainty Calibration Error (UCE) defined by Equation (2.100), the Uncertainty Calibration Quality Score (UCQS) specified by Equation (2.101), and the

Area Under the Sparsification Error curve (AUSE_S) characterized in Section 2.4.5 – utilizing the Shannon entropy S for sorting – are considered as entropy-based calibration measures. On the contrary, the Expected Calibration Error (ECE) determined by Equation (2.99), the Confidence Calibration Quality Score (CCQS) defined by Equation (2.101), and the AUSE_V utilizing the variation ratio V for sorting are employed as point-wise calibration measures.

8.1 Related work

As experimentally demonstrated in Section 5.5 and addressed meticulously by the PIPTS approach in Chapter 6, modern neural networks typically exhibit a bias in the predictive confidence/uncertainty estimation as they are prone to overconfidence [70]. Consequently, it is required to properly calibrate the predictive confidence/uncertainty estimates. The violation of the calibration condition introduced in Equation (2.97) for confidence calibration and specified by Equation (2.98) for uncertainty calibration is quantifiable by different calibration quality measures, as defined in Section 2.4.5. Suppressing calibration errors is of utmost importance in safety-critical domains, like medical imaging [13] or autonomous driving [194, 190], where risk quantified by uncertainty influences life-and-death decisions.

The first evidence for a decoupling between neural network calibration measures was found by Nixon et al. [137]. They demonstrated that the minimum of the NLL over temperature curve does not coincide with minimum of the ECE over temperature curve [137] due to the bias introduced by the ECE binning mechanism. If the bias varies among different calibration quality measures, then those measures decouple and distinct optimal temperatures will be observed.

Wang et al. [176] alleviate the impact of the binning bias of ECE by introducing a novel calibration methodology. In detail, they tackle this issue by modulating the loss function of an arbitrary backbone network by a focal loss regularization factor with instance-wise gamma exponents [176]. For obtaining those instance-wise gamma values, an auxiliary neural network is trained with a differentiable calibration quality surrogate measure as a loss function. This surrogate measure is similar to the smoothed ECE proposed in Chapter 6 and serves the goal of ensuring continuous differentiability (C^1) for backpropagation. In summary, Wang et al. [176] achieved superior calibration quality by utilizing an unbiased kernel density estimate for the ECE and by training an auxiliary network to predict instance-wise gamma exponents to focus the backbone loss.

Islam et al. [85] pointed to another issue of model calibration, namely that the optimal temperature is typically class-dependent. This becomes important if the model is trained on a long-tailed dataset, where classes are highly imbalanced. As a result, the model tends

to be overconfident for high-frequency classes [85]. In order to address this issue they propose to incorporate the knowledge about the class distribution into the temperature scaling approach. In detail, they replace the scalar temperature by a temperature vector, which embeds class specific temperatures. Nevertheless, the degree of freedom of the temperature scaling approach remains at one, as the class dependency is solely accounted for by an additive term that consists of the max-normalized class frequencies [85]. They experimentally demonstrate that this class distribution-aware temperature scaling approach is superior in terms of the ECE and UCE [85].

In addition, Kirchenbauer et al. [97] observed that many common calibration methods, which suppress the calibration error in terms of the ECE, do not coherently minimize the generalized ECE (gECE) hinting on a decoupling between different calibration measures. The gECE evaluates the average performance of the ECE over the top-k classes and is more suitable for probabilistic classifiers [170], where the entire distribution is relevant for downstream decision making. As this thesis focuses on semantic segmentation, we will not consider the gECE as a calibration quality metric in our studies.

The current state of scientific knowledge raises some puzzling questions: Is the decoupling effect also noticeable for different neural network calibration measures? What factors drive the decoupling between those measures? What are the practical implications of calibration measure decoupling for safety-critical applications?

As a first contribution to tackle these questions, we will systematically analyze the phenomenon of decoupling among distinct neural network calibration measures by demonstrating experimentally that the optimal calibration temperature depends on the calibration measure under consideration. This analysis will comprise sparsification-based calibration measures (AUSE_V and AUSE_S) as well as measures derived from the reliability diagram (ECE, CCQS, UCE, and UCQS).

Secondly, we will identify several mechanisms responsible for calibration measure decoupling and provide experimental evidence. Specifically, we will demonstrate that the skewed frequency distribution of the uncertainty/confidence estimates (see Section 8.3), the class imbalances due to varying class representations within the dataset (see Section 8.4), the target domain mismatch regarding the performance metric (e.g., accuracy vs. IoU) between distinct calibration measures (see Section 8.5), and the binning procedure for reliability diagram-based measures (see Section 8.6) drive the decoherence of the optimal temperature and generate the observed decoupling.

Thirdly, we will propose the AUSE utilizing the cross-entropy for sorting as a measure for quantifying the residual uncertainty and we will elaborate on the risk associated with the decoupling of neural network calibration measures for autonomous driving.

8.2 Evaluation setup

For the evaluation of different calibration measures, a customized UNET architecture [149] is utilized. The UNET layout is based on the baseline multi-task network presented in Section 6.2 with five downsampling stages. Since the semantic segmentation task is the primary subject of interest, only the segmentation head is employed. As elaborated in detail in Section 2.4.1, there are three central pillars in CNN-based neural network design, namely depth, wideness and finesse. For the customized UNET architecture, the depth extends over 48 layers excluding the input and output layer. Furthermore, the wideness of the network doubles before every downsample step to conserve information and reaches a maximum of 304 channels at the UNET bottleneck. Additionally, the finesse of the input is given by 960×604 pixels and reaches a minimum in the UNET bottleneck of 30×19 pixels. In order to encourage the network to learn orthonormal features, kernel orthonormality regularization [95] is employed as a weight tensor regularizer in the convolutional layers. Moreover, the loss function is given by the balanced cross-entropy and utilizes a focal loss modulation in order to down-weight the loss assigned to well-classified instances. The cross-entropy is balanced by weighting the classes according to their frequency in the A2D2 dataset, similar to Dreissig et al. [42]. In total, the network comprises 7,271,110 trainable parameters and 1,178 non-trainable parameters, which track the first two statistical moments of the batch distribution during training mode for the batch normalization and are not subject to the loss backpropagation.

The training of the UNET was performed on two Nvidia RTX A6000 with 48GB and a learning rate schedule was applied similar to Section 6.2. Furthermore, the A2D2 segmentation dataset [56] served as the training dataset with a test ratio of 20%. Due to limitations in the computational resources, an evaluation dataset was selected from the test dataset by randomly selecting 200 images. On this subset, a mIoU of 74.8% was achieved with the UNET. This performance is significantly higher than the one observed with the segmentation head of the baseline multi-task model, presented in Section 6.2, as the training dataset is not augmented by optical aberrations of the Fourier-optical threat model. This decreases the generalization capabilities of the UNET due to a less heterogeneous training dataset but it leads to a higher baseline performance for in-domain samples. Consequently, the mIoU performance is higher but the robustness against optical perturbations induced by the windshield is reduced.

In alignment with Chapter 6, the groundtruth data was relabeled to match the 19 classes labeling taxonomy of the Cityscapes dataset [32] for enhancing comparability.

8.3 Decoupling due to skewed frequency distributions

Figure 8.1 illustrates the impact of temperature on the reliability diagram-based metrics (ECE, CCQS, UCE and UCQS). The optimal temperature ($T = 0.4$) is given by the location of the minimum in the mean ECE over temperature graph, which coincides with the optimal temperature of the mean UCE. It can be noticed that the enclosed areas A_{CC} and A_{UC}, which relate to the CCQS and UCQS, are not simultaneously minimized for the optimal temperature of the ECE/UCE. This is caused by the extreme skewness in the distribution of the predicted average confidences as well as the distribution of the predicted average uncertainties respectively, which is highlighted in Figure 8.2.

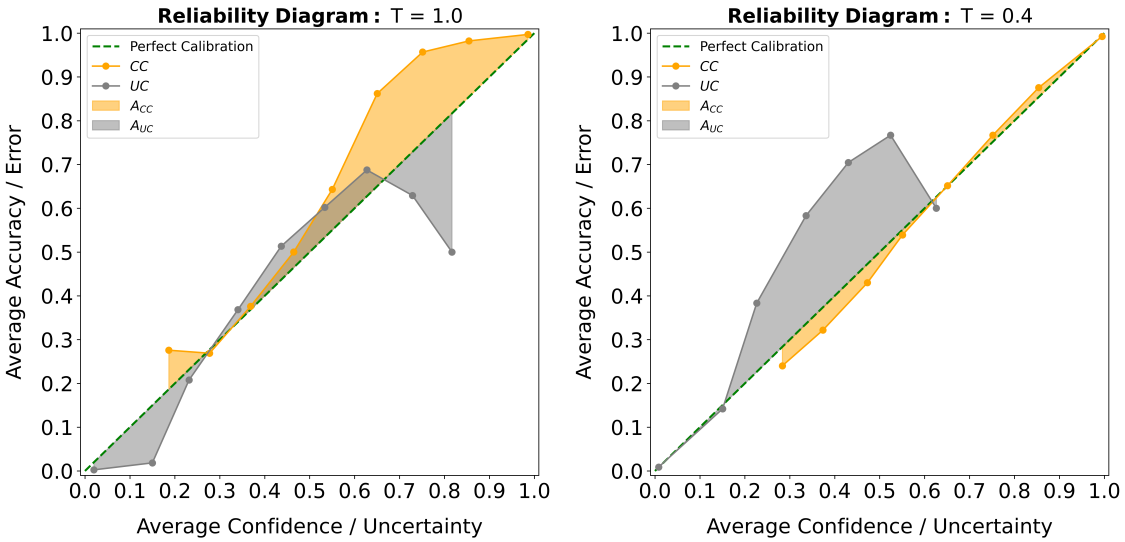


Figure 8.1: The reliability diagrams for the initial calibration state of the neural network ($T = 1.0$, left) and for the optimal temperature ($T = 0.4$, right) are illustrated along with the enclosed areas A_{CC} and A_{UC}.

The skewness accounts for the degree of asymmetry observed in the frequency distribution and is quantifiable by the third standardized moment. As a consequence, the predictions in the last bin (for the ECE) and the first bin (for the UCE) are dominating the calibration process. In a nutshell, the CCQS or the UCQS as a calibration measure target on the fulfillment of the calibration condition in Equations (2.97) and (2.98) over the entire average confidence/uncertainty range, whereas the ECE and the UCE are implicitly focused to a subdomain due to the skewed frequency distribution. In safety-critical situations, such a focus on highly confident predictions is not favorable because the sparsely observed scenarios with the lowest confidence will be the most critical ones for preventing an accident in autonomous driving.

The results presented in Figure 8.1 are separated into individual graphs in Figure 8.2. By doing so, the relationship between the frequency distribution of the confidence/uncertainty estimates and the bin-wise calibration error becomes evident. As stated above, the

skewness of the frequency distribution implies a focusing of the ECE and UCE on bins with high confidence/low uncertainty. Consequently, the areal calibration measures, CCQS and the UCQS, decouple from the implicitly focused measures, ECE and UCE, if the frequency distribution is skewed.

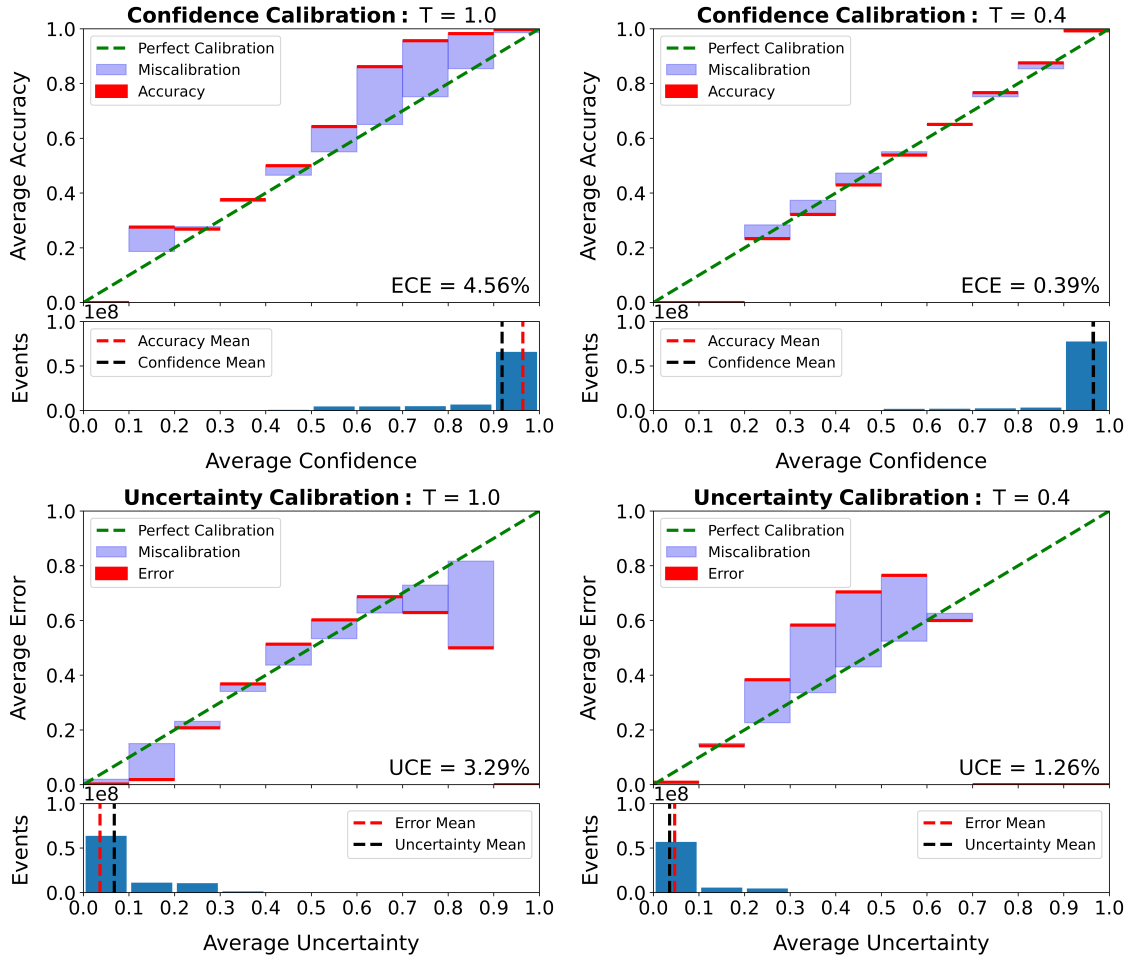


Figure 8.2: The binned reliability diagrams for the initial calibration state of the neural network ($T = 1.0$, left column) and for the optimal temperature ($T = 0.4$, right column) are visualized. The upper row utilizes the confidence estimate for evaluating the bin-wise violation of the calibration condition in Equation (2.97) and the lower row employs the Shannon entropy as a predictive uncertainty measure to assess the bin-wise breach of the calibration condition in Equation (2.98). The ideal case of perfect calibration is indicated as a green dashed line and the distance from this line to the bin-wise average accuracy/error (indicated as a red vertical line) is visualized as a blue bar. In addition, the frequency distributions of the confidence/uncertainty estimates are illustrated on the bottom of each reliability diagram. It is evident that all frequency distributions demonstrate a significant skewness.

8.4 Decoupling due to class-imbalances

In this section, another driving factor for the decoupling of calibration measures is studied, namely class-imbalances. For this, the ECE and the UCE are analyzed on a class-wise basis and they are contrasted against two equivalent metrics from the sparsification space, the AUSE calculated using the variation ratio as the sorting measure (AUSE_V) and the AUSE calculated using the Shannon entropy as the sorting measure (AUSE_S) respectively. For the computation of the AUSE, the IoU was used as a merit function. The optimal temperature that minimizes these four metrics differs from one class to another. This is attributable to the level of representation of a particular class in the dataset and ultimately to how well each class' features have been learned by the model during training. For the ECE, this class conditionality was noticed previously by Nixon et al. [137] and aligns well with the observations presented in Table 8.1, where the optimal calibration temperature T_{cal} is stated for the most represented and the least represented classes. At this point, it should be highlighted that this class-wise optimal temperature dependency is still observable even though the UNET was trained by employing a class balancing schema (see Equation (6.1)) and by utilizing a focal loss to accentuate instances with high prediction uncertainty. As a consequence, the observation of distinct class-wise temperatures might be better explainable in terms of the epistemic uncertainty in the class-specific feature embeddings. This dependency is not taken into account by the point-wise variation ratio nor by the Shannon entropy, wherefore the bias w.r.t. the epistemic uncertainty contribution can vary across the classes, which imposes the causal reason for measuring a temperature dependence on classes.

In addition to the observed class dependency of the calibration temperature in Table 8.1, it is evident that the class-wise optimal temperatures obtained by the reliability diagram-based measures (ECE and UCE) and the sparsification-based measures (AUSE_V and AUSE_S) do not coincide. This points to another decoupling effect, which will be addressed in Section 8.5.

Class	**Reliability**				**Sparsification**			
	$\underset{T}{\text{argmin}}$ ECE		$\underset{T}{\text{argmin}}$ UCE		$\underset{T}{\text{argmin}}$ AUSE_V		$\underset{T}{\text{argmin}}$ AUSE_S	
Nature	0.4	± 0.05	0.4	± 0.05	1.8	± 0.17	0.9	± 0.38
Sky	0.4		0.4		1.4		0.9	
Road	0.5		0.5		1.7		0.1	
Pedestrian	0.3	± 0.31	0.4	± 0.40	0.9	± 0.52	0.9	± 0.17
Small Vehicle	0.2		0.3		1.9		1.0	
Animal	0.9		1.2		0.7		0.6	

Table 8.1: T_{cal}, which minimizes the respective calibration metric is listed for the three most and least represented classes in the evaluation set (top three and bottom three respectively). It is noticeable that the standard deviation of the class-wise T_{cal} is amplified if a subset of less represented classes is considered.

8.5 Decoupling due to different target domains

In this section, the decoherence between sparsification-based measures (AUSE_V and AUSE_S) and reliability diagram-based measures (ECE and UCE), observed in Section 8.4, will be further investigated. In order to retrieve a holistic picture, the average calibration error is computed across all classes for calibration temperatures in the range of $T \in [0.1, 10]$. For example, the class-wise ECEs are calculated at every temperature and their means are plotted in Figure 8.3.

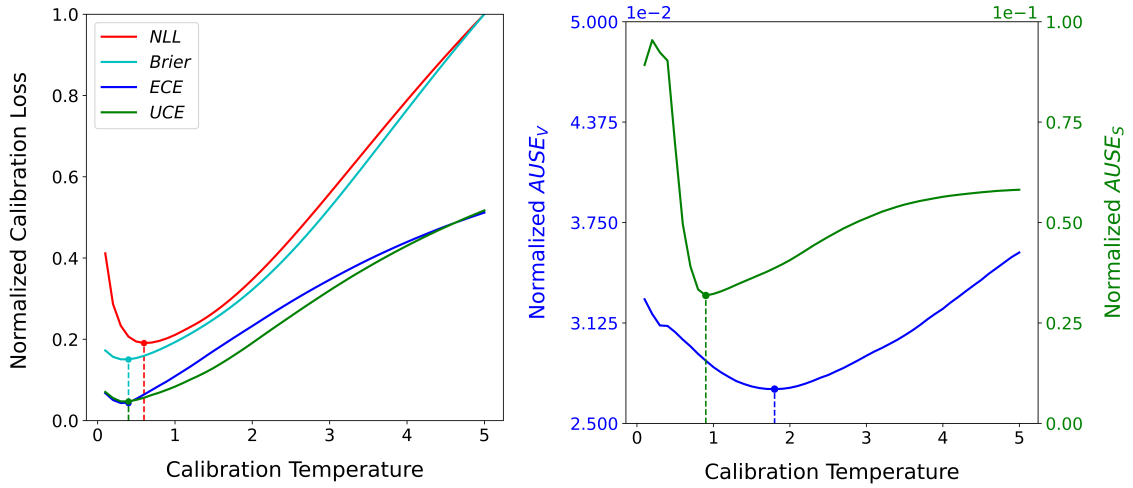


Figure 8.3: Left: The normalized calibration loss surface of NLL, Brier score, the mean of class-wise ECE and the UCE are plotted. The optimal temperature for the Brier score, ECE and UCE coincide at $T = 0.4$ while the NLL indicates a minimum at $T = 0.6$. Right: The normalized calibration loss surface of AUSE_V and AUSE_S indicate minima at T=1.8 and T=0.9 respectively.

It should be stressed that there is a difference between evaluating the ECE/UCE holistically over all classes, as done in Section 8.3, and assessing the average of the individual ECE/UCE obtained on a class by class basis, as analyzed in Section 8.4. This difference was initially addressed by Kull et al. [104] and arises from the fact that the binning process in the reliability diagram does not commute with the averaging operation. In this study, the average of the class-wise ECE/UCE scores is selected for quantifying the total calibration quality in the multi-class classification problem as the AUSE is based on the IoU, which requires a class-wise analysis. This requirement is based on the observation that the sparsification and the oracle curve can intersect if the mIoU is used as a merit function. The intersection might occur due to class imbalances, wherefore the n-th wrongly classified pixel that is removed in the sparsification process increases the mIoU non-uniformly. Hence, the underlying assumption for generating the oracle curve, namely that wrongly classified pixels, which are randomly dropped, are equally detrimental, is violated as changes in the mIoU are class-sensitive.

The temperature dependency of the calibration measures from the sparsification space are contrasted against the reliability diagram-based measures in Figure 8.3. For comparability reasons, the individual calibration measures were normalized. It is evident that decoherence between sparsification-based measures (AUSE_V and AUSE_S) and reliability diagram-based measures (ECE and UCE) also holds true if all classes are considered. This decoupling effect might arise from the fact that optimizing the accuracy is not equivalent to optimizing the IoU highlighting the target domain disparity. As a reminder, the ECE measures the bin-wise confidence to accuracy mismatch and UCE measures the bin-wise uncertainty to error mismatch. Hence, the calibration error is determined in the error/accuracy domain and integrated by considering the bin cardinalities. On the other hand, the AUSE measures quantify the calibration error in terms of the IoU mismatch between the sparsification and the oracle curve, integrated over the entire domain. As an additional contributor for the observed decoupling, the binning operation should be considered. The AUSE acts on a pixel-per-pixel basis, whereas the ECE and UCE rely on the reliability diagram binning. This factor will be discussed in the subsequent section.

As an aside, the optimal temperature retrieved for the ECE and UCE by averaging over all class-wise scores coincides with the holistically calculated ECE and UCE temperature determined in Section 8.3, which is a remarkable coincidence and is not expected as elaborated above. On the contrary, the coherence between the ECE and UCE does not hold true for the corresponding counterpart in the sparsification framework, as the AUSE_V and the AUSE_S demonstrate distinct minima.

8.6 Decoupling due to binning

Figure 8.3 also visualizes the NLL and the Brier score as a function of temperature. According to the original work of Guo et al. [70] on temperature scaling, the optimal calibration temperature is given if the NLL is minimized. It is evident that the optimal temperature retrieved from the NLL analysis does not coincide with the temperature associated with the ECE minimum. This conclusion is congruent with previous studies of Nixon et al. [137] and points to another driving force for decoupling, namely the binning mechanism of ECE and UCE. As the bin-wise average accuracy/error is evaluated, the predictions are implicitly smoothed, which introduces a bias w.r.t. the NLL. On the other hand, the clustering of instances across the confidence/uncertainty domain is required for an accurate calibration quality evaluation, as the limit of number of bins equals the number of instances, results in an imprecise binary accuracy/error estimate. In this particular case, the ECE is identical to the Brier score. In a nutshell, the degree of freedom introduced by the number of bins in the reliability diagram is another source for decoupling.

8.7 AUSE as an uncertainty estimator

As introduced in Section 2.4.5, the AUSE is suppressed to zero if the sparsification curve aligns with its supremum, the oracle curve. If this holds true, then the associated uncertainty measure is unbiased and the calibration error vanishes. In the sparsification process, the predictive uncertainty measure under test should not rely on the groundtruth labels as they are not available for unknown instances during employment.

Nevertheless, if the AUSE methodology is transformed to serve as an uncertainty measure itself, instead of quantifying the bias of a given uncertainty measure, then the sparsification curve may rely on the groundtruth labels. As the supremum of the sparsification curve is imposed by the oracle curve, the AUSE_{CE} obtained by utilizing the cross-entropy for sorting the predictions – which is the standard loss function for multi-class semantic segmentation – is proposed as an estimator for the residual uncertainty. The residual uncertainty is non-reducible if the network architecture is fixed and comprises the aleatoric uncertainty of the underlying data generation process as well as the model uncertainty [80] due to the limitations in the hypothesis space induced by the particular choice of the neural network architecture. This holds true if and only if the model training converged to the infimum in the loss landscape, which is extremely unlikely. Therefore, the residual uncertainty also entails a second component of epistemic nature, namely the approximation uncertainty, which quantifies the uncertainty arising from the difference between the local minimum, where the training converged, and the infimum of the loss landscape. Hence, the AUSE_{CE} is supposed to decline as the network training progresses.

In order to provide first evidence for this hypothesis, an ensemble consisting of five UNET models was trained with i.i.d. initialization and equivalent hyperparameterization. The individual model training was terminated after a fixed and distinct number of epochs and the residual uncertainty was evaluated in terms of the AUSE_{CE}. For this, the most prevalent classes, based on their representation in the chosen evaluation set, are considered for the evaluation. The AUSE_{CE} over the number of epochs is plotted in Figure 8.4 and indicates strictly monotonic decreasing graphs, which supports the hypothesis.

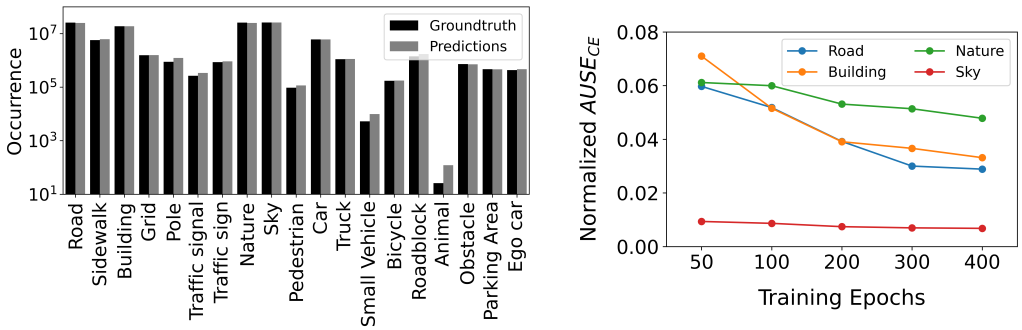


Figure 8.4: Only classes with an occurrence of $> 10^7$ are considered. It is evident that the class-wise AUSE_{CE} converges to the lowest value for the 400-epoch model.

8.8 Interim conclusion

In this section, the optimal temperature for temperature scaling was determined by considering several distinct calibration quality measures. Based on evaluations with a customized UNET model on the A2D2 dataset, evidence on the decoupling between well-established calibration measures was found. In detail, four key observations were made that drive the decoherence between different calibration measures.

First of all, Section 8.3 highlights that the areal calibration measures CCQS and UCQS decouple from the bin-cardinality weighted measures ECE and UCE due to the tremendous skewness in the frequency distribution. Nevertheless, by minimizing the ECE and UCE, computed across all classes, the same optimal calibration temperature was observed. This is remarkable considering that the UCE uses the entire prediction vector and the ECE relies exclusively on the probability mass allocated to the most likely class. As a second driving factor for the decoupling effect between neural network calibration measures, the class dependency of the optimal temperature was identified in Section 8.4. The class conditionality of ECE has already been known from previous studies by Nixon et al. [137] but the results presented in Section 8.4 indicate that this dependency also applies to the AUSE. Thirdly, Section 8.5 provides evidence on the decoupling between the sparsification world and accuracy-based measures. This might be explainable by the difference in the target domain since the AUSE considers the IoU mismatch, whereas the ECE penalizes deviations in the accuracy. Only in the case of vanishing true negative predictions, these two measures are congruent, wherefore it is expected that in such a scenario the decoupling effect between the AUSE and the ECE is minimized but still observable as the binning of the ECE also induces a bias. This gives rise to a fourth decoupling mechanism, driven by the implicit smoothing during the reliability diagram binning, which inherently biases the ECE and UCE in comparison to the NLL as elaborated in Section 8.6.

Furthermore, the AUSE_{CE} is proposed as a measure of the residual uncertainty if predictions are sparsified according to the cross-entropy. The residual uncertainty quantifies the remaining uncertainty originating from the data generation process (aleatoric nature) as well as the model uncertainty due to limitations in the hypothesis space given a fixed neural network architecture (epistemic nature), if and only if, the training converged to the infimum of the loss landscape. If this constraint is fulfilled, then the model uncertainty can be isolated by subtracting the aleatoric uncertainty component, which is quantified by the expected conditional entropy according to Equation (2.95).

In a nutshell, the results of this section provide evidence on the decoupling between several calibration measures, which gives rise to one or more degrees of freedom in the confidence/uncertainty calibration of neural networks. This undermines the goal of determining unique prediction intervals and generates an ambiguity that compromises trustworthiness and ultimately prevents the homologation of safety-critical applications based on neural networks. As a way forward, it should first be identified, which calibration

measures are the least biased distance measures with respect to the confidence/uncertainty calibration condition in Equation (2.97) and (2.98) from a mathematical point of view. Furthermore, it should be considered to account for additional aspects in the definition of calibration quality measures. For example, it might be advisable to increase the sensitivity for sparsely observed scenarios – with low confidence and high uncertainty – as those instances correspond to safety-critical long-tail scenarios. Consequently, the bias of the confidence/uncertainty estimates is more harmful for autonomous driving functionalities within this domain, which indicates that a focused calibration measure might be beneficial, similar to loss-focusing [112].

9 Benefits for autonomous driving

The motivation for this thesis was given in Section 1.1 by highlighting the limitations of windshields in terms of optical quality. In order to handle the aberrations induced by the windshield, a plan of action was presented in Figure 1.1, which decomposed the perturbation operator $\hat{\mathcal{H}}$ into a tilting operator $\hat{\mathcal{T}}$ and a blurring operator $\hat{\mathcal{B}}$.

The tilting operator $\hat{\mathcal{T}}$ was addressed by a non-parametric, grid-based approach to restore conformality by rectifying the local optical distortions induced by the windshield. For providing the required high-resolution measurement information about the point-wise deflection vectors, BOS imaging was proposed. As BOS imaging implicitly accounts for the varying incident angle across the field of view, it is predestined for capturing the windshield's optical distortions as seen by the ADAS camera. Hence, BOS imaging provides unbiased measurement information about the local windshield's distortions, which enables the rectification of the tilting operator $\hat{\mathcal{T}}$ for ensuring conformality. Consequently, the perturbation operator $\hat{\mathcal{H}}$ simplifies to the blurring operator $\hat{\mathcal{B}}$ because the residual tilting operator after correction is supposed to be neglectable.

As the blurring operator $\hat{\mathcal{B}}$ shall be suppressed by imposing corresponding quality requirements on the windshield, the question arose whether the current quality measures in place are adequate. This question was tackled by starting to investigate the fundamental differences between present optical measures. In detail, the local refractive power evaluates local wavefront curvature modulations. On the contrary, the MTF is a global property of the optical system taking into account the wavefront aberrations across the entire aperture surface. This key distinction was underpinned by analytical arguments and experimental evidence. As a result, only the entire local refractive power distribution across the aperture surface entails sufficient information to draw conclusions about the optical system performance. The local refractive power itself, measured exclusively along the canonical axes, does not even comprise sufficient information to determine the local Gaussian curvature, which correlates to the areal spread of the geometrical PSF. Furthermore, it was demonstrated that the MTF does not allow for part-level measurements if the ADAS camera objective lens and the windshield entail optical aberrations. Consequently, neither the local refractive power nor the MTF are adequate optical quality measures for characterizing part-level requirements for the windshield, such that the blurring operator $\hat{\mathcal{B}}$ can effectively be suppressed. The magnitude of residual blurring that is acceptable for the perception system needs to be derived from a pre-defined threshold on the level of risk tolerable in terms of the target KPI. For semantic segmentation, such a threshold can be translated into a system-level requirement for the maximal allowed

residual blurring by exploiting the regression function presented in Equation (5.1). In the ideal case, the blurring operator $\hat{\mathcal{B}}$ will solely consist of the convolution operator with the diffraction-limited PSF as the kernel function. In a nutshell, in order to account for interference effects, future part-level requirements might need to be defined in the complex-valued domain of the electromagnetic field, quantified by a maximum tolerable wavefront modulation in terms of Zernike coefficients.

On the contrary, for selecting an appropriate optical quality measure to impose system-level requirements, the correlation of those measures to the AI performance needed to be established. By utilizing the Fourier optical threat model presented in Section 2.1.2, it was found that optical measures, which incorporate additional information about the PSF (esp. Strehl ratio and OIG), demonstrate a superior correlation to neural network KPIs, esp. mIoU and mECE. The degree of correlation was quantified by the Chatterjee rank correlation measure, which denotes the fraction of explainable variance over the total variance. The non-linear correlation was cast into a functional relationship given by Equation (5.1), which adequately captures the interplay between neural network KPIs and PSF-based optical measures. Those results were consistent across three different CNN-based architectures (HRNet, MTL and UNET) as well as three different datasets (KITTI, CARIAD and A2D2), which leads to the hypothesis that the conclusions are network architecture independent as long as CNN-based models are considered. Consequently, by adopting PSF-based optical measures (esp. Strehl ratio and OIG) suitable ADAS working requirements can be derived on a system-level. By doing so, the impact of the blurring operator on the imaging process can be controlled and suppressed if sufficiently strong quality requirements on the windshield are imposed. On the contrary, the system-level MTF evaluated at half-Nyquist frequency is an inadequate quality measure as it does not demonstrate a functional relationship with the AI performance, which is driven by the non-monotonicity of the MTF curve in the large aberration regime. The system performance degradation can be broken down to the impact of the Zernike polynomials of second radial order. Among those aberration patterns, defocus is the dominant contributor in terms of Shapley values, which sets the agenda for optimizing the manufacturing process of the windshield.

As observed in the correlation study above, the impact of the blurring operator does not only degenerate the mIoU performance but also the neural network confidence calibration. This observation is mainly driven by the lack of robustness of temperature scaling under dataset shifts [39]. As optical aberrations of the windshield also induce a dataset shift between the training dataset and the instances captured during deployment, this highlights another safety-critical aspect that has to be accounted for in the plan of action presented in Figure 1.1. To address this issue, it was proposed to enrich the state-of-the art calibration method PTS, which utilizes a post-hoc calibration network to predict instance-wise temperatures, by a physical inductive bias that correlates with the dataset shift magnitude. It was experimentally demonstrated that embedding the Zernike

coefficient vector as a physical inductive bias into the PTS network leads to superior calibration robustness against dataset shifts. By employing this PIPTS framework, the dataset shift related bias in the confidence estimates can be mitigated, which strengthens the trustworthiness in the baseline model's predictions. As the system-level Zernike coefficient vector has to be estimated by the baseline model alongside the target application in order to apply PIPTS, a system health online monitoring of the perception system is inherently inbuilt. To utilize this feature, safety margins in terms of the Zernike coefficients need to be derived by exploiting the discovered functional relationship between PSF-based optical measures and neural network KPIs. If the predicted Zernike coefficients undermine the safety requirements, then the perception chain is compromised and car maintenance might be necessary. Hence, the dynamical real-time monitoring of the hazard potential reduces the risk of encountering situations that could jeopardize safety. This comprises a wide spectrum of degradation-related dataset shifts. For example, external effects, like rock chips or severe weather conditions, influence the effective optical quality of the perception system. Furthermore, internal factors of the windshield, like thermal effects due to windshield heating or windshield ageing, also contribute to a shift in the imaging quality. Overall, PIPTS alleviates the impact of aberration-related dataset shifts on the confidence calibration by incorporating a physical inductive bias, which correlates with the bias in the uncertainty measure. Hence, PIPTS enhances the robustness of AI-based autonomous driving functionalities under optical aberrations – which ultimately permits for a wider definition of part-specific requirements – and fosters their trustworthiness in real-world deployment.

In order to leverage the synergies of PIPTS, the Zernike coefficient vector characterizing the wavefront modulation induced by the windshield must be known a priori. It was shown by a proof of concept study that BOS imaging can be adapted to measure the wavefront aberrations of the windshield if specific resolution requirements are fulfilled. With this information, the PIPTS training can be initiated and the data augmentation subspace can be enriched by artificially generating subsamples utilizing a Fourier-optical degradation model, as presented in Section 2.1.2. Hence, the heterogeneity of the training dataset can be enhanced, which amplifies the model generalizability as the neural network learns more universal feature representations that reduce the overfitting-related variance. The boost in robustness against optical aberrations by increasing the sample density in the data augmentation subspace typically requires experimental data captured behind different windshield configurations, which is expensive, time consuming and finally generates a huge CO_2 footprint in the development phase if the test mules are not powered by climate neutral energy sources. Here, BOS imaging enables the deployment of the Fourier-optical degradation model as the Zernike coefficient vector characterizing the windshield's wavefront modulation becomes available independent of the ADAS camera. This augmentation strategy only relies on a baseline training dataset, which was initially captured without a windshield in the optical path (e.g. A2D2 [56], KITTI [55]).

In a nutshell, with the contributions presented in this thesis, it is possible to correct for the tilting operator $\hat{\mathcal{T}}$ by utilizing high-resolution measurements of point-wise deflection vector maps obtained through BOS imaging. Furthermore, it was experimentally demonstrated that PSF-based optical measures (esp. Strehl ratio and OIG) can be utilized to specify optical system requirements. Those system-level requirements can be backpropagated to part-level requirements by inverting the deterministic Fourier optical model. It has been shown that the part-level requirements need to be defined in the complex-valued domain of the electromagnetic field in order to account for interference effects. In summary, by imposing sufficiently strong optical quality requirements, the effect of the blurring operator $\hat{\mathcal{B}}$ on the imaging process can be limited such that a maximal tolerable mIoU drop / mECE rise is not exceeded. With the proposed PIPTS calibration procedure, the robustness of the confidence calibration under dataset shifts induced by wavefront aberrations is enhanced, which might allow for the relaxation of part-level requirements without violating the safety condition in terms of the mECE.

Are we done?

Not, yet. By training the PIPTS calibrator, distinct minima for the ECE over temperature and the AUREC over temperature curve were identified. This contradicts the goal of ensuring model trustworthiness by imposing requirements on the neural network calibration error in terms of the mECE. As both minima correspond to different optimal temperatures, the estimated bias of the predicted confidences varies. This indicates a decoupling of neural network calibration measures, which ultimately prevents the issuing of safety guarantees. Within this thesis, four driving factors for the decoherence between different calibration measures were experimentally identified. First of all, the enormous skewness in the frequency distribution drives the decoupling between the areal calibration measures, e.g. CCQS and UCQS, and the bin-cardinality weighted measures, e.g. ECE and UCE. Secondly, the optimal temperature retrieved from the sparsification analysis as well as the temperatures determined by the reliability diagram-based measures are class dependent. Thirdly, the target domain disparity gives rise to a decoherence between accuracy-based measures and sparsification measures. Finally, the binning mechanism of the ECE and UCE imposes a bias w.r.t. the NLL. In a nutshell, ensuring the trustworthiness of the model's predictions by imposing safety requirements on the mECE is insufficient as the calibration quality measures itself are subject to biases. Hence, the observed decoupling effect generates a degree of freedom in the confidence/uncertainty calibration of neural networks that needs to be lifted in order to allow for a holistic safety strategy. This underpins the necessity for future research into the field of unbiased uncertainty quantification for neural network predictions. Those research efforts should aim for determining unique prediction intervals, which satisfy the condition of perfect calibration formulated in Equation (2.97-2.98). If this mission is accomplished, then safety guarantees for the predictive performance of neural networks can be granted on a pre-specified

confidence level, which is an indispensable requirement for establishing trustworthiness in autonomous driving.

Furthermore, even if we successfully resolve the aforementioned puzzle of determining unique prediction intervals, the question might come up, if the production-related quality limitations of the windshield allow us to reach the requested confidence level. In this thesis, several strategies have been presented to tackle this task, e.g. Chapter 7 paves the way for increasing the performance robustness of the mIoU under optical aberrations by applying data augmentation on Fourier-optical principles and Chapter 6 demonstrated that the dataset shift driven bias of the confidence estimation is reducible by a physical inductive bias. These efforts significantly contribute to our objective. Nevertheless, they might be insufficient for reaching the desired safety level. A possible way forward to further alleviate the impact of optical aberrations of the windshield is given by optimizing the perception system holistically, end-to-end. This conception is typically referred to as a deep optics [25, 168, 198] approach. In deep optics, it is the idea to assign optical parameters as trainable hyperparameters to configure an optical perception system that suits the target application the most. As a result, the system might not be optimized such that the Kullback–Leibler divergence between the system PSF and the diffraction-limited PSF is minimized. Instead, the most optically informative PSF [98] is learned. For instance, when estimating the longitudinal distance of objects from the camera by monocular images, coding an artificial defocus blur into the PSF can be advantageous [25]. This blur will then introduce depth-dependent sharpness variations, thereby encoding depth-related information. As a result, if optical aberrations are properly tuned, the feature embeddings and the neural network performance can be enhanced. As the windshield exhibit significant part by part variations due to the limited controllability of the production process [84], a standard deep optics approach, where the trainable optical hyperparameters are scalars is insufficient. Nevertheless, the deep optics approach can be adjusted to consider those production tolerances a priori by assigning the trainable optical hyperparameters as distributions with pre-defined variances. This idea was recently proposed by Dai et al. [36] and is similar to variational Bayesian neural networks, except that the means of the Gaussian distributions are learned exclusively. This is the most promissing next step forward in order to grant safety guarantees on high confidence levels in the light of the production-related optical quality limitations of windshields.

Bibliography

[1] Franz Adunka. *Messunsicherheiten: Theorie und Praxis*. Vulkan Verlag, 2007. ISBN: 3-8027-2205-1.

[2] Andrea Agostinelli, Michal Pándy, Jasper Uijlings, Thomas Mensink, and Vittorio Ferrari. "How Stable Are Transferability Metrics Evaluations?" In: *Computer Vision – ECCV 2022*. Cham: Springer Nature Switzerland, 2022, pp. 303–321. ISBN: 978-3-031-19830-4.

[3] B. B. Alchagirov and A. M. Chochaeva. "Temperature dependence of the density of liquid tin". In: *High Temperature* 38.1 (2000), pp. 44–48. DOI: 10.1007/BF02755565. URL: https://doi.org/10.1007/BF02755565.

[4] Torben E. Andersen, Mette Owner-Petersen, and Anita Enmark. "Image-based wavefront sensing for astronomy using neural networks". In: *Journal of Astronomical Telescopes, Instruments, and Systems* 6.3 (2020), p. 034002. DOI: 10.1117/1.JATIS.6.3.034002. URL: https://doi.org/10.1117/1.JATIS.6.3.034002.

[5] Reza Azad, Moein Heidary, Kadir Yilmaz, Michael Hüttemann, Sanaz Karimijafarbigloo, Yuli Wu, Anke Schmeink, and Dorit Merhof. *Loss Functions in the Era of Semantic Segmentation: A Survey and Outlook*. 2023. arXiv: 2312.05391 [cs.CV]. URL: https://arxiv.org/abs/2312.05391.

[6] Chayan Banerjee, Kien Nguyen, Clinton Fookes, and Karniadakis George. "Physics-Informed Computer Vision: A Review and Perspectives". In: *ACM Comput. Surv.* 57.1 (Oct. 2024). ISSN: 0360-0300. DOI: 10.1145/3689037. URL: https://doi.org/10.1145/3689037.

[7] David Barber and Christopher M. Bishop. "Ensemble Learning in Bayesian Neural Networks". In: *Neural Networks and Machine Learning* (1998), pp. 215–237.

[8] Matthias Bartelmann, Björn Feuerbacher, Timm Krüger, Dieter Lüst, Anton Rebhan, and Andreas Wipf. *Theoretische Physik*. Springer Spektrum Berlin, Heidelberg, 2015. URL: http://dx.doi.org/10.1007/978-3-642-54618-1.

[9] Christian Bauckhage and Daniel Speicher. *Lecture Notes on Machine Learning: Neurons with Non-Monotonic Activation Functions*. Tech. rep. University of Bonn, July 2019.

[10] A. B. Bhatia and E. Wolf. "On the circle polynomials of Zernike and related orthogonal sets". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 50.1 (1954), pp. 40–48. DOI: 10.1017/S0305004100029066.

[11] A Bichal and B S Thurow. "On the application of background oriented schlieren for wavefront sensing". In: *Measurement Science and Technology* 25 (1 Jan. 2014), p. 015001. ISSN: 0957-0233. DOI: 10.1088/0957-0233/25/1/015001. URL: https://iopscience.iop.org/article/10.1088/0957-0233/25/1/015001.

[12] Abhishek Bichal and Brian Thurow. "Development of a Background Oriented Schlieren-Based Wavefront Sensor for Aero-Optics". In: *40th Fluid Dynamics Conference and Exhibit* (June 2010). DOI: 10.2514/6.2010-4842. URL: https://arc.aiaa.org/doi/10.2514/6.2010-4842.

[13] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer New York, NY, 2006. ISBN: 978-0-387-31073-2.

[14] Sebastian Bordt, Uddeshya Upadhyay, Zeynep Akata, and Ulrike von Luxburg. "The Manifold Hypothesis for Gradient-Based Explanations". In: *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2023, pp. 3697–3702. DOI: 10.1109/CVPRW59228.2023.00378.

[15] Max Born, Emil Wolf, A. B. Bhatia, P. C. Clemmow, D. Gabor, A. R. Stokes, A. M. Taylor, P. A. Wayman, and W. L. Wilcock. *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. 7th ed. Cambridge University Press, 1999. DOI: 10.1017/CBO9781139644181.

[16] Alexander Braun. "Automotive mass production of camera systems: Linking image quality to AI performance". In: *tm - Technisches Messen* 90.3 (2023), pp. 205–218. DOI: doi:10.1515/teme-2022-0029. URL: https://doi.org/10.1515/teme-2022-0029.

[17] I.N. Bronstein and K.A. Semendjaev. *Taschenbuch der Mathematik*. 10th ed. Verlag Europa-Lehrmittel, Nourney, Vollmer GmbH & Co. KG, 2016.

[18] D. C. Brown. "Close-range camera calibration". In: *Photogrammetric Engeneering* 37 (1971), pp. 855–866.

[19] D. C. Brown. "Decentering distortion of lenses". In: *Photogrammetric Engeneering* 22 (1966), pp. 444–462.

[20] D. C. Brown. "The Bundle Adjustment - Progress and Prospectives". In: *International Archives of Photogrammetry* 21 (1976), pp. 444–462.

[21] David E. Burmaster and Andrew M. Wilson. "An introduction to second order random variables in human health risk assessments". In: *Human and Ecological Risk Assessment: An International Journal* 2.4 (1996), pp. 892–919. DOI: 10.1080/10807039609383655.

[22] Campbell, H.I. and Greenaway, A.H. "Wavefront Sensing: From Historical Roots to the State-of-the-Art". In: *EAS Publications Series* 22 (2006), pp. 165–185. DOI: 10.1051/eas:2006131. URL: https://doi.org/10.1051/eas:2006131.

[23] Manfredo Perdigão do Carmo. *Differential geometry of curves and surfaces*. Prentice-Hall, 1976.

[24] Stanley H. Chan. "Tilt-Then-Blur or Blur-Then-Tilt? Clarifying the Atmospheric Turbulence Model". In: *IEEE Signal Processing Letters* 29 (2022), pp. 1833–1837. DOI: 10.1109/LSP.2022.3200551.

[25] Julie Chang and Gordon Wetzstein. "Deep Optics for Monocular Depth Estimation and 3D Object Detection". In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2019, pp. 10192–10201. DOI: 10.1109/ICCV.2019.01029.

[26] Sourav Chatterjee. "A New Coefficient of Correlation". In: *Journal of the American Statistical Association* 116.536 (2021), pp. 2009–2022. DOI: 10.1080/01621459.2020.1758115.

[27] Bowen Cheng, Maxwell D. Collins, Yukun Zhu, Ting Liu, Thomas S. Huang, Hartwig Adam, and Liang-Chieh Chen. "Panoptic-DeepLab: A Simple, Strong, and Fast Baseline for Bottom-Up Panoptic Segmentation". In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020, pp. 12472–12482. DOI: 10.1109/CVPR42600.2020.01249.

[28] Roberto Cipolla, Yarin Gal, and Alex Kendall. "Multi-task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2018, pp. 7482–7491. DOI: 10.1109/CVPR.2018.00781.

[29] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. "Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)". In: *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*. 2016. URL: http://arxiv.org/abs/1511.07289.

[30] W. G. Cochran. "The distribution of quadratic forms in a normal system, with applications to the analysis of covariance". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 30.2 (1934), pp. 178–191. DOI: 10.1017/S0305004100016595.

[31] A. E. Conrady. "Decentred Lens-Systems". In: *Monthly Notices of the Royal Astronomical Society* 79.5 (Mar. 1919), pp. 384–390. ISSN: 0035-8711. DOI: 10.1093/mnras/79.5.384. URL: https://doi.org/10.1093/mnras/79.5.384.

[32] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. "The Cityscapes Dataset for Semantic Urban Scene Understanding". In: *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.

[33] R. Courant and D. Hilbert. *Methods of Mathematical Physics*. John Wiley & Sons, 2024. ISBN: 978-3-527-41447-5. DOI: 10.1002/9783527617210.

[34] Thomas M Cover. *Elements of information theory*. John Wiley & Sons, 1999.

[35] Harald Cramér. *Mathematical Methods of Statistics*. Princeton: Princeton University Press, 1946. ISBN: 9781400883868. DOI: doi:10.1515/9781400883868. URL: https://doi.org/10.1515/9781400883868.

[36] Jun Dai, Liqun Chen, Xinge Yang, Yuyao Hu, Jinwei Gu, and Tianfan Xue. *Tolerance-Aware Deep Optics*. 2025. arXiv: 2502.04719 [cs.CV]. URL: https://arxiv.org/abs/2502.04719.

[37] A. Deitmar and S. Echterhoff. *Principles of Harmonic Analysis*. Springer New York. ISBN: 9780387854687. 2008.

[38] Stefan Depeweg, José Miguel Hernández-Lobato, Finale Doshi-Velez, and Steffen Udluft. "Decomposition of Uncertainty in Bayesian Deep Learning for Efficient and Risk-sensitive Learning". In: *International Conference on Machine Learning*. 2017. URL: https://api.semanticscholar.org/CorpusID:3461501.

[39] Jessica Deuschel, Andreas Foltyn, Karsten Roscher, and Stephan Scheele. "The Role of Uncertainty Quantification for Trustworthy AI". In: *Unlocking Artificial Intelligence: From Theory to Applications*. Ed. by Christopher Mutschler, Christian Münzenmayer, Norman Uhlmann, and Alexander Martin. Cham: Springer Nature Switzerland, 2024, pp. 95–115. ISBN: 978-3-031-64832-8. DOI: 10.1007/978-3-031-64832-8_5. URL: https://doi.org/10.1007/978-3-031-64832-8_5.

[40] DIN-52305. "Determination of the optical deviation and refractive power of safety glass for vehicle glazing". In: *German Institute for Standardization* (1995). URL: https://www.beuth.de/de/norm/din-52305/2533097.

[41] Alexey Dosovitskiy et al. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale". In: *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. 2021.

[42] Mariella Dreissig, Florian Piewak, and Joschka Boedecker. "On the Calibration of Uncertainty Estimation in LiDAR-based Semantic Segmentation". In: *IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)* (2023), pp. 4798–4805.

[43] Mariella Dreissig, Florian Piewak, and Joschka Boedecker. "On the calibration of underrepresented classes in LiDAR-based semantic segmentation". In: *IROS 2022 Workshop on Probabilistic Robotics in the Age of Deep Learning*. 2022. URL: https://openreview.net/forum?id=YzRNccnAdd6.

[44] Simon Ek, Vassily Kornienko, Adrian Roth, Edouard Berrocal, and Elias Kristensson. "High-speed videography of transparent media using illumination-based multiplexed schlieren". In: *Scientific Reports* 12.1 (2022), p. 19018. DOI: 10.1038/s41598-022-23198-6. URL: https://doi.org/10.1038/s41598-022-23198-6.

[45] Badr Ben Elallid, Nabil Benamar, Abdelhakim Senhaji Hafid, Tajjeeddine Rachidi, and Nabil Mrani. "A Comprehensive Survey on the Application of Deep and Reinforcement Learning Approaches in Autonomous Driving". In: *Journal of King Saud University - Computer and Information Sciences* 34.9 (2022), pp. 7366–7390. ISSN: 1319-1578. DOI: https://doi.org/10.1016/j.jksuci.2022.03.013. URL: https://www.sciencedirect.com/science/article/pii/S1319157822000970.

[46] Kerkhof F. "Optische Wirkungen von Flachglas mit unebenen Oberflächen". In: *Glastechnische Berichte* 25 (1952). URL: https://www.tib.eu/en/search/id/ei-backfile:c84_f4ab65fd80746efdM7a0719817173212/Optical-effects-of-flat-glass-with-uneven-surfaces?cHash=396abec3815801caf47fb55f77b43461.

[47] R. W. Farebrother. "Further Results on the Mean Square Error of Ridge Regression". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 38.3 (1976), pp. 248–250. ISSN: 00359246. URL: http://www.jstor.org/stable/2984971.

[48] Charles Fefferman, Sanjoy Mitter, and Hariharan Narayanan. "Testing the manifold hypothesis". In: *Journal of the American Mathematical Society* 29.4 (Oct. 2016), pp. 983–1049. ISSN: 0894-0347. DOI: 10.1090/jams/852.

[49] Stanislav Fort, Huiyi Hu, and Balaji Lakshminarayanan. *Deep Ensembles: A Loss Landscape Perspective*. 2020. arXiv: 1912.02757 [stat.ML]. URL: https://arxiv.org/abs/1912.02757.

[50] Maurice Fréchet. "Sur l'extension de certaines evaluations statistiques au cas de petits echantillons". In: *Revue de l'Institut International de Statistique* 11.3 (1943), pp. 182–205. ISSN: 03731138. URL: http://www.jstor.org/stable/1401114.

[51] Benoît Frénay and Ata Kabán. "A comprehensive introduction to label noise". In: *The European Symposium on Artificial Neural Networks (ESANN)*. 2014. URL: https://www.esann.org/sites/default/files/proceedings/legacy/es2014-10.pdf.

[52] Yarin Gal and Zoubin Ghahramani. "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning". In: *Proceedings of The 33rd International Conference on Machine Learning*. Ed. by Maria Florina Balcan and Kilian Q. Weinberger. Vol. 48. Proceedings of Machine Learning Research. New York, New York, USA: PMLR, June 2016, pp. 1050–1059.

[53] Jakob Gawlikowski et al. "A survey of uncertainty in deep neural networks". In: *Artificial Intelligence Review* 56.1 (2023), pp. 1513–1589. DOI: 10.1007/s10462-023-10562-9. URL: https://doi.org/10.1007/s10462-023-10562-9.

[54] James H Gawron, Gregory A Keoleian, Robert D De Kleine, Timothy J Wallington, and Hyung Chul Kim. "Life Cycle Assessment of Connected and Automated Vehicles: Sensing and Computing Subsystem and Vehicle Level Effects". In: *Environmental science and technology* 52 (2018), pp. 3249–3256. DOI: 10.1021/acs.est.7b04576.

[55] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. "Vision meets Robotics: The KITTI Dataset". In: *Inter. Journal of Robotics Research (IJRR)* (2013).

[56] Jakob Geyer et al. "A2D2: Audi Autonomous Driving Dataset". In: *arXiv* (2020). URL: https://www.a2d2.audi.

[57] Steve Gibson, Tsu-Chin Tsao, Michael Roggemann, Timothy Schulz, Allen Tannenbaum, Eric Magee, Matthew Whiteley, Ben Fitzpatrick, Yun Wang, and Mikhail Belenkii. "Atmospheric Propagation of High Energy Lasers: Modeling, Simulation, Tracking, and Control". In: *Final Report of AFOSR Grant F49620-02-01-0319* (Apr. 2008), p. 56. URL: https://www.researchgate.net/publication/235041191_Atmospheric_Propagation_of_High_Energy_Lasers_Modeling_Simulation_Tracking_and_Control.

[58] LaVision GmbH. *Background Oriented Schlieren: BOS Imaging*. Retrieved on March 17th, 2025, from https://www.lavision.de/en/products/flow-bos/bos/.

[59] Tilmann Gneiting and Adrian E Raftery. "Strictly proper scoring rules, prediction, and estimation". In: *Journal of the American statistical Association* 102.477 (2007), pp. 359–378.

[60] Herbert Goldstein, Charles Poole, and John Safko. *Classical mechanics*. 3rd Edition. Addison Wesley, 2000.

[61] Joseph W. Goodman. *Introduction to Fourier optics*. Stanford University, San Francisco: McGraw-Hill, 1968. ISBN: 0-07-023776-X.

[62] David L. Goodstein. *States of matter*. Prentice-Hall physics series. Englewood Cliffs, NJ: Prentice-Hall, 1975.

[63] Google. *HRNet based model trained on the semantic segmentation dataset KITTI*. Retrieved on April 3rd, 2025, from https://www.kaggle.com/models/google/hrnet/tensorFlow2/hrnet-kitti-hrnetv2-w48/.

[64] Hiromasa Goto. "Moiré patterns allow us to visualize the interference between propagating waves". In: *Physics Education* 44.4 (July 2009), p. 338. DOI: 10.1088/0031-9120/44/4/F02. URL: https://dx.doi.org/10.1088/0031-9120/44/4/F02.

[65] IEEE P2020 Working Group. *IEEE P2020 Automotive Imaging White Paper*. Tech. rep. IEEE, 2018. URL: https://ieeexplore.ieee.org/document/8439102.

[66] Volkswagen Group. *Nachhaltigkeitsbericht 2023*. Mar. 2024. URL: https://www.volkswagen-group.com/de/publikationen/weitere/nachhaltigkeitsbericht-2023-2674.

[67] Cornelia Gruber, Patrick Oliver Schenk, Malte Schierholz, Frauke Kreuter, and Göran Kauermann. *Sources of Uncertainty in Supervised Machine Learning – A Statisticians' View*. 2025. arXiv: 2305.16703 [stat.ML]. URL: https://arxiv.org/abs/2305.16703.

[68] Joint Committee for Guides in Metrology (JCGM). "Evaluation of measurement data, Guide to the expression of uncertainty in measurement (GUM)". In: *International Bureau of Weights and Measures (BIPM)* (2008). URL: https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf/cb0ef43f-baa5-11cf-3f85-4dcd86f77bd6.

[69] C. Guillemet. "Annealing and tempering of glass". In: *Journal of Non-Crystalline Solids* 123.1 (1990). XVth International Congress on Glass, pp. 415–426. ISSN: 0022-3093. DOI: https://doi.org/10.1016/0022-3093(90)90813-2. URL: https://www.sciencedirect.com/science/article/pii/0022309390908132.

[70] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. "On calibration of modern neural networks". In: *Proceedings of the 34th International Conference on Machine Learning (ICML)*. ICML'17. PMLR. Sydney, NSW, Australia, 2017, pp. 1321–1330.

[71] Fredrik K. Gustafsson, Martin Danelljan, and Thomas B. Schon. "Evaluating Scalable Bayesian Deep Learning Methods for Robust Computer Vision". In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. June 2020.

[72] Schardin H. and Stamm G. "Prüfung von Flachglas mit Hilfe eines farbigen Schlierenverfahrens". In: *Glastechnische Berichte* 20 (1942).

[73] Karen M. Hampson, Raphaël Turcotte, Donald T. Miller, Kazuhiro Kurokawa, Jared R. Males, Na Ji, and Martin J. Booth. "Adaptive optics for high-resolution imaging". In: *Nature Reviews Methods Primers* 1.1 (2021), p. 68. DOI: 10.1038/s43586-021-00066-7. URL: https://doi.org/10.1038/s43586-021-00066-7.

[74] William Harris. "The matrix representation of dioptric power. Part 1: An introduction". In: *South African Optometrist* 47 (Jan. 1988), pp. 19–23.

[75] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep Residual Learning for Image Recognition". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), pp. 770–778. DOI: 10.1109/CVPR.2016.90.

[76] Eugene Hecht. *Optik*. 7. Auflage. Berlin, Boston: De Gruyter, 2018.

[77] Dan Hendrycks and Thomas Dietterich. "Benchmarking Neural Network Robustness to Common Corruptions and Perturbations". In: *International Conference on Learning Representations (ICLR)*. 2019. URL: https://openreview.net/forum?id=HJz6tiCqYm.

[78] Dan Hendrycks and Kevin Gimpel. "Gaussian Error Linear Units (GELUs)". In: *arXiv* (2023). URL: https://arxiv.org/abs/1606.08415.

[79] G. E. Hinton and R. R. Salakhutdinov. "Reducing the Dimensionality of Data with Neural Networks". In: *Science* 313.5786 (2006), pp. 504–507. DOI: 10.1126/science.1127647.

[80] Eyke Huellermeier and Willem Waegeman. "Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods". In: *Machine Learning* 110 (2021), pp. 457–506. URL: https://doi.org/10.1007/s10994-021-05946-3.

[81] Henry Alexander Ignatious, Hesham-El- Sayed, and Manzoor Khan. "An overview of sensors in Autonomous Vehicles". In: *Procedia Computer Science* 198 (2022). 12th International Conference on Emerging Ubiquitous Systems and Pervasive Networks / 11th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare, pp. 736–741. ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2021.12.315. URL: https://www.sciencedirect.com/science/article/pii/S1877050921025540.

[82] G. Indebetouw and R. Czarnek. "Selected Papers on Optical Moiré and Applications". In: *SPIE Optical Engineering Press* (1992).

[83] British Standards Institution. *ISO 6041. Road Vehicles. Glazing Materials. Method for the Determination of Optical Characteristics of Glazing Camera Zones*. Tech. rep. British Standards Institution, 2023. URL: https://books.google.de/books?id=DuQ30AEACAAJ.

[84] M. J. Irland. *Windshield Optics*. Sept. 1969. DOI: 10.1364/AO.8.001787. URL: https://opg.optica.org/ao/abstract.cfm?URI=ao-8-9-1787.

[85] Mobarakol Islam, Lalithkumar Seenivasan, Hongliang Ren, and Ben Glocker. "Class-Distribution-Aware Calibration for Long-Tailed Visual Recognition". In: *ICML 2021 Workshop on Uncertainty and Robustness in Deep Learning* (2021). eprint: 2109.05263. URL: https://arxiv.org/abs/2109.05263.

[86] ISO12233. "Resolution and spatial frequency responses". In: *International Organization for Standardization* (Feb. 2023).

[87] SAE International Standard J3016-202104. "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles". In: *International Society of Automotive Engineers* (Apr. 2021).

[88] Ajay Jaiswal, Xingguang Zhang, Stanley H. Chan, and Zhangyang Wang. "Physics-Driven Turbulence Image Restoration with Stochastic Refinement". In: *2023 IEEE/CVF International Conference on Computer Vision (ICCV)* (Oct. 2023), pp. 12136–12147. DOI: 10.1109/ICCV51070.2023.01118. URL: https://doi.ieeecomputersociety.org/10.1109/ICCV51070.2023.01118.

[89] Tom Joy, Francesco Pinto, Ser-Nam Lim, Philip H. S. Torr, and Puneet K. Dokania. "Sample-dependent adaptive temperature scaling for improved calibration". In: *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*. AAAI'23/IAAI'23/EAAI'23. AAAI Press, 2023. ISBN: 978-1-57735-880-0. DOI: 10.1609/aaai.v37i12.26742. URL: https://doi.org/10.1609/aaai.v37i12.26742.

[90] Eric J. Jumper, Mike Zenk, Stanislav Gordeyev, David Cavalieri, and Matthew R. Whiteley. "The Airborne Aero-Optics Laboratory, AAOL". In: *Acquisition, Tracking, Pointing, and Laser Systems Technologies XXVI* 8395 (2012). Ed. by William E. Thompson and Paul F. McManamon, p. 839507. DOI: 10.1117/12.922734. URL: https://doi.org/10.1117/12.922734.

[91] George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. "Physics-informed machine learning". In: *Nature Reviews Physics* 3 (2021), pp. 422–440. ISSN: 2522-5820. DOI: https://doi.org/10.1038/s42254-021-00314-5.

[92] Alex Kendall and Yarin Gal. "What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?" In: *Advances in Neural Information Processing Systems (NeurIPS)*. Ed. by I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett. Vol. 30. Curran Associates, Inc., 2017. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/2650d6089a6d640c5e85b2b88265dc2b-Paper.pdf.

[93] Nitish Shirish Keskar, Dheevatsa Mudigere, Jorge Nocedal, Mikhail Smelyanskiy, and Ping Tak Peter Tang. "On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima". In: *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings* (2017).

[94] Cassius J. Keyser. "Mathematical Philosophy - A Study of Fate and Freedom - Lectures for Educated Laymen". In: *University of Michigan Historical Math Collection* (1922). URL: https://name.umdl.umich.edu/ACA0682.0001.001.

[95] Taehyeon Kim and Se-Young Yun. "Revisiting Orthogonality Regularization: A Study for Convolutional Neural Networks in Image Classification". In: *IEEE Access* 10 (2022), pp. 69741–69749.

[96] Diederik P. Kingma and Max Welling. "Auto-Encoding Variational Bayes". In: *CoRR* abs/1312.6114 (2013). URL: https://api.semanticscholar.org/CorpusID:216078090.

[97] John Kirchenbauer, Jacob Oaks, and Eric Heim. "What is Your Metric Telling You? Evaluating Classifier Calibration under Context-Specific Definitions of Reliability". In: *ICLR 2022 Workshop on ML Evaluation Standards* (2022). eprint: 2205.11454. URL: https://arxiv.org/abs/2205.11454.

[98] Klinghoffer, Tzofi and Somasundaram, Siddharth and Tiwary, Kushagra and Raskar, Ramesh. *Physics vs. Learned Priors: Rethinking Camera and Algorithm Design for Task-Specific Imaging*. Los Alamitos, CA, USA, Aug. 2022. DOI: 10.1109/ICCP54855.2022.9887681. URL: https://doi.ieeecomputersociety.org/10.1109/ICCP54855.2022.9887681.

[99] Steven G. Krantz. *A Panaroma of Harmonic Analysis*. Mathematical Association of America, 1999. ISBN: 978-1-614-44026-0. DOI: https://doi.org/10.5948/UPO9781614440260.

[100] Christian Krebs, Patrick Müller, and Alexander Braun. "Impact of Windshield Optical Aberrations on Visual Range Camera Based Classification Tasks Performed by CNNs". In: *London Imaging Meeting* (2021). DOI: 10.2352/issn.2694-118X.2021.LIM-83. URL: https://library.imaging.org/lim/articles/2/1/art00016.

[101] Dilip Krishnan and Rob Fergus. "Fast Image Deconvolution using Hyper-Laplacian Priors". In: *Advances in Neural Information Processing Systems (NeurIPS)* 22 (2009), pp. 1033–1041.

[102] Gerhard Krumpl, Henning Avenhaus, Horst Possegger, and Horst Bischof. "ATS: Adaptive Temperature Scaling for Enhancing Out-of-Distribution Detection Methods". In: *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 2024, pp. 3852–3861. DOI: 10.1109/WACV57701.2024.00382.

[103] Horst Kuchling. *Taschenbuch der Physik*. 19., aktualisierte Auflage. Carl Hanser Verlag, 2007.

[104] Meelis Kull, Miquel Perello Nieto, Markus Kängsepp, Telmo Silva Filho, Hao Song, and Peter Flach. "Beyond temperature scaling: Obtaining well-calibrated multi-class probabilities with dirichlet calibration". In: *Advances in Neural Information Processing Systems (NeurIPS)* 32 (2019).

[105] Kuraray Europe GmbH. *Manual TROSIFOL*. Nov. 2007.

[106] *LABSCAN-SCREEN, Fortschrittliches Offline-System für hochauflösende Messergebnisse*. ISRA VISION GmbH. URL: https://www.isravision.com/de/flat-panel-display-glas/anwendungen/floatglas-fuer-duennglas/labscan-screen/.

[107] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. "Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles". In: *Advances in Neural Information Processing Systems (NeurIPS)* 30 (2017). URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/9ef2ed4b7fd2c810847ffa5fa85bce38-Paper.pdf.

[108] Steven Landgraf, Markus Hillemann, Theodor Kapler, and Markus Ulrich. "Efficient multi-task uncertainties for joint semantic segmentation and monocular depth estimation". In: *German Conference on Pattern Recognition (GCPR)* (2024).

[109] Steven Landgraf, Kira Wursthorn, Markus Hillemann, and Markus Ulrich. "DUDES: Deep Uncertainty Distillation using Ensembles for Semantic Segmentation". In: *PFG –Journal of Photogrammetry, Remote Sensing and Geoinformation Science* 92.2 (2024), pp. 101–114. DOI: 10.1007/s41064-024-00280-4. URL: https://doi.org/10.1007/s41064-024-00280-4.

[110] Max-Heinrich Laves, Sontje Ihler, Karl-Philipp Kortmann, and Tobias Ortmaier. "Well-calibrated model uncertainty with temperature scaling for dropout variational inference". In: *NeurIPS 2019 Workshop on Bayesian Deep Learning* (2019). URL: https://bayesiandeeplearning.org/2019/papers/77.pdf.

[111] Zonghui Li, Yongsheng Dong, Longchao Shen, Yafeng Liu, Yuanhua Pei, Haotian Yang, Lintao Zheng, and Jinwen Ma. "Development and challenges of object detection: A survey". In: *Neurocomputing* 598 (2024), p. 128102. ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2024.128102. URL: https://www.sciencedirect.com/science/article/pii/S0925231224008737.

[112] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar. "Focal Loss for Dense Object Detection". In: *IEEE International Conference on Computer Vision (ICCV)* (2017), pp. 2999–3007. DOI: 10.1109/ICCV.2017.324. URL: https://doi.ieeecomputersociety.org/10.1109/ICCV.2017.324.

[113] J. David Logan. *Applied Mathematics*. John Wiley & Sons, 2013. ISBN: 978-1-118-47580-5.

[114] Bjoern Ludwig. "GUM-compliant neural network robustness verification". In: *Master thesis at the Technical University of Berlin, Zuse Institute Berlin and Physikalisch-Technische Bundesanstalt* (2023).

[115] Thomas Luhmann. "Erweiterte Verfahren zur geometrischen Kamerakalibrierung in der Nahbereichsphotogrammetrie". German. PhD thesis. Technische Universität Dresden, 2010. ISBN: 978-3-7696-5057-0.

[116] Scott M. Lundberg and Su-In Lee. "A Unified Approach to Interpreting Model Predictions". In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NIPS'17. Long Beach, California, USA: Curran Associates Inc., 2017, pp. 4768–4777. ISBN: 9781510860964.

[117] Kira Maag, Matthias Rottmann, and Hanno Gottschalk. "Time-dynamic estimates of the reliability of deep semantic segmentation networks". In: *IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI)* (2020), pp. 502–509.

[118] David J. C. MacKay. "A Practical Bayesian Framework for Backpropagation Networks". In: *Neural Computation* 4.3 (May 1992), pp. 448–472. ISSN: 0899-7667. DOI: 10.1162/neco.1992.4.3.448. URL: https://doi.org/10.1162/neco.1992.4.3.448.

[119] Virendra N. Mahajan. *Optical Imaging and Aberrations: Part I. Ray Geometrical Optics*. SPIE Press, 1998. DOI: 10.1117/3.265735. URL: https://doi.org/10.1117/3.265735.

[120] Justin Marshall and Johannes Oberwinkler. "The colourful world of the mantis shrimp". In: *Nature* 401.6756 (1999), pp. 873–874. ISSN: 1476-4687. DOI: https://doi.org/10.1038/44751.

[121] Patrick C. McGuire, David G. Sandler, Michael Lloyd-Hart, and Troy A. Rhoadarmer. "Adaptive optics: Neural network wavefront sensing, reconstruction, and prediction". In: *Scientific Applications of Neural Nets*. Ed. by John W. Clark, Thomas Lindenau, and Manfred L. Ristig. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 97–138. ISBN: 978-3-540-48980-1.

[122] Pankaj Mehta, Marin Bukov, Ching-Hao Wang, Alexandre G.R. Day, Clint Richardson, Charles K. Fisher, and David J. Schwab. "A high-bias, low-variance introduction to Machine Learning for physicists". In: *Physics Reports* 810 (2019). A high-bias, low-variance introduction to Machine Learning for physicists, pp. 1–124. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2019.03.001. URL: https://www.sciencedirect.com/science/article/pii/S0370157319300766.

[123] Marina Meila and Hanyu Zhang. "Manifold Learning: What, How, and Why". In: *Annual Review of Statistics and Its Application* 11.Volume 11, 2024 (2024), pp. 393–417. ISSN: 2326-831X. DOI: https://doi.org/10.1146/annurev-statistics-040522-115238. URL: https://www.annualreviews.org/content/journals/10.1146/annurev-statistics-040522-115238.

[124] Thomas Mensink, Jasper Uijlings, Alina Kuznetsova, Michael Gygli, and Vittorio Ferrari. "Factors of Influence for Transfer Learning Across Diverse Appearance Domains and Task Types". In: *IEEE Transactions on Pattern Analysis & Machine Intelligence* 44.12 (Dec. 2022), pp. 9298–9314. ISSN: 1939-3539. DOI: 10.1109/TPAMI.2021.3129870. URL: https://doi.ieeecomputersociety.org/10.1109/TPAMI.2021.3129870.

[125] Fritz Merkle and Norbert Hubin. "Adaptive Optics for the ESO Very Large Telescope". In: *Adaptive Optics for Large Telescopes*. Optica Publishing Group, 1992. DOI: 10.1364/AOLT.1992.ATuA4. URL: https://opg.optica.org/abstract.cfm?URI=AOLT-1992-ATuA4.

[126] Shervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. "Image Segmentation Using Deep Learning: A Survey". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.7 (2022), pp. 3523–3542. DOI: 10.1109/TPAMI.2021.3059968.

[127] Christoph Molnar, Gunnar König, Julia Herbinger, Timo Freiesleben, Susanne Dandl, Christian A. Scholbeck, Giuseppe Casalicchio, Moritz Grosse-Wentrup, and Bernd Bischl. "General Pitfalls of Model-Agnostic Interpretation Methods for Machine Learning Models". In: *xxAI - Beyond Explainable AI: International Workshop, Held in Conjunction with ICML 2020, July 18, 2020, Vienna, Austria, Revised and Extended Papers*. Springer International Publishing, 2022, pp. 39–68. ISBN: 978-3-031-04083-2. DOI: 10.1007/978-3-031-04083-2_4. URL: https://doi.org/10.1007/978-3-031-04083-2_4.

[128] Bálint Mucsányi, Michael Kirchhof, and Seong Joon Oh. "Benchmarking Uncertainty Disentanglement: Specialized Uncertainties for Specialized Tasks". In: *Advances in Neural Information Processing Systems (NeurIPS)*. Vol. 37. Curran Associates, Inc., 2024, pp. 50972–51038. URL: https://proceedings.neurips.cc/paper_files/paper/2024/file/5afa9cb1e917b898ad418216dc726fbd-Paper-Datasets_and_Benchmarks_Track.pdf.

[129] Patrick Mueller and Alexander Braun. "MTF as a performance indicator for AI algorithms?" In: *Electronic Imaging* 35.16 (2023), pp. 125-1–125-1. DOI: 10.2352/EI.2023.35.16.AVM-125. URL: https://library.imaging.org/ei/articles/35/16/AVM-125.

[130] Jishnu Mukhoti, Viveka Kulharia, Amartya Sanyal, Stuart Golodetz, Philip Torr, and Puneet Dokania. "Calibrating deep neural networks using focal loss". In: *Advances in Neural Information Processing Systems (NeurIPS)* 33 (2020), pp. 15288–15299.

[131] I. Müller and W. Weiss. *Entropy and Energy: A Universal Competition*. Interaction of Mechanics and Mathematics. Springer Berlin, Heidelberg, 2006. ISBN: 9783540323808.

[132] Patrick Müller and Alexander Braun. "Simulating optical properties to access novel metrological parameter ranges and the impact of different model approximations". In: *2022 IEEE International Workshop on Metrology for Automotive (MetroAutomotive)*. 2022, pp. 133–138. DOI: 10.1109/MetroAutomotive54295.2022.9855079.

[133] Patrick Müller, Alexander Braun, and Margret Keuper. "Impact of realistic properties of the point spread function on classification tasks to reveal a possible distribution shift". In: *NeurIPS 2022 Workshop on DistShift* (2022).

[134] Rafael Müller, Simon Kornblith, and Geoffrey E Hinton. "When does label smoothing help?" In: *Advances in Neural Information Processing Systems (NeurIPS)* 32 (2019).

[135] Sai Munikoti, Deepesh Agarwal, Laya Das, and Balasubramaniam Natarajan. "A general framework for quantifying aleatoric and epistemic uncertainty in graph neural networks". In: *Neurocomputing* 521 (Feb. 2023), pp. 1–10. ISSN: 0925-2312. DOI: 10.1016/j.neucom.2022.11.049. URL: https://doi.org/10.1016/j.neucom.2022.11.049.

[136] Vinod Nair and Geoffrey E. Hinton. "Rectified linear units improve restricted boltzmann machines". In: *Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML)* (2010), pp. 807–814.

[137] Jeremy Nixon, Michael W. Dusenberry, Linchuan Zhang, Ghassen Jerfel, and Dustin Tran. "Measuring Calibration in Deep Learning". In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. June 2019, pp. 38–41.

[138] Yaniv Ovadia, Emily Fertig, Jie Ren, Zachary Nado, D Sculley, Sebastian Nowozin, Joshua Dillon, Balaji Lakshminarayanan, and Jasper Snoek. *Can You Trust Your Model's Uncertainty? Evaluating Predictive Uncertainty Under Dataset Shift*. Advances in Neural Information Processing Systems (NeurIPS). 2019. URL: https://papers.nips.cc/paper/2019/hash/8558cb408c1d76621371888657d2eb1d-Abstract.html.

[139] Mahdi Pakdaman Naeini, Gregory Cooper, and Milos Hauskrecht. "Obtaining Well Calibrated Probabilities Using Bayesian Binning". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 29.1 (Feb. 2015). DOI: 10.1609/aaai.v29i1.9602. URL: https://ojs.aaai.org/index.php/AAAI/article/view/9602.

[140] Bernd Pesch. *Bestimmung der Messunsicherheit nach GUM, Grundlagen der Metrologie*. Books on Demand, 2003. ISBN: 3-8330-1039-8.

[141] B. C. Platt and R. Shack. "History and principles of Shack-Hartmann wavefront sensing". In: *Journal of refractive surgery* 17 (5 Sept. 2001), pp. 1033–1041. DOI: 10.3928/1081-597X-20010901-13.

[142] John Platt et al. "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods". In: *Advances in large margin classifiers* 10.3 (1999), pp. 61–74.

[143] Serban Rares Pop. *Modeling and simulation of the float glass process*. Technische Universität Kaiserslautern, 2005. URL: https://nbn-resolving.de/urn:nbn:de:hbz:386-kluedo-18823.

[144] Markus Raffel. "Background-oriented schlieren (BOS) techniques". In: *Experiments in Fluids* 56 (3 Mar. 2015), p. 60. ISSN: 1432-1114. DOI: 10.1007/s00348-015-1927-5. URL: http://link.springer.com/10.1007/s00348-015-1927-5.

[145] M. Raissi, P. Perdikaris, and G.E. Karniadakis. "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations". In: *Journal of Computational Physics* 378 (2019), pp. 686–707. ISSN: 0021-9991. DOI: https://doi.org/10.1016/j.jcp.2018.10.045.

[146] C.R. Rao. "Information and the Accuracy Attainable in the Estimation of Statistical Parameters". In: *Bulletin of Calcutta Mathematical Society* 37 (1945), pp. 81–91.

[147] Samik Raychaudhuri. "Introduction to Monte Carlo simulation". In: *Winter Simulation Conference* (2008), pp. 91–100. DOI: 10.1109/WSC.2008.4736059.

[148] *Regulation No. 43 of the Economic Commission for Europe of the United Nations (UN/ECE) - Uniform provisions concerning the approval of safety glazing materials and their installation on vehicles*. European Union. Feb. 2014, pp. 1–126. URL: http://data.europa.eu/eli/reg/2014/43(2)/oj.

[149] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (2015). Ed. by Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, pp. 234–241.

[150] Benedek Rozemberczki, Lauren Watson, Peter Bayer, Hao-Tsung Yang, Oliver Kiss, Sebastian Nilsson, and Rik Sarkar. "The Shapley Value in Machine Learning". In: *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*. Ed. by Lud De Raedt. Survey Track. International Joint Conferences on Artificial Intelligence Organization, July 2022, pp. 5572–5579. DOI: 10.24963/ijcai.2022/778. URL: https://doi.org/10.24963/ijcai.2022/778.

[151] A. Rudolph, J. Krois, and K. Hartmann. *Statistics and Geodata Analysis using Python (SOGA-Py), Department of Earth Sciences, Freie Universitaet Berlin*. 2023. URL: https://www.geo.fu-berlin.de/en/v/soga-py/Basics-of-statistics/Hypothesis-Tests/Introduction-to-Hypothesis-Testing/Critical-Value-and-the-p-Value-Approach/index.html (visited on 11/28/2024).

[152] C. E. Shannon. "A mathematical theory of communication". In: *The Bell System Technical Journal* 27.3 (1948), pp. 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

[153] Lloyd S. Shapley. *A Value for N-Person Games*. RAND Corporation, Santa Monica, CA. URL: https://www.rand.org/pubs/papers/P295.html. 1952.

[154] L. Shartsis and S. Spinner. "Viscosity and density of molten optical glasses". In: *Journal of Research of the National Bureau of Standards* 46.3 (1951). DOI: https://doi.org/10.1016/0022-3093(90)90813-2.

[155] H Shi, M Drton, and F Han. "On the power of Chatterjee's rank correlation". In: *Biometrika* 109.2 (Oct. 2021), pp. 317–333. ISSN: 1464-3510. DOI: 10.1093/biomet/asab028.

[156] Connor Shorten and Taghi M. Khoshgoftaar. "A survey on Image Data Augmentation for Deep Learning". In: *Journal of Big Data* 6.60 (2019). DOI: 10.1186/s40537-019-0197-0. URL: https://doi.org/10.1186/s40537-019-0197-0.

[157] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *Journal of Machine Learning Research* 15.56 (2014), pp. 1929–1958. URL: http://jmlr.org/papers/v15/srivastava14a.html.

[158] Carsten Steger, Markus Ulrich, and Christian Wiedemann. *Machine Vision Algorithms and Applications*. 2nd. Weinheim: Wiley-VCH, 2018. ISBN: 9783527413652.

[159] Marco Taboga. *Ridge regression, Lectures on probability theory and mathematical statistics. Kindle Direct Publishing*. 2021. URL: https://www.statlect.com/fundamentals-of-statistics/ridge-regression (visited on 11/29/2024).

[160] Gerhard Teicher. *Glas im Automobil II*. Vol. 23. Fachbuch, 23. Renningen: Expert Verlag, 2003. ISBN: 3816922155. URL: https://www.tib.eu/de/suchen/id/TIBKAT%3A365276456.

[161] Marvin Tom Teichmann, Manasi Datar, Lisa Kratzke, Fernando Vega, and Florin C. Ghesu. "Towards Integrating Epistemic Uncertainty Estimation into the Radiotherapy Workflow". In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024* (2024), pp. 729–738.

[162] C. M. Theobald. "Generalizations of Mean Square Error Applied to Ridge Regression". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 36.1 (1974), pp. 103–106. ISSN: 00359246. URL: http://www.jstor.org/stable/2984775.

[163] Larry N Thibos. "Calculation of the geometrical point-spread function from wavefront aberrations". In: *Ophthalmic and Physiological Optics* 39.4 (2019), pp. 232–244. DOI: https://doi.org/10.1111/opo.12619. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/opo.12619. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/opo.12619.

[164] Larry N. Thibos, Xin Hong, Arthur Bradley, and Raymond A Applegate. "Accuracy and precision of objective refraction from wavefront aberrations". In: *Journal of Vision* 4.4 (Apr. 2004), pp. 329–51. ISSN: 1534-7362. DOI: 10.1167/4.4.9. URL: https://doi.org/10.1167/4.4.9.

[165] Hanne H Thoen, Martin J How, Tsyr-Huei Chiou, and Justin Marshall. "A different form of color vision in mantis shrimp". In: *Science* 343.6169 (2014), pp. 873–874. ISSN: 1095-9203. DOI: https://doi.org/10.1126/science.1245824.

[166] LiFeng Tian, ShiHe Yi, YuXin Zhao, Lin He, and Zhi Chen. "Aero-optical wavefront measurement technique based on BOS and its applications". In: *Chinese Science Bulletin* 56 (22 Aug. 2011), pp. 2320–2326. ISSN: 1001-6538. DOI: 10.1007/s11434-011-4583-z. URL: http://link.springer.com/10.1007/s11434-011-4583-z.

[167] Christian Tomani, Daniel Cremers, and Florian Buettner. "Parameterized Temperature Scaling for Boosting the Expressive Power in Post-Hoc Uncertainty Calibration". In: *European Conference on Computer Vision (ECCV)* (2022), pp. 555–569.

[168] Ethan Tseng, Ali Mosleh, Fahim Mannan, Karl St-Arnaud, Avinash Sharma, Yifan Peng, Alexander Braun, Derek Nowrouzezahrai, Jean-François Lalonde, and Felix Heide. "Differentiable Compound Optics and Processing Pipeline Optimization for End-to-end Camera Design". In: *ACM Transactions on Graphics* 40.2 (June 2021). ISSN: 0730-0301. DOI: 10.1145/3446791. URL: https://doi.org/10.1145/3446791.

[169] G. Upton and I. Cook. *A Dictionary of Statistics*. Oxford Paperback Reference. OUP Oxford, 2008. ISBN: 9780199541454. URL: https://books.google.de/books?id=u97pzxRjaCQC.

[170] Juozas Vaicenavicius, David Widmann, Carl Andersson, Fredrik Lindsten, Jacob Roll, and Thomas Schön. "Evaluating model calibration in classification". In: *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*. Vol. 89. Proceedings of Machine Learning Research. PMLR, Apr. 2019, pp. 3459–3467. URL: https://proceedings.mlr.press/v89/vaicenavicius19a.html.

[171] Andreas Veit, Michael Wilber, and Serge Belongie. "Residual networks behave like ensembles of relatively shallow networks". In: *Proceedings of the 30th International Conference on Neural Information Processing Systems* (2016), pp. 550–558.

[172] L. Venkatakrishnan and G. E. A. Meier. "Density measurements using the Background Oriented Schlieren technique". In: *Experiments in Fluids* 37.2 (2004), pp. 237–247. ISSN: 1432-1114. DOI: 10.1007/s00348-004-0807-1. URL: https://doi.org/10.1007/s00348-004-0807-1.

[173] Frank Verbiest, Marc Proesmans, and Luc Van Gool. "Modeling the Effects of Windshield Refraction for Camera Calibration". In: *ECCV* (2020).

[174] Jason P. de Villiers, F. Wilhelm Leuschner, and Ronelle Geldenhuys. "Modeling of radial asymmetry in lens distortion facilitated by modern optimization techniques". In: *Intelligent Robots and Computer Vision XXVII: Algorithms and Techniques*. Ed. by David P. Casasent, Ernest L. Hall, and Juha Röning. Vol. 7539. International Society for Optics and Photonics. SPIE, 2010, 75390J. DOI: 10.1117/12.838804. URL: https://doi.org/10.1117/12.838804.

[175] David D. Walden, Thomas M. Shortell, Garry J. Roedler, Bernardo A. Delicado, Odile Mornas, Yip Yew-Seng, and David Endler. *INCOSE Systems Engineering Handbook*. 5th Edition. Wiley, June 2023, p. 368. ISBN: 978-1-119-81431-3. URL: https://www.wiley.com/en-ie/INCOSE+Systems+Engineering+Handbook%2C+5th+Edition-p-9781119814313.

[176] Cheng Wang and Jacek Golebiowski. "Towards Unbiased Calibration using Meta-Regularization". In: *Transactions on Machine Learning Research* (2024). ISSN: 2835-8856. URL: https://www.amazon.science/publications/towards-unbiased-calibration-using-meta-regularization.

[177] Guotai Wang, Wenqi Li, Michael Aertsen, Jan Deprest, Sebastien Ourselin, and Tom Vercauteren. "Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks". In: *Neurocomputing* 338 (2019), pp. 34–45. ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2019.01.103. URL: https://www.sciencedirect.com/science/article/pii/S0925231219301961.

[178] Hanjing Wang and Qiang Ji. "Epistemic Uncertainty Quantification for Pretrained Neural Networks". In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, June 2024, pp. 11052–11061. DOI: 10.1109/CVPR52733.2024.01051. URL: https://doi.ieeecomputersociety.org/10.1109/CVPR52733.2024.01051.

[179] Jingdong Wang et al. "Deep High-Resolution Representation Learning for Visual Recognition". In: *IEEE Transactions on Pattern Analysis & Machine Intelligence* 43.10 (Oct. 2021), pp. 3349–3364. ISSN: 1939-3539. DOI: 10.1109/TPAMI.2020.2983686. URL: https://doi.ieeecomputersociety.org/10.1109/TPAMI.2020.2983686.

[180] Wenhai Wang et al. "InternImage: Exploring Large-Scale Vision Foundation Models with Deformable Convolutions". In: *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2023, pp. 14408–14419. DOI: 10.1109/CVPR52729.2023.01385.

[181] Ansgar Wego. "Elektroauto vs. Verbrennerfahrzeug im einjaehrigen Alltagsbetrieb". In: *Magazine of the University of Applied Sciences Wismar* (Aug. 2022). URL: https://fiw.hs-wismar.de/storages/hs-wismar/_FIW/Bereich_EUI/AB_AET/Professoren/Wego/Elektroautos_vs._Verbrennerfahrzeug_im_einjaehrigen_Alltagsbetrieb_-_2022.pdf.

[182] Karl Weierstrass. "Über die analytische Darstellbarkeit sogenannter willkürlicher Functionen einer reellen Veränderlichen". In: *Sitzungsberichte der Königlich-Preussischen Akademie der Wissenschaften zu Berlin* (1885).

[183] N.A. Weiss, P.T. Holmes, and M. Hardy. *A Course in Probability*. Pearson Addison Wesley, 2005, pp. 385–386. ISBN: 9780321189547.

[184] Tobias Werner. "Simulation, Aufbau und Charakterisierung von autostereoskopischen Display-Systemen im Fahrzeugbereich". German. PhD thesis. Karlsruher Institut für Technologie (KIT), 2017. 264 pp. ISBN: 978-3-7315-0617-1. DOI: 10.5445/KSP/1000063916.

[185] Gordon Wetzstein, Aydogan Ozcan, Sylvain Gigan, Shanhui Fan, Dirk Englund, Marin Soljačić, Cornelia Denz, David A. B. Miller, and Demetri Psaltis. "Inference in artificial intelligence with deep optics and photonics". In: *Nature* 588 (2020), pp. 39–47. ISSN: 1476-4687. DOI: https://doi.org/10.1038/s41586-020-2973-6.

[186] Folker Wientapper, Harald Wuest, Pavel Rojtberg, and Dieter Fellner. "A camera-based calibration for automotive augmented reality Head-Up-Displays". In: *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 2013, pp. 189–197. DOI: 10.1109/ISMAR.2013.6671779.

[187] A.M. Wildberger. "Alleviating the opacity of neural networks". In: *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)*. Vol. 4. 1994, 2373–2376 vol.4. DOI: 10.1109/ICNN.1994.374590.

[188] Lisa Wimmer, Yusuf Sale, Paul Hofman, Bernd Bischl, and Eyke Hüllermeier. "Quantifying aleatoric and epistemic uncertainty in machine learning: Are conditional entropy and mutual information appropriate measures?" In: *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*. Vol. 216. Proceedings of Machine Learning Research. PMLR, July 2023, pp. 2282–2292. URL: https://proceedings.mlr.press/v216/wimmer23a.html.

[189] Dominik Werner Wolf, Prasannavenkatesh Balaji, Alexander Braun, and Markus Ulrich. "Decoupling of neural network calibration measures". In: *German Conference on Pattern Recognition (GCPR)* (2024).

[190] Dominik Werner Wolf, Alexander Braun, and Markus Ulrich. *Optical aberrations in autonomous driving: Physics-informed parameterized temperature scaling for neural network uncertainty calibration*. (Under review). 2024. arXiv: 2412.13695 [cs.CV]. URL: https://arxiv.org/abs/2412.13695.

[191] Dominik Werner Wolf, Boris Thielbeer, Markus Ulrich, and Alexander Braun. "Wavefront aberration measurements based on the Background Oriented Schlieren method". In: *Measurement: Sensors* (2024), p. 101509. ISSN: 2665-9174. DOI: https://doi.org/10.1016/j.measen.2024.101509. URL: https://www.sciencedirect.com/science/article/pii/S2665917424004859.

[192] Dominik Werner Wolf, Markus Ulrich, and Alexander Braun. "Novel developments of refractive power measurement techniques in the automotive world". In: *Metrologia* 60 (6 Sept. 2023). ISSN: 0026-1394. DOI: 10.1088/1681-7575/acf1a4. URL: https://iopscience.iop.org/article/10.1088/1681-7575/acf1a4.

[193] Dominik Werner Wolf, Markus Ulrich, and Alexander Braun. "Windscreen Optical Quality for AI Algorithms: Refractive Power and MTF not Sufficient". In: *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)* (Sept. 2023), pp. 5190–5197. DOI: 10.1109/ITSC57777.2023.10421970. URL: https://ieeexplore.ieee.org/document/10421970/.

[194] Dominik Werner Wolf, Markus Ulrich, and Nikhil Kapoor. "Sensitivity analysis of AI-based algorithms for autonomous driving on optical wavefront aberrations induced by the windshield". In: *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)* (Oct. 2023), pp. 4102–4111. DOI: 10.1109/ICCVW60793.2023.00443. URL: https://ieeexplore.ieee.org/document/10350923/.

[195] Jeffrey Wong. *Lecture Notes on PDEs, part II: Laplace's equation, the wave equation and more*. Tech. rep. Duke University, 2018. URL: https://services.math.duke.edu/~jtwong/math353-2018/lectures/Notes-PDEs2.pdf.

[196] K. Wursthorn, M. Hillemann, and M. Ulrich. "Uncertainty Quantification with Deep Ensembles for 6D Object Pose Estimation". In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* X-2-2024 (2024), pp. 223–230. DOI: 10.5194/isprs-annals-X-2-2024-223-2024. URL: https://isprs-annals.copernicus.org/articles/X-2-2024/223/2024/.

[197] Johnathan Xie, Annie S Chen, Yoonho Lee, Eric Mitchell, and Chelsea Finn. "Calibrating Language Models with Adaptive Temperature Scaling". In: *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*. Ed. by Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen. Miami, Florida, USA: Association for Computational Linguistics, Nov. 2024, pp. 18128–18138. URL: https://aclanthology.org/2024.emnlp-main.1007.

[198] Xinge Yang, Qiang Fu, Yunfeng Nie, and Wolfgang Heidrich. *Image Quality Is Not All You Want: Task-Driven Lens Design for Image Classification*. May 2023. DOI: 10.48550/arXiv.2305.17185. URL: https://doi.org/10.48550/arXiv.2305.17185.

[199] Ai Ping Yow, Damon Wong, Yueqian Zhang, Christoph Menke, Ralf Wolleschensky, and Peter Török. "Artificial intelligence in optical lens design". In: *Artificial Intelligence Review* 57.8 (2024). ISSN: 1573-7462. DOI: https://doi.org/10.1007/s10462-024-10842-y.

[200] Jiachuan Yu, Zhisheng Zhang, Han Sun, Zhijie Xia, and Haiying Wen. "Reevaluating the Underlying Radial Symmetry Assumption of Camera Distortion". In: *IEEE Transactions on Instrumentation and Measurement* 73 (2024), pp. 1–10. DOI: 10.1109/TIM.2024.3451594.

[201] Yuhui Yuan, Xilin Chen, and Jingdong Wang. "Object-Contextual Representations for Semantic Segmentation". In: *Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI*. Springer-Verlag, 2020, pp. 173–190. ISBN: 978-3-030-58538-9. DOI: 10.1007/978-3-030-58539-6_11. URL: https://doi.org/10.1007/978-3-030-58539-6_11.

[202] Alireza Zaeemzadeh, Nazanin Rahnavard, and Mubarak Shah. "Norm-Preservation: Why Residual Networks Can Become Extremely Deep?" In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43 (2018), pp. 3980–3990. URL: https://api.semanticscholar.org/CorpusID:29155803.

[203] M. Zare, S. M. Mousavi, A. Mohseni, and A. H. Nikjoo. "On the Application of Endoscopic Background-Oriented Schlieren Technique for Near-Field Investigation of Density Field". In: *Journal of Applied Fluid Mechanics* 13.2 (2020), pp. 667–678. ISSN: 1735-3572. DOI: 10.29252/jafm.13.02.30408.

[204] F. Zernike and F.J.M. Stratton. "Diffraction Theory of the Knife-Edge Test and its Improved Form, The Phase-Contrast Method". In: *Monthly Notices of the Royal Astronomical Society* 94.5 (Mar. 1934), pp. 377–384. ISSN: 0035-8711. DOI: 10.1093/mnras/94.5.377. eprint: https://academic.oup.com/mnras/article-pdf/94/5/377/3785471/mnras94-0377.pdf. URL: https://doi.org/10.1093/mnras/94.5.377.

[205] Frits Zernike. "Beugungstheorie des schneidenver-fahrens und seiner verbesserten form, der phasenkontrastmethode". In: *Physica* 1.7 (1934), pp. 689–704. ISSN: 0031-8914. DOI: https://doi.org/10.1016/S0031-8914(34)80259-5. URL: https://www.sciencedirect.com/science/article/pii/S0031891434802595.

[206] Nan Zhan, Xiaochuan Zhang, Jingchen Ye, Tingting Wang, Ziwen Dong, and Zhenxing Song. "High-precision parameter-free optical distortion measurement and correction via digital image correlation". In: *Optics & Laser Technology* 177 (2024), p. 111129. ISSN: 0030-3992. DOI: https://doi.org/10.1016/j.optlastec.2024.111129. URL: https://www.sciencedirect.com/science/article/pii/S0030399224005875.

[207] Jize Zhang, Bhavya Kailkhura, and T. Yong-Jin Han. "Mix-n-Match: ensemble and compositional methods for uncertainty calibration in deep learning". In: *Proceedings of the 37th International Conference on Machine Learning*. ICML'20. JMLR.org, 2020.

[208] Xiong Zhang, Hongmin Xu, Hong Mo, Jianchao Tan, Cheng Yang, Lei Wang, and Wenqi Ren. "DCNAS: Densely Connected Neural Architecture Search for Semantic Image Segmentation". In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2021, pp. 13951–13962. DOI: 10.1109/CVPR46437.2021.01374.

[209] Yuxiao Zhang, Alexander Carballo, Hanting Yang, and Kazuya Takeda. "Perception and sensing for autonomous vehicles under adverse weather conditions: A survey". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 196 (2023), pp. 146–177. ISSN: 0924-2716. DOI: https://doi.org/10.1016/j.isprsjprs.2022.12.021. URL: https://www.sciencedirect.com/science/article/pii/S0924271622003367.

[210] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. "Loss Functions for Image Restoration With Neural Networks". In: *IEEE Transactions on Computational Imaging* 3.1 (2017), pp. 47–57. DOI: 10.1109/TCI.2016.2644865.

List of Figures

List of Tables

Colophon

This thesis was typeset with LaTeX 2ε. It uses the *Clean Thesis* style developed by Ricardo Langner. The design of the *Clean Thesis* style is inspired by user guide documents from Apple Inc.

Download the *Clean Thesis* style at http://cleanthesis.der-ric.de/.