

# Deep Multimodality Image-Guided System for Assisting Neurosurgery

This dissertation is submitted for the degree of

Doctor of Engineering

Faculty of Computer Science  
Karlsruhe Institute of Technology (KIT)

approved

Dissertation

by

M.Sc. Ramy Ashraf Salaheldin Zeineldin

Born in Nancy, France

Day of the oral examination: February 3, 2023

First Supervisor:

Prof. Dr. Franziska Mathis-Ullrich

Second Supervisor:

Prof. Dr.-Ing. Oliver Burgert

Third Supervisor:

Prof. Dr. med. Christian Rainer Wirtz



This document is licensed under a Creative Commons  
Attribution-Non Commercial 4.0 International License (CC BY-NC 4.0):  
<https://creativecommons.org/licenses/by-nc/4.0/deed.en>

# Deep Multimodality Image-Guided System for Assisting Neurosurgery

Zur Erlangung des akademischen Grades eines  
Doktors der Ingenieurwissenschaften  
von der KIT-Fakultät für Informatik  
des Karlsruher Instituts für Technologie (KIT)

genehmigte

Dissertation

von

M.Sc. Ramy Ashraf Salaheldin Zeineldin  
Geboren in Nancy, Frankreich

Tag der mündlichen Prüfung: 3. Februar 2023

Erster Gutachterin:

Prof. Dr. Franziska Mathis-Ullrich

Zweiter Gutachter:

Prof. Dr.-Ing. Oliver Burgert

Dritter Gutachter:

Prof. Dr. med. Christian Rainer Wirtz



Dieses Werk ist lizenziert unter einer Creative Commons  
Namensnennung-Nicht kommerziell 4.0 International Lizenz (CC BY-NC 4.0):  
<https://creativecommons.org/licenses/by-nc/4.0/deed.de>

# Kurzfassung

Intrakranielle Hirntumoren gehören zu den zehn häufigsten bösartigen Krebsarten und sind für eine erhebliche Morbidität und Mortalität verantwortlich. Die größte histologische Kategorie der primären Hirntumoren sind die Gliome, die ein äußerst heterogenes Erscheinungsbild aufweisen und radiologisch schwer von anderen Hirnläsionen zu unterscheiden sind. Die Neurochirurgie ist meist die Standardbehandlung für neu diagnostizierte Gliom-Patienten und kann von einer Strahlentherapie und einer adjuvanten Temozolomid-Chemotherapie gefolgt werden.

Die Hirntumor Chirurgie steht jedoch vor großen Herausforderungen, wenn es darum geht, eine maximale Tumorentfernung zu erreichen und gleichzeitig postoperative neurologische Defizite zu vermeiden. Zwei dieser neurochirurgischen Herausforderungen werden im Folgenden vorgestellt. Erstens ist die manuelle Abgrenzung des Glioms einschließlich seiner Unterregionen aufgrund seines infiltrativen Charakters und des Vorhandenseins einer heterogenen Kontrastverstärkung schwierig. Zweitens verformt das Gehirn seine Form—die so genannte "Hirnverschiebung"—als Reaktion auf chirurgische Manipulationen, Schwellungen durch osmotische Medikamente und Anästhesie, was den Nutzen präoperativer Bilddaten für die Steuerung des Eingriffs einschränkt.

Bildgesteuerte Systeme bieten Ärzten einen unschätzbaren Einblick in anatomische oder pathologische Ziele auf der Grundlage moderner Bildgebungsmodalitäten wie Magnetresonanztomographie (MRT) und Ultraschall (US). Bei den bildgesteuerten Instrumenten handelt es sich hauptsächlich um computergestützte Systeme, die mit Hilfe von Computer-Vision-Methoden die Durchführung perioperativer chirurgischer Eingriffe erleichtern. Die Chirurgen müssen jedoch immer noch den Operationsplan aus präoperativen Bildern gedanklich mit Echtzeitinformationen zusammenführen, während sie die chirurgischen Instrumente im Körper manipulieren und die Zielerreichung überwachen. Daher war die Notwendigkeit einer Bildführung während neurochirurgischer Eingriffe schon immer ein wichtiges Anliegen der Ärzte.

Ziel dieser Forschungsarbeit ist die Entwicklung eines neuartigen Systems für die peri-operative bildgeführte Neurochirurgie (IGN), nämlich DeepIGN, mit dem die erwarteten Ergebnisse der Hirntumor Chirurgie erzielt werden können, wodurch die Gesamtüberlebensrate maximiert und die postoperative neurologische Morbidität minimiert wird. Im Rahmen dieser Arbeit werden zunächst neuartige Methoden für die Kernbestandteile des DeepIGN-Systems der Hirntumor-Segmentierung im MRT und der multimodalen präoperativen MRT zur intraoperativen US-Bildregistrierung (iUS) unter Verwendung der jüngsten Entwicklungen im Deep Learning vorgeschlagen. Anschließend wird die Ergebnisvorhersage der verwendeten Deep-Learning-Netze weiter interpretiert und untersucht, indem für den Menschen verständliche, erklärbare Karten erstellt werden. Schließlich wurden Open-Source-Pakete entwickelt und in weithin anerkannte Software integriert, die für die Integration von Informationen aus Tracking-Systemen, die Bildvisualisierung und -fusion sowie die Anzeige von Echtzeit-Updates der Instrumente in Bezug auf den Patientenbereich zuständig ist.

Die Komponenten von DeepIGN wurden im Labor validiert und in einem simulierten Operationssaal evaluiert. Für das Segmentierungsmodul erreichte DeepSeg, ein generisches entkoppeltes Deep-Learning-Framework für die automatische Abgrenzung von Gliomen in der MRT des Gehirns, eine Genauigkeit von 0,84 in Bezug auf den Würfelkoeffizienten für das Bruttotumorvolumen. Leistungsverbesserungen wurden bei der Anwendung fortschrittlicher Deep-Learning-Ansätze wie 3D-Faltungen über alle Schichten, regionenbasiertes Training, fliegende Datenerweiterungstechniken und Ensemble-Methoden beobachtet.

Um Hirnverschiebungen zu kompensieren, wird ein automatisierter, schneller und genauer deformierbarer Ansatz, iRegNet, für die Registrierung präoperativer MRT zu iUS-Volumen als Teil des multimodalen Registrierungsmoduls vorgeschlagen. Es wurden umfangreiche Experimente mit zwei Multi-Location-Datenbanken durchgeführt: BITE und RESECT. Zwei erfahrene Neurochirurgen führten eine zusätzliche qualitative Validierung dieser Studie durch, indem sie MRT-iUS-Paare vor und nach der deformierbaren Registrierung überlagerten. Die experimentellen Ergebnisse zeigen, dass das vorgeschlagene iRegNet schnell ist und die besten Genauigkeiten erreicht. Darüber hinaus kann das vorgeschlagene iRegNet selbst bei nicht trainierten Bildern konkurrenzfähige Ergebnisse liefern, was seine Allgemeingültigkeit unter Beweis stellt und daher für die intraoperative neurochirurgische Führung von Nutzen sein kann.

Für das Modul "Erklärbarkeit" wird das NeuroXAI-Framework vorgeschlagen, um das Vertrauen medizinischer Experten in die Anwendung von KI-Techniken und tiefen neuronalen Netzen zu erhöhen. Die NeuroXAI umfasst sieben Erklärungsmethoden, die Visualisierungskarten bereitstellen, um tiefe Lernmodelle transparent zu machen. Die experimentellen Ergebnisse zeigen, dass der vorgeschlagene XAI-Rahmen eine gute Leistung bei der Extraktion lokaler und globaler Kontexte sowie bei der Erstellung erklärbarer Salienzkarten erzielt, um die Vorhersage des tiefen Netzwerks zu verstehen. Darüber hinaus werden Visualisierungskarten erstellt, um den Informationsfluss in den internen Schichten des Encoder-Decoder-Netzwerks zu erkennen und den Beitrag der MRI-Modalitäten zur endgültigen Vorhersage zu verstehen. Der Erklärungsprozess könnte medizinischen Fachleuten zusätzliche Informationen über die Ergebnisse der Tumorsegmentierung liefern und somit helfen zu verstehen, wie das Deep-Learning-Modell MRT-Daten erfolgreich verarbeiten kann.

Außerdem wurde ein interaktives neurochirurgisches Display für die Eingriffsführung entwickelt, das die verfügbare kommerzielle Hardware wie iUS-Navigationsgeräte und Instrumentenverfolgungssysteme unterstützt. Das klinische Umfeld und die technischen Anforderungen des integrierten multimodalen DeepIGN-Systems wurden mit der Fähigkeit zur Integration von (1) präoperativen MRT-Daten und zugehörigen 3D-Volumenrekonstruktionen, (2) Echtzeit-iUS-Daten und (3) positioneller Instrumentenverfolgung geschaffen. Die Genauigkeit dieses Systems wurde anhand eines benutzerdefinierten Agar-Phantom-Modells getestet, und sein Einsatz in einem vorklinischen Operationssaal wurde simuliert. Die Ergebnisse der klinischen Simulation bestätigten, dass die Montage des Systems einfach ist, in einer klinisch akzeptablen Zeit von 15 Minuten durchgeführt werden kann und mit einer klinisch akzeptablen Genauigkeit erfolgt.

In dieser Arbeit wurde ein multimodales IGN-System entwickelt, das die jüngsten Fortschritte im Bereich des Deep Learning nutzt, um Neurochirurgen präzise zu führen und prä- und intraoperative Patientenbilddaten sowie interventionelle Geräte in das chirurgische Verfahren einzu-

beziehen. DeepIGN wurde als Open-Source-Forschungssoftware entwickelt, um die Forschung auf diesem Gebiet zu beschleunigen, die gemeinsame Nutzung durch mehrere Forschungsgruppen zu erleichtern und eine kontinuierliche Weiterentwicklung durch die Gemeinschaft zu ermöglichen. Die experimentellen Ergebnisse sind sehr vielversprechend für die Anwendung von Deep-Learning-Modellen zur Unterstützung interventioneller Verfahren - ein entscheidender Schritt zur Verbesserung der chirurgischen Behandlung von Hirntumoren und der entsprechenden langfristigen postoperativen Ergebnisse.





# Abstract

Intracranial brain tumors are one of the ten most common malignant cancers and account for substantial morbidity and mortality. The largest histological category of primary brain tumors is the gliomas which occur with an ultimate heterogeneous appearance and can be challenging to discern radiologically from other brain lesions. Neurosurgery is mostly the standard of care for newly diagnosed glioma patients and may be followed by radiation therapy and adjuvant temozolomide chemotherapy.

However, brain tumor surgery faces fundamental challenges in achieving maximal tumor removal while avoiding postoperative neurologic deficits. Two of these neurosurgical challenges are presented as follows. First, manual glioma delineation, including its sub-regions, is considered difficult due to its infiltrative nature and the presence of heterogeneous contrast enhancement. Second, the brain deforms its shape, called “brain shift,” in response to surgical manipulation, swelling due to osmotic drugs, and anesthesia, which limits the utility of pre-operative imaging data for guiding the surgery.

Image-guided systems provide physicians with invaluable insight into anatomical or pathological targets based on modern imaging modalities such as magnetic resonance imaging (MRI) and Ultrasound (US). The image-guided toolkits are mainly computer-based systems, employing computer vision methods to facilitate the performance of peri-operative surgical procedures. However, surgeons still need to mentally fuse the surgical plan from pre-operative images with real-time information while manipulating the surgical instruments inside the body and monitoring target delivery. Hence, the need for image guidance during neurosurgical procedures has always been a significant concern for physicians.

This research aims to develop a novel peri-operative image-guided neurosurgery (IGN) system, namely DeepIGN, that can achieve the expected outcomes of brain tumor surgery, thus maximizing the overall survival rate and minimizing post-operative neurologic morbidity. In the scope of this thesis, novel methods are first proposed for the core parts of the DeepIGN system of brain tumor segmentation in MRI and multimodal pre-operative MRI to the intra-operative US (iUS) image registration using the recent developments in deep learning. Then, the output prediction of the employed deep learning networks is further interpreted and examined by providing human-understandable explainable maps. Finally, open-source packages have been developed and integrated into widely endorsed software, which is responsible for integrating information from tracking systems, image visualization, image fusion, and displaying real-time updates of the instruments relative to the patient domain.

The components of DeepIGN have been validated in the laboratory and evaluated in the simulated operating room. For the segmentation module, DeepSeg, a generic decoupled deep learning framework for automatic glioma delineation in brain MRI, achieved an accuracy of 0.84 in terms of the dice coefficient for the gross tumor volume. Performance improvements were observed when employing advancements in deep learning approaches such as 3D convolutions

over all slices, region-based training, on-the-fly data augmentation techniques, and ensemble methods.

To compensate for brain shift, an automated, fast, and accurate deformable approach, iRegNet, is proposed for registering pre-operative MRI to iUS volumes as part of the multimodal registration module. Extensive experiments have been conducted on two multi-location databases: the BITE and the RESECT. Two expert neurosurgeons conducted additional qualitative validation of this study through overlaying MRI-iUS pairs before and after the deformable registration. Experimental findings show that the proposed iRegNet is fast and achieves state-of-the-art accuracies. Furthermore, the proposed iRegNet can deliver competitive results, even in the case of non-trained images, as proof of its generality and can therefore be valuable in intra-operative neurosurgical guidance.

For the explainability module, the NeuroXAI framework is proposed to increase the trust of medical experts in applying AI techniques and deep neural networks. The NeuroXAI includes seven explanation methods providing visualization maps to help make deep learning models transparent. Experimental findings showed that the proposed XAI framework achieves good performance in extracting both local and global contexts in addition to generating explainable saliency maps to help understand the prediction of the deep network. Further, visualization maps are obtained to realize the flow of information in the internal layers of the encoder-decoder network and understand the contribution of MRI modalities in the final prediction. The explainability process could provide medical professionals with additional information about tumor segmentation results and therefore aid in understanding how the deep learning model is capable of processing MRI data successfully.

Furthermore, an interactive neurosurgical display has been developed for interventional guidance, which supports the available commercial hardware such as iUS navigation devices and instrument tracking systems. The clinical environment and technical requirements of the integrated multi-modality DeepIGN system were established with the ability to incorporate: (1) pre-operative MRI data and associated 3D volume reconstructions, (2) real-time iUS data, and (3) positional instrument tracking. This system's accuracy was tested using a custom agar phantom model, and its use in a pre-clinical operating room is simulated. The results of the clinical simulation confirmed that system assembly was straightforward, achievable in a clinically acceptable time of 15 min, and performed with a clinically acceptable level of accuracy.

In this thesis, a multimodality IGN system has been developed using the recent advances in deep learning to accurately guide neurosurgeons, incorporating pre- and intra-operative patient image data and interventional devices into the surgical procedure. DeepIGN is developed as open-source research software to accelerate research in the field, enable ease of sharing between multiple research groups, and continuous developments by the community. The experimental results hold great promise for applying deep learning models to assist interventional procedures – a crucial step towards improving the surgical treatment of brain tumors and the corresponding long-term post-operative outcomes.

# Acknowledgment

The research conducted over the past years at CaMed (Hochschule Reutlingen) and HERA (KIT), which has led to this dissertation, would not have been possible without continuous guidance, encouragement, and inspiration of my advisors, mentors, colleagues, friends, and family. The research documented in this doctoral thesis has been funded by the German Academic Exchange Service (DAAD) under Scholarship No. 91705803.

First and foremost, I would like to sincerely thank Prof. Dr. Franziska Mathis-Ullrich as well as Prof. Dr.-Ing. Oliver Burgert for guiding me through this work, having interesting and inspiring discussions, as well as their continuous support for all topics related to this Ph.D. thesis. Further, I am very grateful for giving me the necessary freedom to pursue new ideas and for offering valuable advice and a fresh perspective at critical moments.

I also owe a large debt of gratitude to Dr. Mohamed Esmail Karar for his constant advice and help and for giving invaluable feedback in the process leading up to this thesis. In particular, I would like to acknowledge the scientific guidance and helpful discussions over the past years.

Furthermore, I would like to thank Prof. Dr. Christian Rainer Wirtz for agreeing to revise and give me feedback on this thesis, which is a great honor since he and his group are frontrunners in the area of neurosurgery. In addition, I would like to thank Prof. Dr. Frank Bellosa, Prof. Dr. Mehdi B. Tahoori, and Prof. Dr. Michael Beigl for their cooperativeness and their willingness to act as the examiner, chairman, and vice chairman.

Many thanks to all the colleagues at Reutlingen University and at KIT for the warm welcome and great time in Germany. I would especially like to thank them for their support, scientific discussions, and friendly atmosphere. I also want to thank all my students who have contributed to my project. Especially: Jana Ickenroth, Arthur Krauß, Tim Mangliers, Alex Pollock, and Pauline Weimann.

I further want to thank our clinical collaborators, Prof. Dr. med. Jan Coburger, Dr. med. Ziad Elshaer, and Dr. med. univ. Markus Schmidhammer, whose expertise and insights were indispensable and helped ground this work in clinical desiderata.

A very warm thank you goes out to my colleagues at the Process Analysis and Technology (PA&T), Faculty of Applied Chemistry, Reutlingen University, in particular Mohammed Al Ktash and Dr. Josune Jimenez Ezenarro for their help and support during the brain phantom fabrication, for which I am truly grateful.

A resounding and heartfelt thank you shall go to my dear family, without whom I could not bring this work to an end. I would like to express my deep appreciation to my father, Prof. Dr. Ashraf Zeineldin, who inspired me to start my academic career path, all of the contributions he has given in my life and allowed me to follow in his footsteps, as my role model. I am infinitely grateful to my mum, Dr. Hanaa El-Demiry, for giving me life and love, and for the millions of prayers she has sent up on my behalf. I feel protected and blessed by their care and devotion.

Big thanks to my siblings Ahmed and Karim who provided me with grounding, stability, and joyful holidays back home whenever I visited.

Lastly, I am deeply indebted to my wife, Asmaa, who was always supportive, never (well, hardly ever) complained about long working hours, and whose open ears and heart were key in getting me this far. To my son Josef for making every day of my life better. THANK YOU!

Reutlingen, December 2022

Ramy Zeineldin

*Dedicated to my dad, mom, and family*



# Table of Contents

<b>Kurzfassung</b>	<b>xiii</b>
<b>Abstract</b>	<b>v</b>
<b>Acknowledgment</b>	<b>vii</b>
<b>Table of Contents</b>	<b>xi</b>
<b>List of Figures</b>	<b>xv</b>
<b>List of Tables</b>	<b>xxi</b>
<b>List of Abbreviations</b>	<b>xxiii</b>
<b>Table of Symbols</b>	<b>xxvii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation	2
1.2 Objectives of the Thesis	4
1.3 Thesis Roadmap	7
<b>2 Medical Background</b>	<b>9</b>
2.1 Medical Imaging Techniques	10
2.1.1 Magnetic Resonance Imaging	10
2.1.2 Ultrasound Imaging	12
2.2 Brain Tumors	13
2.2.1 Diagnosis	14
2.2.2 Treatment	16
<b>3 Technical Background</b>	<b>19</b>
3.1 Deep Learning	20
3.1.1 Artificial Neural Networks	20
3.1.2 Convolutional Neural Networks	22
3.1.3 CNN Architectures	23
3.1.4 Transformers	26
3.1.5 Deep Network Training	27
3.2 Image-guided Neurosurgery	28
3.2.1 IGN Components	28
3.2.2 IGN Software	29
3.3 Challenges and Limitations	30
3.3.1 Neurosurgical Challenges	30
3.3.2 Technical Challenges	31
<b>4 Brain Tumor Segmentation</b>	<b>33</b>
4.1 Introduction	34
4.2 Related Work	34
4.3 Methods	36
4.3.1 DeepSeg	36
4.3.2 3D DeepSeg	37
4.3.3 Ensemble Methods	38
4.3.4 Post-processing	39

4.4	Experiments	39
4.4.1	Data	39
4.4.2	Experimental setup	42
4.4.3	Evaluation Metrics	43
4.4.4	Ablation Study	44
4.5	Results and Discussion	45
4.5.1	Statistical Evaluation	45
4.5.2	Qualitative Output	47
4.5.3	Ensemble Results	49
4.6	Summary	52
<b>5</b>	<b>Multimodal Registration</b>	<b>53</b>
5.1	Introduction	54
5.2	Related Work	54
5.3	Methods	55
5.3.1	Deformable Image Registration	55
5.3.2	Learning-based Registration Framework	56
5.3.3	Self-supervised iRegNet	59
5.4	Experiments	60
5.4.1	Data	60
5.4.2	Registration Procedure	61
5.4.3	Experimental Setup and Evaluation	62
5.5	Results and Discussion	64
5.5.1	Quantitative Registration Results	64
5.5.2	Generality Evaluation	64
5.5.3	Processing Time Analysis	66
5.5.4	Comparison with Other MRI-iUS Studies	68
5.5.5	Qualitative Analysis by Neurosurgeons	70
5.5.6	Self-supervised Results	72
5.6	Summary	75
<b>6</b>	<b>Explainability of Deep Neural Networks</b>	<b>77</b>
6.1	Introduction	78
6.2	Related Work	78
6.3	Methods	80
6.3.1	NeuroXAI	80
6.3.2	TransXAI	84
6.4	Experiments	87
6.4.1	Data	87
6.4.2	Experimental Setup	88
6.5	Results and Discussion	89
6.5.1	Showcase I: Application to Classification	89
6.5.2	Showcase II: Application to Segmentation	92
6.6	Clinical Relevance	101
6.7	Summary	102
<b>7</b>	<b>DeepIGN Integration and Neuronavigation Display</b>	<b>103</b>
7.1	Introduction	104
7.2	Related Work	104
7.3	System Design	104
7.3.1	Design Requirements	105
7.3.2	System Components	106



---

7.4	Results and Discussion	110
7.4.1	Pre-operative Application to BraTS Dataset	110
7.4.2	Phantom Study	111
7.4.3	Comparison with IGN Platforms	118
7.5	Summary	120
<b>8</b>	<b>Conclusions and Outlook</b>	<b>121</b>
8.1	Conclusions	122
8.2	Outlook	124
8.2.1	Multimodal Registration	124
8.2.2	Need for large datasets	125
8.2.3	Mixed Reality	125
8.2.4	Explainable Deep Models	126
<b>Appendix A. Brain Phantom Preparation</b>		<b>127</b>
<b>Appendix B. List of Publications</b>		<b>131</b>
<b>References</b>		<b>133</b>



# List of Figures

Figure 1.1:	Cranial navigation using the Brainlab image-guided system with optical tracking and surgical displays. Courtesy of Brainlab AG. ....	3
Figure 1.2:	Overview of the contribution areas and the underlying research questions of this Ph.D. thesis.....	5
Figure 2.1:	Hydrogen nuclei spinning phenomena. (a) In the absence of a magnetic field, the magnetic moments of the nuclei are distributed randomly and thus the net magnetization factor is zero. (b) In the case of a strong external magnetic field, the spinning nuclei parallel (spin up) or antiparallel to the external field (spin down) ( $B_0$ ) (van Geuns et al., 1999).....	11
Figure 2.2:	MRI Scanner at Ulm University Hospital on a 1.5 Tesla Magnetom Symphony (Siemens Medical, Germany) with a 12-channel head coil.....	12
Figure 2.3:	An example of a B-mode US scanner (a) and its output images (b) showing the brain tumor which appears in bright grayscale in axial, sagittal, and coronal views (Shetty & Moiyadi, 2016; Xiao, Fortin, Unsgard, Rivaz, & Reinertsen, 2017).....	13
Figure 2.4:	Overview of different MRI modalities showing the characteristics of brain Glioma from the BraTS dataset (Menze et al., 2015). Image patches show the different modalities of T1 (a), T1Gd (b), T2 (c), and FLAIR (d).....	15
Figure 2.5:	Fusion of iUS images ( <i>green</i> ) over pre-operative MRI ( <i>grey</i> ) of a sample patient from the BITE dataset (Mercier, Del Maestro, et al., 2012). From left to right: FLAIR MRI, B-Mode US image, and the corresponding overlay of MRI and US images. ....	16
Figure 2.6:	Overview of traditional approaches as well as new hybrid strategies for glioblastoma therapy (Shah & Heiss, 2022).....	17
Figure 3.1:	Sample three-layer neural network with an input layer ( <i>blue</i> ), two hidden layers ( <i>orange</i> ), and an output layer ( <i>green</i> ) that are fully connected only in the forward direction. ....	20
Figure 3.2:	Concept of artificial neural networks, a perceptron ( <i>up</i> ), which was originally inspired by biological neural systems, a biological neuron ( <i>down</i> ) (Zhang, Yu, Barbiero, Wang, & Gu, 2019).....	21
Figure 3.3:	Two ideas employed by a convolutional layer to reduce the number of parameters and thus the network complexity: sparse connection and parameter sharing. The first row shows layer interconnection for a fully connected layer as well as the two enhancements by the convolution operation. Their respective weight matrices are presented in the second row.....	23

Figure 3.4:	Architecture of a convolutional network for image classification (Kang, Ullah, & Gwak, 2021). .....	23
Figure 3.5:	Comparison of the basic blocks for different convolutional networks. (a) the residual block of ResNet; (b) a 3-layer dense block, (c) a depth-wise based module of MobileNet, (d) a simplified Inception module, (e) an extreme module of Inception (Xception module), and (f) MobileNetV2 blocks with two stride values.....	24
Figure 3.6:	U-Net network consists of convolutional blocks ( <i>blue boxes</i> ), maximum pooling ( <i>orange arrows</i> ), upsampling ( <i>grey arrows</i> ), and softmax output ( <i>green block</i> ).....	26
Figure 3.7:	Overview of the Transformer architecture (Vaswani et al., 2017). .....	27
Figure 4.1:	DeepSCAN architectures, as applied to brain tumor segmentation in BraTS 2018 challenge (McKinley, Meier, & Wiest, 2019). .....	35
Figure 4.2:	nnU-Net network consists of strided convolution blocks ( <i>grey boxes</i> ), and upsampling as convolution transposed ( <i>blue arrows</i> ). The input patch size was set to $128 \times 128 \times 128$ and the maximum filter size is 320 (Fabian Isensee et al., 2021). .....	36
Figure 4.3:	DeepSeg architecture for using different feature extractor models of MRI brain tumors. ....	37
Figure 4.4:	3D DeepSeg network consists of convolution neural blocks ( <i>blue boxes</i> ), downsampling using maximum pooling ( <i>orange arrows</i> ), upsampling using up convolution ( <i>blue arrows</i> ), and softmax output layer ( <i>green block</i> ). The input patch size was set to $128 \times 128 \times 128$ . .....	38
Figure 4.5:	A sample of the mpMRI BraTS 2021 training set. Shown are image slices in two different MRI modalities T1Gd (a), T2 (b), FLAIR (c), and the ground truth segmentation (d). The color labels indicate Edema ( <i>yellow</i> ), enhancing solid tumor ( <i>red</i> ), non-enhancing tumor core, and necrotic core ( <i>green</i> ). Images were obtained by using the 3D Slicer software (Fedorov et al., 2012).....	40
Figure 4.6:	Random augmented image transformation. The first row shows the original image. The next three rows present horizontal and vertical flipping, scaling, translation, rotation, and shearing methods. The elastic transformation is presented in the last row. ....	42
Figure 4.7:	Brain tumor segmentation results. FLAIR, ground truth, and output of Original U-Net, Modified U-Net, VGGNet, ResNet, and DenseNet. Tumor regions are indicated in red. ....	47
Figure 4.8:	Brain tumor segmentation results. FLAIR, ground truth, and output of Xception, MobileNet, NASNet, and MobileNetV2. Tumor regions are indicated in red. ....	48
Figure 4.9:	Sample qualitative validation set results of the ensemble model. The best, median, and worse cases are shown in the rows. Columns display the T2, FLAIR, and overlay of the predicted segmentation on the	

	FLAIR image. WT includes all visible labels (green, yellow and red labels), TC is a union of green and red, while the ET class is shown in red.....	51
Figure 5.1:	A representative workflow of the proposed MRI-iUS deformable registration approach, where dashed red arrows indicate the data flow only required in the training stage.....	56
Figure 5.2:	The enhanced 3D CNN architecture for predicting the deformation field $\phi$ . The network consists of 3D convolution with batch normalization layers ( <i>blue blocks</i> ), maximum pooling ( <i>orange arrows</i> ), up-sampling ( <i>grey arrows</i> ), and concatenate connections ( <i>dashed grey lines</i> ). The number of features is doubled in each step of the encoder part, while halved in the decoder part. ....	57
Figure 5.3:	The MRI-iUS registration procedure (a) Input MRI; (b) Input iUS; (c) Cropped MRI; (d) An overlap between iUS and corrected MRI.....	62
Figure 5.4:	Processing time analysis for the proposed approaches on two different multiple sites datasets. (a) Training time in minutes; (b) Test time in seconds using CPU ( <i>in Blue</i> ) and GPU ( <i>in Red</i> ).....	67
Figure 5.5:	A comparison of the registration error ( <i>mTRE</i> ) for the proposed iRegNet methods and the state-of-the-art methods on the BITE dataset.....	69
Figure 5.6:	A comparison of the registration error ( <i>mTRE</i> ) for the proposed iRegNet methods and the state-of-the-art methods on the RESECT dataset.....	69
Figure 5.7:	Alignment of pre-operative MRI ( <i>gray color mode</i> ) to iUS ( <i>green color mode</i> ) in six different cases from (a) BITE (cases #6, #8, and #11) and (b) RESECT (cases #23, #14, and #21). Column designations: pre-operative FLAIR MRI, intra-operative US, initial misaligned MRI over iUS before registration, and the final aligned MRI over iUS after registration, and the deformed grids. Yellow arrows indicate expert-labeled landmarks while <i>mTRE</i> values are shown ( <i>bottom right</i> ).....	71
Figure 5.8:	Example registration results from three validation cases (patients 141, 148, and 152). From left to right: (a) and (b) the post- and pre-operative MRI T1ce, (c) the follow-up to pre-operative affine registration of BRAINSFit, (d) the follow-up to pre-operative deformable registration of the iRegNet, (e) the pre-operative FLAIR scans, only for visualization purposes, and (f) determinant of the Jacobian of the displacement field are shown, respectively. The red box highlights regions of major differences.....	73
Figure 5.9:	Boxplots of the mean landmark errors. For each method, the landmark errors are computed against the fixed landmarks of the BraTS-Reg dataset. From left to right, mean absolute registration errors are shown for the initial dataset, affine, and the enhanced iRegNet, respectively. On each box, the red line is the median and the green triangle is the mean. ....	73

Figure 6.1:	Pipeline of the proposed NeuroXAI framework.....	80
Figure 6.2:	Applying Guided Backpropagation to a sample brain tumor segmentation CNN model.....	81
Figure 6.3:	Applying Grad-CAM to a sample brain tumor segmentation CNN model.....	83
Figure 6.4:	Applying Guided Grad-CAM to a sample brain tumor segmentation CNN model.....	84
Figure 6.5:	Overall proposed TransXAI pipeline for visual justification of glioma segmentation in brain MRI using a hybrid CNN-Transformer architecture. ....	84
Figure 6.6:	The architecture of the hybrid CNN-Transformer brain segmentation network from mpMRI volumes. The input is a 2D multi-modal MRI of T1, T1Gd, T2, and FLAIR with a patch spatial resolution of $192 \times 192 \times 4$ . The network has 8 convolution neural blocks ( <i>blue boxes</i> ), each consisting of two successive convolutional layers $3 \times 3$ , BN layer, and ReLU activation.....	85
Figure 6.7:	Glioma sub-regions in a sample scan from the BraTS 2019 challenge database. Image patches show the different modalities of T1 (a), T1Gd (b), T2 (c), FLAIR (d), and annotated expert-labeled tumor segmentation (e). Ground truth segmentation is provided for the enhancing tumor ( <i>blue</i> ) surrounding the non-enhancing necrotic tumor core ( <i>green</i> ) visible in T1Gd, and (b) the peritumoral Edema ( <i>yellow</i> ) visible in the FLAIR, respectively.....	87
Figure 6.8:	The architecture of the enhanced 3D brain segmentation network (DeepSeg) for brain tumor segmentation from mpMRI volumes. The input is a 3D multi-modal MRI of T1, T1Gd, T2, and FLAIR with a patch spatial resolution of $192 \times 224 \times 160$ . The CNN network has 24 convolution neural blocks ( <i>blue boxes</i> ), four downsampling blocks ( <i>orange boxes</i> ), four upsampling blocks ( <i>grey boxes</i> ), and a final softmax output layer ( <i>green box</i> ).....	89
Figure 6.9:	Comparing different XAI visualization methods for brain glioma classification. Sensitivity maps are presented for four HGG cases. ....	90
Figure 6.10:	Comparing different XAI visualization methods for brain glioma classification. Sensitivity maps are presented for three LGG cases. ....	91
Figure 6.11:	Comparing different XAI visualization methods for brain glioma segmentation.....	94
Figure 6.12:	Visualization of the information flow in the segmentation CNN internal layers. The input MRI sequences are shown in (a). (b) – (d) show implicit concepts for which no ground truth labels are available in addition to explicit concepts (e) – (g) with trained labels. L stands for convolutional layer. ....	95

Figure 6.13:	Impact of MRI input modality in the detection of different tumor labels. The first row shows the input MRI sequences and the ground truth annotations. The following rows correspond to label 1 (the necrotic tumor core), label 2 (the peritumoral edema), and label 4 (the enhancing tumor). In the saliency maps, warmer regions represent a high score for the specified label detection. ....96	96
Figure 6.14:	Saliency maps for implicit concepts learned by individual filters of the CNN model. It is interesting to note that there are no labels for these concepts in the training dataset. Warmer regions represent a high score for the specified concept in the prediction map. Note that EB and DB denote the encoder and decoder block layers, individually. ....97	97
Figure 6.15:	Saliency maps for explicit concepts learned by individual filters of the CNN model which are labeled in the training dataset. Warmer regions represent a high score for the specified concept in the prediction map. Note that EB and DB denote the encoder and decoder block layers, individually. ....98	98
Figure 6.16:	Prediction failure detection DSC metrics on 70 selected subjects from the BraTS 2021 validation database. Each point per color represents one of the brain tumor sub-regions. WT, TC, and ET stand for Whole Tumor, Tumor Core, and Enhancing Tumor, respectively. ....99	99
Figure 6.17:	Overview of the relevant gradient-based class activation maps for failure cases. ....100	100
Figure 7.1:	Existing open-source platforms dedicated to IGS and their dependencies. From left to right: (1) Common libraries for interfacing with the hardware in the operating room. (2) General imaging platforms that provide visualization capabilities with customizable interfaces. (3) Add-on plugins for the general-purpose platforms in level (2) to provide image-guidance capabilities. (4) Fully integrated IGS platforms for specific target applications. ....105	105
Figure 7.2:	Overview of the proposed DeepIGN navigation system for assisting neurosurgery. NDI Vicra system is tracking the three optical markers: attached to the brain phantom, pointer tool (stylus), and the ultrasound transducer probe. Each marker is composed of four optically reflective spheres. ....107	107
Figure 7.3:	The hierarchy of transformations within the 3D slicer software ( <i>left</i> ), and their representations ( <i>right</i> ). <i>purple transforms</i> represent coordinate transformations from all tracked tools to the tracking camera, whereas PLUS Toolkit is responsible for generating the transforms of tracked tools to the reference ( <i>blue transforms</i> ). <i>Orange transforms</i> show fixed transforms that are the result of all registration procedures (Preiswerk, Brinker, McDannold, & Mariano, 2019). ....108	108
Figure 7.4:	A high-level overview of the Proposed AI-powered IGN platform architecture and its relationship with 3D Slicer software and other	

---

	libraries. Each extension has its dependencies for image processing, tracking information, resampling, or registration. ....	110
Figure 7.5:	DeepSeg module during visualization of the resultant tumor boundaries. The GUI panel is shown on the left, where the user can specify inputs, output volumes, model parameters, and other 3D visualization options. Brain tumor segmentation results ( <i>green</i> ) of a sample HGG case from the BraTS 2021 challenge are presented on the right. The application layout shows the axial ( <i>top middle</i> ), sagittal ( <i>bottom middle</i> ), and coronal ( <i>bottom right</i> ), as well as 3D views ( <i>top right</i> ).....	110
Figure 7.6:	Validation of the DeepIGN system for use in guided neurosurgery in the simulated OR using the fabricated brain phantom. (a) custom brain phantom; (b) pre-operative FLAIR MRI with tumor highlighted in <i>green</i> ; (c) live iUS image captured with the Esaote MyLabSat ultrasound linear transducer.....	112
Figure 7.7:	Interactive segmentation using the segment editor module as a built-in module of 3D Slicer. Tools such as threshold, paint, draw, erase, and grow from seed are available on the left ( <i>module panel</i> ). ....	113
Figure 7.8:	The NeuroXAI module provides sensitivity maps to explain the prediction of the tumor segmentation in brain MRI. ....	113
Figure 7.9:	Automatic image-to-image registration using the iRegNet module. MRI and iUS scans are utilized as moving and fixed images, respectively, since the main goal is to reflect the brain shift in the pre-operative MRI data. ....	114
Figure 7.10:	Illustration of the stylus calibration using the VTK landmark registration method as part of the SlicerIGT extension. (a) – (d) show different fiducial landmarks, whereas the resultant registered stylus pointer is displayed in <i>cyan</i> (e).....	115
Figure 7.11:	Illustration of the live US acquisition as part of the DeepIGN system. (a) provides US view with recording and reconstruction of 3D US images. (b) provide a view of the tracked US probe and its location with respect to the pre-operative MRI. ....	116
Figure 7.12:	Intra-operative simulation of 3D Ultrasound reconstruction using patient data and a patient-specific phantom model.....	117



# List of Tables

Table 3.1:	Comparison of the available platforms for image-guided interventions regarding visualization, segmentation, registration, acquisition, and iUS navigation. ....	30
Table 4.1:	List of the applied data augmentation methods. ....	41
Table 4.2:	A comparative performance of the employed models. Average computational times for each encoder of 35 results during training and validation phases. ....	45
Table 4.3:	Mean DSC, Sensitivity, Specificity, and Hausdorff distance scores of testing different encoders on BraTS 2019 training data. ....	46
Table 4.4:	Mean DSC, Sensitivity, and Specificity scores of applied models on BraTS 2019 validation data. ....	46
Table 4.5:	Results of the five-fold cross-validation models on BraTS 2021 validation cases. All reported values were computed by the online evaluation platform Synapse. The average of DSC and HD95 scores are computed and used for ranking the methods locally. ....	50
Table 5.1:	The deformable CNN architecture details. ....	58
Table 5.2:	A detailed description of the two databases used in this study. ....	61
Table 5.3:	Pre-and post-registration TREs for the proposed models in the BITE dataset. <i>Model B, Model R, and Model C</i> represent the proposed models trained on the BITE, RESECT, and both datasets, correspondingly. Test cases are shown in <i>bold underlined</i> and the lowest error in each row is highlighted in <i>bold</i> . ....	65
Table 5.4:	Pre- and post-registration TREs for the proposed models in the RESECT dataset. <i>Model B, Model R, and Model C</i> represent the proposed models trained on the BITE, RESECT, and both datasets, correspondingly. Test cases are shown in <i>bold underlined</i> and the lowest error in each row is highlighted in <i>bold</i> . ....	65
Table 5.5:	Statistical assessment of the MRI-iUS alignment of 36 cases by two expert neurosurgeons. ....	70
Table 5.6:	The ablation study of MRI modalities on the BraTS-Reg 2022 validation cases. Bold highlights the best scores. ....	72
Table 5.7:	Quantitative results of the proposed method and the baseline affine method on the BraTS-Reg challenge validation set. MAE denotes the average of median absolute error between the predicted coordinates and the ground truth coordinates, whereas Robustness represents the successful rate of measuring how many landmarks have improved MAE after the registration. ....	74

Table 6.1:	List of the added top layers to the standard ResNet-50. ....	88
Table 6.2:	Quantitative results of CNN models on the BraTS 2021 Challenge database. ....	93
Table 7.1:	Design requirements for a multimodal image-guided navigation system for assisting neurosurgery. CRX and TRX stand for clinical and technical requirements, respectively.....	106
Table 7.2:	Runtime measurement comparing the two integrated DNNs: DeepSeg and nnU-Net on both CPU and GPU implementations. ....	111
Table 7.3:	Comparative analysis of the proposed DeepIGN platform against the available image guidance systems. ....	119

# List of Abbreviations

AI	Artificial Intelligence
ANNs	Artificial neural networks
ASPP	Atrous spatial pyramid pooling
BN	Batch normalization
BraTS	Multimodal Brain Tumor Segmentation Challenge
CAM	Class Activation Mapping
CC	Correlation Coefficient
CE	Cross-entropy
CNN	Convolutional neural network
CR	Correlation Ratio
CuRIOUS	Correction of Brain shift with Intra-Operative Ultrasound challenge
DenseNet	Dense Convolutional Network
DNN	Deep neural network
DSC	Dice similarity coefficient
ED	Peritumoral edematous/invaded tissue
EOR	Extent of resection
ET	Enhancing tumor
FLAIR	T2-weighted fluid-attenuated inversion recovery
FOV	Field of view
GAN	Generative adversarial network
GBM	Glioblastoma
GBP	Guided Backpropagation
GCAM	Grad-CAM
GDPR	General Data Protection Regulation

GGCAM	Guided Grad-CAM
GIG	Guided Integrated Gradients
GMM	Gaussian mixture model
GPU	Graphical processing unit
HD	Hausdorff distance
HGGs	High-grade gliomas
IG	Integrated Gradients
IG	Integrated Gradients
IGN	Image-guided neurosurgery
IGSTK	Image-Guided Surgery Toolkit
ILSVRC	ImageNet Large-Scale Visual Recognition Challenge
iMRI	intra-operative MRI
iUS	Intra-operative ultrasound images
LC2	Linear Correlation of Linear Combination
LGGs	Low-grade gliomas
LN	layer normalization
LReLU	Leaky ReLU
MI	Mutual Information
MICCAI	Medical Image Computing and Computer-Assisted Intervention
MITK	Medical Imaging Interaction Toolkit
MLP	Multilayer perceptron network
MRI	Magnetic resonance imaging
MSA	Multi-head self-attention mechanism
mTRE	Mean target registration errors
NCC	Normalized Correlation Coefficient
NCR	Necrotic components of the core tumor
NLP	Natural language processing

NMR	Nuclear magnetic resonance
RBM	Restricted Boltzmann Machine system
ReLU	Rectified Linear Unit
ResNet	Residual Neural Network
RF	Radiofrequency waves
RF	Random Forest
SGD	Stochastic gradient descent
SHAP	Shapely additive explanations
SSA	Sub-Saharan Africa population
SSC	Self-Similarity Correlation
STAPLE	Simultaneous Truth and Performance Level Estimation
T1	T1-weighted sequences
T1Gd	T1-weighted sequences with a gadolinium-based contrast agent
T2	T2-weighted images
TC	Tumor core
VG	Vanilla Gradient
VGGNet	Visual Geometry Group Network
ViT	Vision Transformer
WT	Whole tumor
XAI	Explainable Artificial Intelligence
XEGK	Explainable ensemble Gaussian kernel



# Table of Symbols

## Calligraphic Symbols

$\mathbb{R}$  regularization term

$\mathcal{S}$  subset of features

## Greek Symbols

$\phi$  deformation field

$\lambda$  regularization term

$\mathcal{B}$  backpropagated gradient

$\mathcal{E}$  energy function

$\mathcal{G}$  Gaussian noise

$\mathcal{L}$  loss function defined on a data pair

$\mathcal{Y}$  points on the guided integrated gradients path

$\alpha$  interpolation constant

$\beta$  multiplying factor of the displacement fields

$\gamma$  gyromagnetic ratio of the nucleus

$\varepsilon$  smooth parameter

$\sigma$  elasticity coefficient

$\sigma^2$  variance

## Roman

$A$  forward pass activation map

$b$  constant bias term

$B_0$  strength of the applied magnetic field

$D$  set of training samples

$d$  Euclidian distance

$d_{pred}$	predicted spatial gradient
$d_{truth}$	ground truth gradient
$F$	resonant frequency
$f$	feature map
$FN$	false positives
$FN$	false negatives
$g$	activation function of the neuron
$G$	deep neural network
$H$	height of the input image
$I_F$	target or fixed image
$I_M$	moving image
$K$	downsampling factor
$L$	loss function of deep neural network
$L_{2D}$	weighted cross-entropy loss
$L_{3D}$	3D loss function
$L_{disp}$	displacement loss function
$L_{overall}$	overall loss function
$L_{sim}$	similarity loss function
$m$	expert-labeled annotations in MRI volume
$M_c$	original sensitivity map
$N_c$	number of classes
$N_m$	number of modalities
$N_p$	number of patches in the images
$N_s$	number of samples
$p$	network softmax predictions
$P$	patch in the image



---

$p_c$	predicted softmax output for class c
$P_c$	maximum gradient value
$R$	resultant reconstructed image
$S$	level of similarity
$TN$	true negatives
$TP$	true positives
$u$	expert-labeled annotations in ultrasound data
$w$	neural network weights
$W$	width of the input image
$w_c$	weight class map
$x$	input signal/ volume
$X^l$	input image
$y$	desired output/ ground truth
$y_c$	true labels for class c
$z_\ell$	output of the $\ell$ -the layer in the Transformer
$\hat{z}_\ell$	encoded image representation of the Transformer



# 1 Introduction

---

## Contents

<b>1.1 Motivation</b> .....	<b>2</b>
<b>1.2 Objectives of the Thesis</b> .....	<b>4</b>
<b>1.3 Thesis Roadmap</b> .....	<b>7</b>

This chapter gives an overview of the main objectives of this dissertation. Furthermore, the main contributions as well as the outline of this thesis are provided.

## 1.1 Motivation

Intra-axial brain tumors are among the ten most common malignancies leading to death (Weller et al., 2021). Although there are no screening or preventive examinations, effective diagnosis, and therapy influence the further course of gliomas. Neurosurgical intervention is the first and sometimes the only therapy for many types of gliomas (Jain, 2018; Pala et al., 2021). In particular, the precise localization of pathological targets (lesions) within the brain anatomy is a major issue in neurosurgery. This challenge is related to the difficulty in visually delineating these pathological structures from healthy tissue. Magnetic resonance imaging (MRI) is an important modality during the diagnosis, management, and care of glioma patients. MRI is the modality of choice for the evaluation of intra-axial, identification of anatomic detail of normal brain structures, peritumoral edema, and detection of tumor-infiltrated regions (Pope & Brandal, 2018; Upadhyay & Waldman, 2011).

Multi-parametric MRI of the brain, including T1-weighted sequences with (T1Gd) and without gadolinium-based contrast agent (T1), T2-weighted images (T2), and T2-weighted fluid-attenuated inversion recovery (FLAIR) sequences, is the gold standard to detect brain gliomas including their sub-regions (Ellingson, Wen, & Cloughesy, 2017). The use of a gadolinium-based contrast agent allows for the detection of areas where the blood-brain barrier is compromised. In general, gliomas are hypointense on T1 images and hyperintense on T2 images. The presence of peripheral contrast enhancement, central necrotic areas, intra-tumoral hemorrhages, ill-defined infiltration of surrounding brain tissue, and extensive perifocal edema is commonly seen in aggressive lesions and raises the possibility of high-grade glial lesions. The presence of peripheral enhancement with central necrotic regions is a common feature of high-grade gliomas (HGGs) or glioblastoma (GBM) (WHO grade IV). However, non-enhancing lesions may represent low-grade gliomas (LGGs).

GBM, the most common and aggressive malignant primary tumor of the brain in adults, occurs with ultimate heterogeneous sub-regions including the enhancing tumor (ET), peritumoral edematous/invaded tissue (ED), and the necrotic components of the core tumor (NCR) (Baid et al., 2021; Louis et al., 2020). Still, accurate GBM localization and its sub-regions in MRI are considered one of the most challenging segmentation problems in the medical field. Manual segmentation is the gold standard for neurosurgical planning, interventional image-guided surgery, follow-up procedures, and monitoring tumor growth. However, identification of the GBM tumor and its sub-regions by hand is time-consuming, subjective, and highly dependent on the experience of clinician's experience.

Image-guided neurosurgery (IGN) technology provides physicians with invaluable insight into anatomical or pathological targets, based on modern imaging modalities. The IGN is mainly a computer-based system with four primary components: segmentation, registration, tracking devices, and visualization (Buchholz, Smith, Laycock, & McDurmont, 2001; Cleary & Peters, 2010). Therefore, the evolution of computers has played an important role in shaping the IGS (Menze et al., 2015; Miner, 2017). Figure 1.1 depicts a typical IGN platform that employs an optical tracking system to relate surgical instruments to the patient's anatomy. Recent developments in computer vision methods have been applied to a variety of clinical applications to improve the view of the patient's anatomy and to facilitate the performance of neurosurgical

procedures. Because no single imaging modality offers real-time physiological and anatomical information, IGN systems use a wide range of imaging modalities for measuring brain shift in the operating room including pre-operative MRI, intra-operative MRI (iMRI), and intra-operative ultrasound images (iUS) (Delorenzo et al., 2010; Gerard, Kersten-Oertel, Hall, Sirhan, & Collins, 2020).



Figure 1.1: Cranial navigation using the Brainlab image-guided system with optical tracking and surgical displays. Courtesy of Brainlab AG<sup>1</sup>.

The key step of image-guided systems for neurosurgery is the generation of 3D pre-operative image data merged with the patient's anatomy by registration. If the registration is accurate, the surgeon can work in the mathematical space (cartesian coordinate system) of the brain image that is the same as the physical space under optimum conditions. Fusion of MRI and iUS can present perfect imaging modalities for planning, guiding, and monitoring neurosurgery, due to their excellent visualization of the brain tissues, tumor tissues, its sub-structure, and surrounding tissues. Nevertheless, the acquired MRI and iUS imaging datasets still require sophisticated image processing algorithms, such as segmentation techniques to accurately extract the required information about a brain tumor.

<sup>1</sup> Brainlab AG, Munich, Germany; <http://www.brainlab.com/>

Recent developments in deep neural networks (DNNs) have demonstrated their effectiveness for processing and analyzing medical images, including those associated with brain tumor segmentation (Baxter, Gibson, Eagleson, & Peters, 2018; Havaei et al., 2017; F. Isensee, Jaeger, Kohl, Petersen, & Maier-Hein, 2021; Luo et al., 2021; Tajbakhsh et al., 2020), image registration (Sedghi et al., 2021), and image classification (Apostolopoulos & Mpesiana, 2020; Mahapatra, Bozorgtabar, & Ge, 2021). Hence, further developments in this technological trend for MRI and/or iUS-guided systems will significantly contribute to leveraging overall neurosurgery performance.

Nevertheless, most machine learning and/or deep learning techniques are still under development for deployment in the clinical field (Angelov, Soares, Jiang, Arnold, & Atkinson, 2021; Xie et al., 2021). This is due to the underlying “black-box” nature of the deep learning methods which are often characterized by the lack of human-like explainable decisions. In addition, these models have a substantial number (within millions) of extracted feature maps in each internal layer which are assumed to contain meaningful information about the input problem and its possible solution. This makes fully understanding deep learning methods highly problematic even for professional experts. Thus, the application of such “black box” models in highly sensitive medical applications is very limited (Angelov et al., 2021; G. Yang, Ye, & Xia, 2022).

## 1.2 Objectives of the Thesis

The focus of this thesis is set on the development of an image-guided system for assisting neurosurgery using deep learning methods. The developed IGN platform incorporates novel and robust methods to enable pre- and intra-operative assistance for brain tumor interventions. Specifically, Fig. 1.2 summarizes the four main modules of the developed IGN system identified with their underlying research questions. Hence, this research work aims to achieve the following main tasks:

- Automatically define the boundary of brain tumors accurately
- Registration of pre-operative MRI to iUS images for brain shift compensation
- Interpreting the developed DNNs using Explainable AI
- Development of an intuitive neuronavigational display in the operating room

For approaching the four above aspects of the thesis, different modules are developed and integrated into an IGN platform. The contributions can be summarized for each of these tasks as follows:

*Automatically define the boundary of brain tumors accurately*

*How to correctly delineate brain tumor boundaries from adjacent healthy structures using automatic deep-learning models?* This module aims at developing a new fully automated MRI brain tumor segmentation based on encoder-decoder deep learning models, including the following contributions:

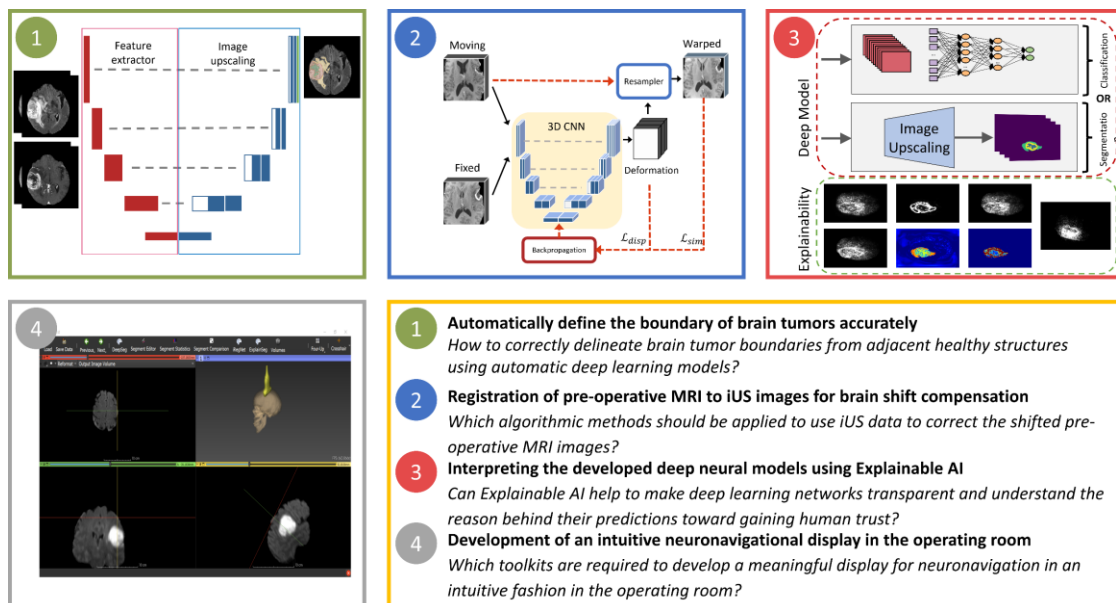


Figure 1.2: Overview of the contribution areas and the underlying research questions of this Ph.D. thesis.

- The development of a generic modular architecture for brain tumor segmentation, called DeepSeg, enables the integration of new network architectures for the segmentation of brain tumors in MRI.
- A detailed ablation study of the state-of-the-art deep learning models highlighting the computational performance during training and prediction processes.
- The improved 3D version of DeepSeg utilizes 3D convolutions over all slices for more robust and accurate results. Moreover, all the available MRI modalities (T1, T1Gd, T2, and FLAIR) are employed so that the brain tumor sub-regions could be detected in comparison with the whole tumor only in the original DeepSeg method.
- Presenting the ensemble convolutional neural networks (CNNs) for glioma segmentation using multimodal MRI, based on the state-of-the-art encoder-decoder methods.
- Validating the proof of concept to apply various deep learning models for assisting the clinical procedures of brain tumor surgery using MRI modality on the Multimodal Brain Tumor Segmentation Challenge (BraTS) datasets.

#### *Registration of pre-operative MRI to iUS images for brain shift compensation*

*Which algorithmic methods should be applied to use iUS data to correct the shifted pre-operative MRI images?* In this module, iRegNet is proposed as an automated deformable MRI to iUS registration workflow using deep learning, intending to provide considerably improved robustness and computational performance toward brain shift compensation for assisting neurosurgeons intra-operatively. The contributions of this chapter include the following advancements:

- Presenting the two-step workflow of the proposed iRegNet method, which first takes two input volumes: pre-operative MRI (the moving image) and the iUS (the fixed image) as input. Then, the DNN generates the corresponding deformation field and corrects the brain-shifted MRI volume using deformable registration.

- Utilizing the truth warped images as the target of the registration rather than the fixed image in the conventional registration method. This provides more precise guided information for training iRegNet and, therefore, contributes to the overall accuracy of the registration results.
- Comprehensively evaluating MRI to iUS registration results using two numerical metrics: the mean target registration errors (mTRE) and the computational processing time. Notably, almost two 3D MRI-iUS pairs per second can be registered on the same graphical processing unit (GPU) using the proposed approaches.
- Competitive registration results when applying the proposed methods on unseen MRI-iUS cases, which evidences their general applicability.
- A detailed comparison with the state-of-the-art non-learning- and learning-based registration algorithms across multi-site volumes.
- Qualitative analysis is done by two experienced neurosurgeons to highlight the clinical applicability of the proposed framework in neurosurgical guidance.

#### *Interpreting the developed deep neural models using Explainable AI*

*Can Explainable AI help to make deep learning networks transparent and understand the reason behind their predictions toward gaining human trust?* Overall, the main focus of recent explainable artificial intelligence (XAI) research in medical image segmentation has been on integrating visual interpretability without considering the clinical evaluation of the resultant visualizations. Besides, little attention has been paid to the incorporation of medical knowledge into the decision approach made by artificial intelligence-based (AI) models. Moreover, the decisions of these models must be consistent with the clinical knowledge to gain the trust of medical professionals and encourage them to adopt AI-based systems. To cope with these challenges, the contributions are divided four-fold:

- An effective NeuroXAI diagnosis generator has been developed to extract 2D and 3D explanations to assist clinicians to understand and trust the performance of brain deep learning algorithms in clinical procedures. The NeuroXAI method is post-hoc and can be applied to any brain CNN model without modifying the underlying architecture or performance degradation.
- NeuroXAI included seven state-of-the-art backpropagating XAI techniques, namely Vanilla Gradient, Guided Backpropagation, Integrated Gradients, Guided Integrated Gradients, SmoothGrad, Grad-CAM, and Guided Grad-CAM, for generating 2D and 3D visual interpretations of CNN output.
- A hybrid CNN-Transformer architecture, called TransXAI, is proposed for the segmentation of brain tumors, which combines high-resolution local representations from CNN and the long-range dependency captured by Transformers.
- A comprehensive evaluation of the proposed NeuroXAI framework demonstrated promising explanation results for two showcases of MRI classification and segmentation of brain tumors.
- Evaluation of the proposed TransXAI framework on the multimodal brain tumor segmentation dataset demonstrates its effectiveness, superiority, and robustness.



- Explainability-driven evaluation by clinical experts showed that the proposed approaches increase surgeons' trust in deep learning systems by providing evidence linked to the results of the TransXAI from the surgical point of view.

#### *Development of an intuitive neuronavigational display in the operating room*

*Which toolkits are required to develop a meaningful display for neuronavigation in an intuitive fashion in the operating room?* In this module, a meaningful display is developed which visualizes all required surgical information simply and intuitively in the operating room with the following contributions:

- Development of an IGN toolkit, namely DeepIGN, as an AI-powered neurosurgical navigation application using a set of open-source platforms.
- DeepIGN allows users, developers, and clinical researchers to employ recent deep learning advances in brain imaging research and image-guided interventions.
- The proposed system is designed to integrate effortlessly with the commercially available neuronavigation systems to facilitate clinical translation.
- The prototype research system was evaluated in a simulated clinical environment using a custom-made tissue-mimicking brain phantom which evaluated the accuracy of the ultrasound navigation and registration.
- As a proof of concept, the system was designed for skull base surgery although it could be used during other cranial neuro-oncology surgeries where multi-modal intra-operative guidance is desired.
- To the best of our knowledge, this is the first publicly available open-source AI toolkit for guiding neurosurgery and brain cancer research using self-contained deep learning methods.

## **1.3 Thesis Roadmap**

The thesis is structured similarly to the research pipeline of Figure 1.2, as follows:

- Chapter 2 provides the medical fundamentals including medical imaging techniques for neurosurgery in addition to brain tumor diagnosis and treatment options.
- Chapter 3 gives an overview of the fundamentals of deep learning that are essential for the remainder of this thesis is presented. Furthermore, technical background on image-guided neurosurgery, challenges, and limitations are given.
- Chapter 4 details algorithmic developments in the analysis of brain medical images and the development of brain tumor segmentation methods along with the comparison to state-of-the-art methods.
- Chapter 5 presents the framework of the proposed multimodal brain imaging registration pipeline with a detailed description, evaluation, and a comparison to existing work in the literature qualitatively and quantitatively.
- Chapter 6 explains the information flow process in the internal layers of DNNs, following the input layers to the output predictions, and highlights the attention maps in the process. This includes two showcases of the most widely applied brain tumor applications of classification and segmentation.

- Chapter 7 describes the proposed neurosurgical display for providing all required surgical information, and patient data as well as its application in the simulated operating room.
- Chapter 9 concludes this work with a review and an outlook on the many areas that remain to be worked on.

A complete list of publications that resulted from this Ph.D. research is provided in Appendix B.

# 2 Medical Background

---

## Contents

<b>2.1</b>	<b>Medical Imaging Techniques</b> -----	<b>10</b>
2.1.1	Magnetic Resonance Imaging-----	10
2.1.2	Ultrasound Imaging -----	12
<b>2.2</b>	<b>Brain Tumors</b> -----	<b>13</b>
2.2.1	Diagnosis -----	14
2.2.2	Treatment-----	16

This chapter first introduces the basic imaging modalities used for brain tumor surgery in Section 2.1. Then, the definition of brain cancer in addition to brain cancer diagnosis and treatment options are presented in Section 2.2.

## 2.1 Medical Imaging Techniques

Before and during brain surgery, a variety of imaging modalities are used for guiding neurosurgeons and measuring brain shift including MRI, and ultrasound (US) (Coburger et al., 2014; Delorenzo et al., 2010). This section briefly describes the basic concepts of the medical imaging modalities used in this thesis.

### 2.1.1 Magnetic Resonance Imaging

MRI is a non-invasive imaging technique that uses strong magnets and low-energy radiofrequency waves (RF) to produce tomographic images of the human body. Since the discovery of nuclear magnetic resonance (NMR) by Lauterbur et al (Lauterbur, 1973), several articles investigated the basic principles of MRI (Liang & Lauterbur, 2000; Pagani, Bizzi, Di Salle, De Stefano, & Filippi, 2008; Rajan, 1997). MRI provides superior image contrast and high-detailed visualization of soft tissue while maintaining a safe procedure since it does not require ionizing radiation to obtain images. MRI is the gold standard for the diagnostic visualization of brain gliomas including their sub-regions (Ellingson et al., 2017).

Based on quantum physics, the production of a magnetic field is based on a simple classical model in which certain nuclei spin around their axes after excitation by an external magnetic field (Pooley, 2005). Most clinical imaging applications of MRI utilize the Hydrogen nucleus which is the most abundant atom in the body because of its high sensitivity to magnetic fields (van Geuns et al., 1999). Under normal circumstances, hydrogen nuclei are spinning randomly around their axis making them act like tiny magnets, as shown in Fig. 2.1 (a). However, upon application of a strong magnetic field, hydrogen protons adopt one of the following possible states: spin-up, which is in the direction of the external field, and spin-down, as presented in Fig. 2.1 (b). Lining up with the magnetic field has the lower energy while the other orientation has the higher energy. This net magnetization generated by the parallelly aligned protons becomes the source of the MRI signal which is used to produce MRI images. Formally, the resonant frequency  $F$  in  $MHz$  is calculated by the Larmor equation:

$$F = \gamma B_0 \tag{2.1}$$

where  $\gamma$  is the gyromagnetic ratio of the nucleus and  $B_0$  is the strength of the applied magnetic field. This frequency is also called the Larmor frequency.

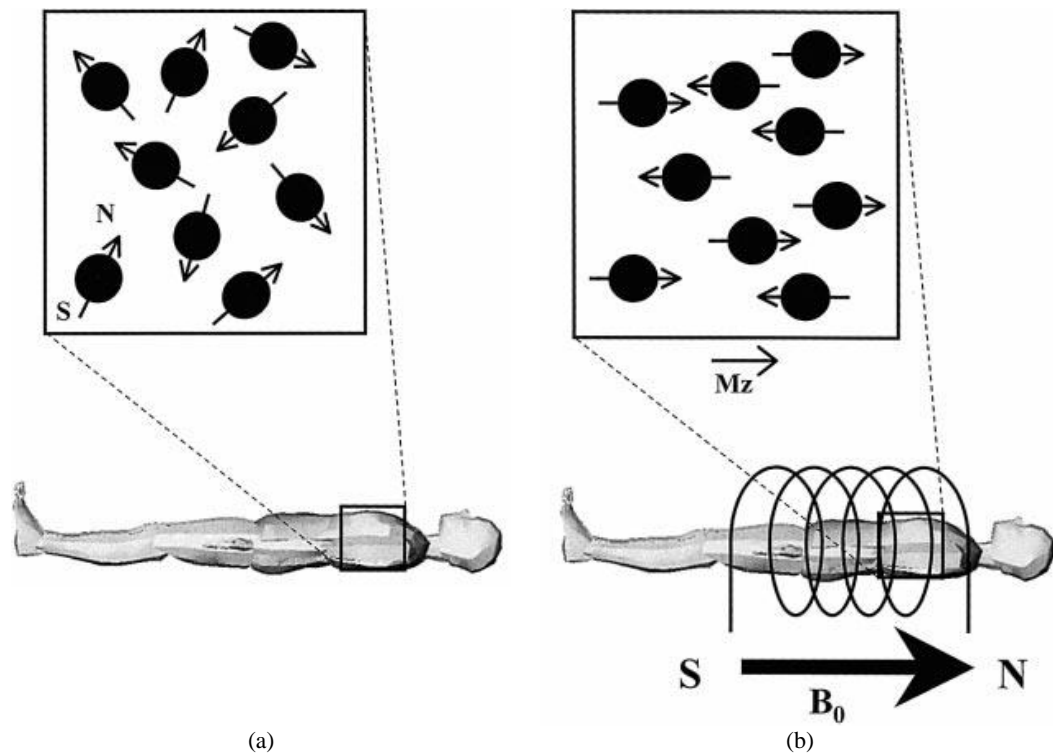


Figure 2.1: Hydrogen nuclei spinning phenomena. (a) In the absence of a magnetic field, the magnetic moments of the nuclei are distributed randomly and thus the net magnetization factor is zero. (b) In the case of a strong external magnetic field, the spinning nuclei parallel (spin up) or antiparallel to the external field (spin down) ( $B_0$ ) (van Geuns et al., 1999).

Figure 2.2 gives an example of an MRI scanner at the Department of Neurosurgery, University of Ulm. Typically, RF waves are transmitted by a transmitting coil, such as the body coil in Fig. 2.2, with the energy of exactly the Larmor frequency of the protons (refer to Equation (2.1)). After the energy is absorbed by the hydrogen nuclei that have the matching Larmor frequency, the net magnetization rotates away from the direction parallel to the imaging magnet (called longitudinal direction). After the RF pulse is turned off, the net magnetization starts to grow back in the longitudinal direction (called T1 relaxation). Depending on protons associated with different soft tissues, the rate at which longitudinal magnetization decays over time varies resulting in the contrast in T1. Similarly, T2 is obtained using transverse relaxation, based on the spins precessing around the RF frequency source generator. Moreover, other MRI modalities are acquired for the precise localization of brain tumors, including T1Gd and FLAIR sequences.



Figure 2.2: MRI Scanner at Ulm University Hospital on a 1.5 Tesla Magnetom Symphony (Siemens Medical, Germany) with a 12-channel head coil<sup>1</sup>.

### 2.1.2 Ultrasound Imaging

US imaging is another non-invasive imaging technique that uses pulsed sound waves at higher frequencies than those detectable by human hearing (greater than 20 kHz) (Novelline & Squire, 2004). In 1942, Dr. Dussik published the use of US transmission attenuation data through the head to diagnose brain cancer (Dussik, 1942). Since that time, US has been adopted widely in many image-guided therapies with the help of evolving multidimensional, multimodality technologies (Sastry et al., 2017; Stoll, 2014). This is primarily due to its high frame rate, portability, low operational costs, and safety. Over the years, a wide variety of US systems have been developed and within this thesis, the B-Mode US ("B" for brightness) is used.

Figure 2.3 (a) shows a typical US scanner and the output B-Mode image with the characteristic of brain tumor type and malignancy is shown in Fig. 2.3 (b). Typical medical US systems utilize the piezoelectric effect, and the US transducer probe emits pulses at a frequency of 1–20 MHz. These pulsed echo signals are transmitted through biological tissues and reflected differently depending on their acoustic impedance. US signals also carry a great deal of information includ-

---

<sup>1</sup> MRI scanner, Ulm University Hospital, Germany; <https://www.uniklinik-ulm.de/>

ing tissue properties which are combined to generate 2D grayscale images. The US modality can assess the dynamic aspects of interventions in real-time, such as tissue deformation and instrument motion, which makes it the preferred modality for intra-operative navigation of brain tumor surgery (Coburger et al., 2014; Delorenzo et al., 2010; Sastry et al., 2017; Stoll, 2014). The use of 3D reconstructed US is introduced in (Miller, Benes, & Sure, 2011), allowing accurate and effective navigation.



Figure 2.3: An example of a B-mode US scanner (a) and its output images (b) showing the brain tumor which appears in bright grayscale in axial, sagittal, and coronal views (Shetty & Moiyadi, 2016; Xiao, Fortin, Unsgard, Rivaz, & Reinertsen, 2017).

## 2.2 Brain Tumors

Brain tumors are one of the leading causes of death for cancer patients, especially children, and young people (Weller et al., 2021). The American Cancer Society reported that 23,820 new brain cancer cases in the United States were discovered in 2019 (Siegel, Miller, & Jemal, 2019).

Brain tumors can be categorized into two types: Primary brain tumors originate in the brain cells, and secondary brain tumors developed through the spreading of malignant cells from other parts of the body to the brain.

Malignant glioma is the most common primary brain tumor representing 75% of primary brain malignancies among adults (Holland, 2001; Lapointe, Perry, & Butowski, 2018). Gliomas normally start at the glial cells of the brain, but they can also invade the surrounding tissues due to their infiltrative nature. HGGs are the most common and aggressive type with a median survival rate of one to two years (Baid et al., 2021; Buckner, 2003; Louis et al., 2020). Another slower-growing LGGs such as astrocytoma have a slightly longer survival time.

### 2.2.1 Diagnosis

The prognosis of glioma cells remains very challenging despite great advancements in neurosurgical therapeutics. The main reason behind that is the highly infiltrative nature and heterogenous appearance of malignant gliomas. Pre-operative imaging is a key factor in the accurate definition of the tumor tissue and its peritumoral regions, and therefore, has a great impact on the overall survival rate (Coburger et al., 2016). In particular, MRI is the modality of choice for the evaluation of brain tumors, identification of anatomic detail of normal brain structures, peritumoral edema, and detection of tumor-infiltrated regions (Coburger et al., 2016; Pope & Brandal, 2018; Upadhyay & Waldman, 2011).

Pre-operative MRI provides superior image contrast and high-resolution soft tissue and thus can be essentially used for IGN (Miner, 2017). Multimodal MRI protocols such as T1, T1Gd, T2, and FLAIR images provide information to neurosurgeons and can be valuable in diagnostics. The use of a gadolinium-based contrast agent in T1Gd allows for the detection of areas where the blood-brain barrier is compromised. In general, gliomas are hypointense on T1 images and hyperintense on T2 images, as shown in Fig. 2.4. The presence of peripheral contrast enhancement, central necrotic areas, intra-tumoral hemorrhages, ill-defined infiltration of surrounding brain tissue, and extensive perifocal edema is commonly seen in aggressive lesions and raises the possibility of high-grade glial lesions or GBM. On the other hand, non-enhancing lesions may represent LGGs.

Although iMRI offers soft-tissue contrast and diffusion-weighted imaging, it requires long scan times and may be associated with high costs. Further, the iMRI scanner limits the physician's access to the operative field and special surgical tools are required due to the strong magnetic field which can affect electronic systems. Alternatively, iUS is portable, inexpensive, requires little preparation, and provides fast data acquisition. Though iUS can visualize interior soft tissue and structures, it has difficulty of low-quality imaging through bones, and its high dependency on inter-operator interpretation may result in image interpretation inconsistency.

Typically, a manual definition of the tumor location and visual examination are used for brain cancer diagnosis. However, interpreting this large amount of data during neurosurgery is a demanding task and an appropriate visualization of lesion structure apart from healthy brain tissues is crucial (Siekman, Lothes, Konig, Wirtz, & Coburger, 2018). Therefore, IGN, the integration of medical imaging modalities with brain surgery has become an essential tool for



assisting neurosurgeons to overcome the above challenge (Coburger & Wirtz, 2019). Figure 2.5 demonstrates the placement of iUS over pre-operative MRI images.

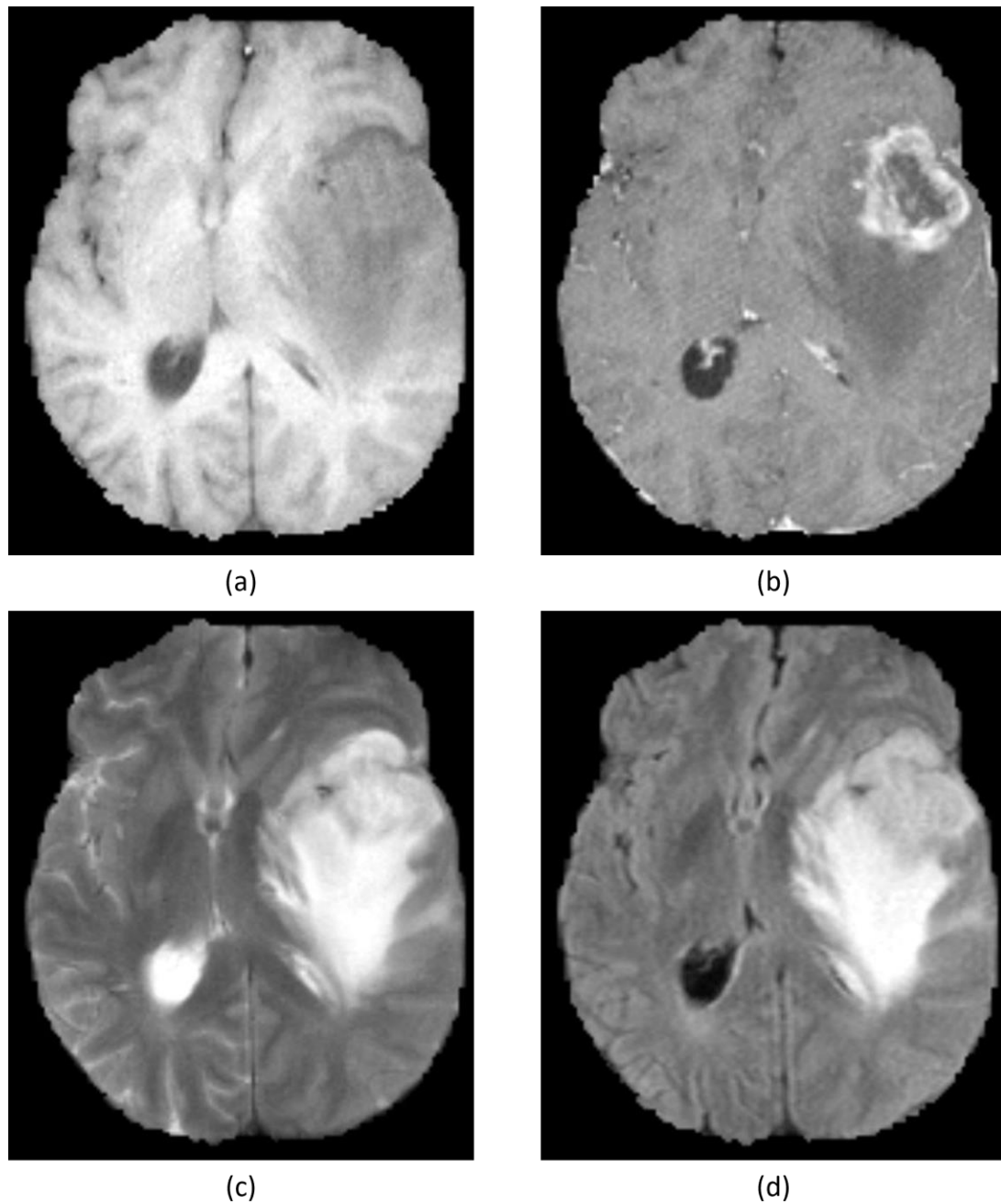


Figure 2.4: Overview of different MRI modalities showing the characteristics of brain Glioma from the BraTS dataset (Menze et al., 2015). Image patches show the different modalities of T1 (a), T1Gd (b), T2 (c), and FLAIR (d).

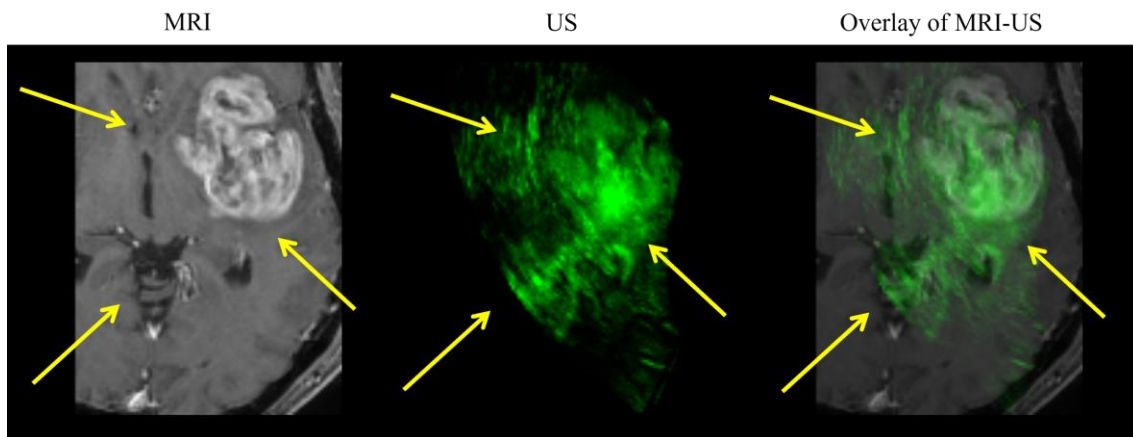


Figure 2.5: Fusion of iUS images (*green*) over pre-operative MRI (*grey*) of a sample patient from the BITE dataset (Mercier, Del Maestro, et al., 2012). From left to right: FLAIR MRI, B-Mode US image, and the corresponding overlay of MRI and US images.

### 2.2.2 Treatment

Treatment of malignant brain tumors is still among the most difficult challenges in cancer remedies (Lapointe et al., 2018). Treatment protocol can vary based on various factors such as the type, size, location, and age of the patient. Neurosurgical intervention is the first and sometimes the only therapy for many types of gliomas, while a combination of surgical removal, radiation therapy, and chemotherapy are the most common forms of treatment (Jain, 2018; Pala et al., 2021).

Brain surgery or neurosurgery aims at the complete or partial resection of the brain tumor by removing as much as possible of the malignant cells without damaging surrounding healthy tissue. Radiation therapy uses high energy radiation to kill cancer cells and shrink the tumor, while chemotherapy uses drugs to kill cancer cells. In some cases, other treatments like immunotherapy, targeted therapy, and proton therapy may also be used, as shown in Fig. 2.6 (Mustaf, Sali, Illzam, Sharifa, & Nang, 2018). It is important to work with a team of specialists such as neurosurgeons, oncologists, and radiation therapists to determine the best course of treatment for each individual patient.

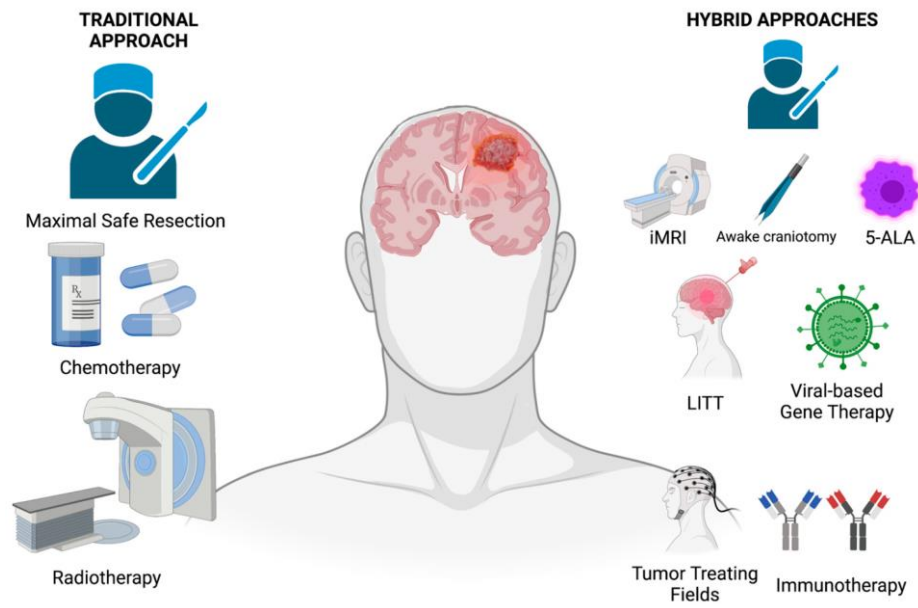


Figure 2.6: Overview of traditional approaches as well as new hybrid strategies for glioblastoma therapy (Shah & Heiss, 2022).



# 3 Technical Background

---

## Contents

<b>3.1</b>	<b>Deep Learning</b>	<b>20</b>
3.1.1	Artificial Neural Networks	20
3.1.2	Convolutional Neural Networks	22
3.1.3	CNN Architectures	23
3.1.4	Transformers	26
3.1.5	Deep Network Training	27
<b>3.2</b>	<b>Image-guided Neurosurgery</b>	<b>28</b>
3.2.1	IGN Components	28
3.2.2	IGN Software	29
<b>3.3</b>	<b>Challenges and Limitations</b>	<b>30</b>
3.3.1	Neurosurgical Challenges	30
3.3.1.1	Precise Tumor Localization	30
3.3.1.2	Brain shift	31
3.3.1.3	Inter-rater Variability	31
3.3.2	Technical Challenges	31
3.3.2.1	Multiple Modules Integration	31
3.3.2.2	Deep Networks Transparency	32
3.3.2.3	Neuronavigational Challenges	32

This chapter provides an overview of the fundamental background knowledge of the technical topics covered in this dissertation. This includes an introduction to deep learning concepts and the state of modern deep learning applications in Section 3.1 and image-guided neurosurgery in Section 3.2. Finally, Section 3.3 summarizes the challenges that face neurosurgical procedures and brain surgery. Parts of the section regarding image-guided neurosurgery have previously appeared in (Ramy A. Zeineldin, Karar, Burgert, & Mathis-Ullrich, 2023; R. A. Zeineldin, Karar, Coburger, Wirtz, & Burgert, 2020; Ramy A. Zeineldin, Karar, Mathis-Ullrich, & Burgert, 2022a; Ramy A. Zeineldin, Alex Pollok, et al., 2022).

## 3.1 Deep Learning

Deep learning, a subfield of AI, has gained a lot of interest over the last decade thanks to its successful applications for processing and analyzing medical images including those associated with brain tumor segmentation (F. Isensee et al., 2021; Liew, Lee, Lan, & Tan, 2021; L. Zhu et al., 2022), image registration (Sedghi et al., 2021), and image classification (Apostolopoulos & Mpesiana, 2020; Le et al., 2021; Mahapatra et al., 2021). Recent advances in computational hardware, availability of large-scale training datasets, and open-source software have led to a renewed interest in analyzing brain cancer, its causes, and its various development phases (Chan, Hadjiiski, & Samala, 2020; Lynch & Liston, 2018). This section describes the basic principles of deep learning including artificial neural networks, convolutional neural networks, transformers, and network training used within this thesis.

### 3.1.1 Artificial Neural Networks

Deep learning is mainly based on artificial neural networks (ANNs) which have been primarily introduced to emulate the biological neural networks of the human brain. The term deep learning refers to a large number of multiple levels of processing layers that can learn representations of high-dimensional data (LeCun, Bengio, & Hinton, 2015). A neural network is a collection of simple processing nodes, called neurons, that are densely interconnected. Figure 3.1 shows a typical neural network that consists of multiple connected layers, namely an input layer, an output layer, and (optionally) multiple middle layers (called also hidden layers). The information flow in the networks in a feedforward manner, or in other words in one direction from the input layer to the output layer through the hidden layers.

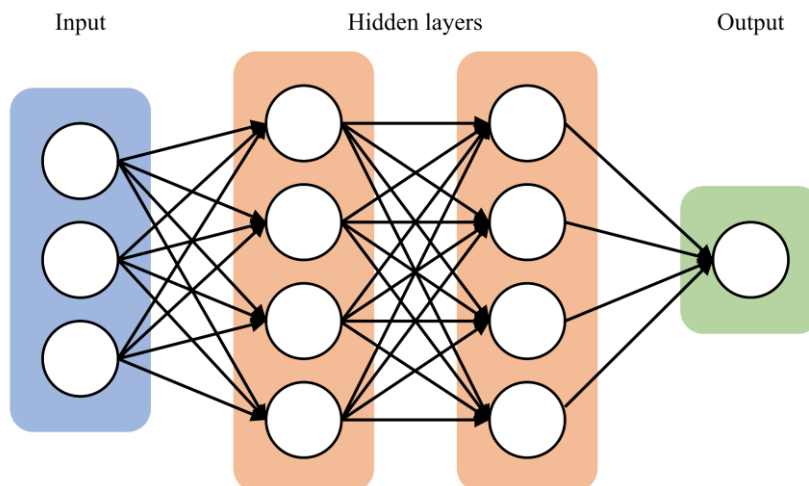


Figure 3.1: Sample three-layer neural network with an input layer (*blue*), two hidden layers (*orange*), and an output layer (*green*) that are fully connected only in the forward direction.

The basic and simplest neural network is a single neuron (or perceptron) which has a minimum of one input (dendrites) and a single output (axon) (Goodfellow, Bengio, & Courville, 2016). Figure 3.2 presents the biological brain neuron in addition to its artificial mathematical model. Mathematically, the artificial neuron output is determined by an input function  $g(x, w)$  dependent on the input signals  $x$ , weights  $w$  based on the multiplicative interaction with the dendrites of other neurons, and constant bias term  $b$  (Habibi Aghdam & Jahani Heravi, 2017).

$$g(x, w) = g(\sum_{i=1}^n x_i w_i + b) \quad (3.1)$$

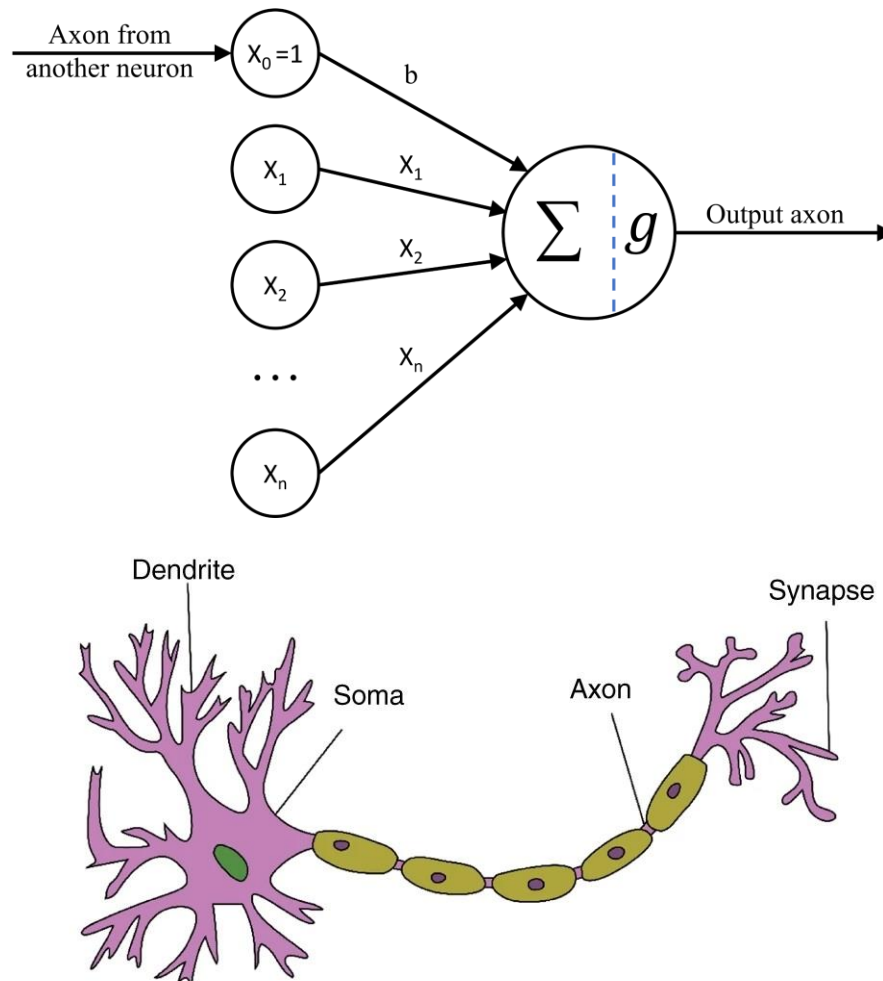


Figure 3.2: Concept of artificial neural networks, a perceptron (*up*), which was originally inspired by biological neural systems, a biological neuron (*down*) (Zhang, Yu, Barbiero, Wang, & Gu, 2019).

The axon conveys the electrical pulse that eventually branches out at multiple synapses which are connected to other neurons. The synaptic strengths or weights are learnable and control the strength of the electrical signals and their direction can be stimulating (positive influence) or inhibiting (negative influence) (Kandel et al., 2000). If the sum of the accumulated inputs is above a certain threshold, the neuron sends out an electrical spike of a certain intensity along its

axon. The function  $g$  is the respective activation function of the neuron, which represents the frequency of these spikes. Common choices for activation functions of neural networks are the sigmoid function (Equation (3.2)) and the hyperbolic tangent function (Equation (3.3)).

$$g(x) = \sigma(x) = \frac{1}{1+e^{-x}} \quad (3.2)$$

$$g(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3.3)$$

On the other hand, the Rectified Linear Unit (ReLU) function (Equation (3.4)) (Dahl, Sainath, & Hinton, 2013) and its variants such as leaky ReLU (LReLU) (Equation (3.5)) (Maas, Hannun, & Ng, 2013) are more popular nonlinearity functions in DNNs with improved performance.

$$g(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases} \quad (3.4)$$

$$g(x) = \begin{cases} 0.01x & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases} \quad (3.5)$$

### 3.1.2 Convolutional Neural Networks

CNNs are a specific type of multi-layer perceptron. A common problem of the single perceptron is that it can only address linear problems and multi-layer perceptron is needed for more complex and high dimensional data encountered in advanced medical image analysis tasks. Due to its inherent fully connected layers, the number of weights increases very fast in the hidden layers making it impossible to handle with modern computational resources. Thanks to the convolutional kernels, CNNs can effectively handle large and dense amounts of data such as images: sparse connection, and parameter sharing, as illustrated in Figure 3.3 (Lecun, Bottou, Bengio, & Haffner, 1998). Sparse connection means that neurons of layer ( $\ell$ ) are only connected to neurons of the previous layer ( $\ell-1$ ) within a fixed local neighborhood rather than being connected to all previous neurons. Another important modification is that CNNs share the same parameters for all neurons within a hidden layer. Both modifications help to significantly improve computational efficiency by reducing the number of parameters and consequently the network complexity (Goodfellow et al., 2016; LeCun et al., 2015).

The architecture of a typical convolutional network (Fig. 3.4) consists of a series of three basic types of layers: convolutional layers, pooling layers, and fully connected layers (Lecun et al., 1998). Convolutional layers are composed of learnable filters, also called channels, to capture sufficient spatial information from the previous layer. Each filter is characterized by its kernel size which defines the corresponding local neighborhood, also known as the receptive field. The output of a convolutional layer is often referred to as feature maps which is the locally weighted sum over all channels via element-wise addition.

Pooling is another important operation of the CNN which reduces the size of the feature maps and therefore increases the receptive field. Pooling layers merge similar elements within the feature map semantically into one by a simple mathematical operation, typically the mean, minimum, or maximum of the local patch. The maximum operation, known as max pooling, is



selected for this thesis. Through convolution and pooling, it is possible to train larger and deeper CNNs for more complex computer vision and medical tasks with increased computational speed. Finally, the output is fed into one or more fully connected layers and the non-linear ReLU activation.

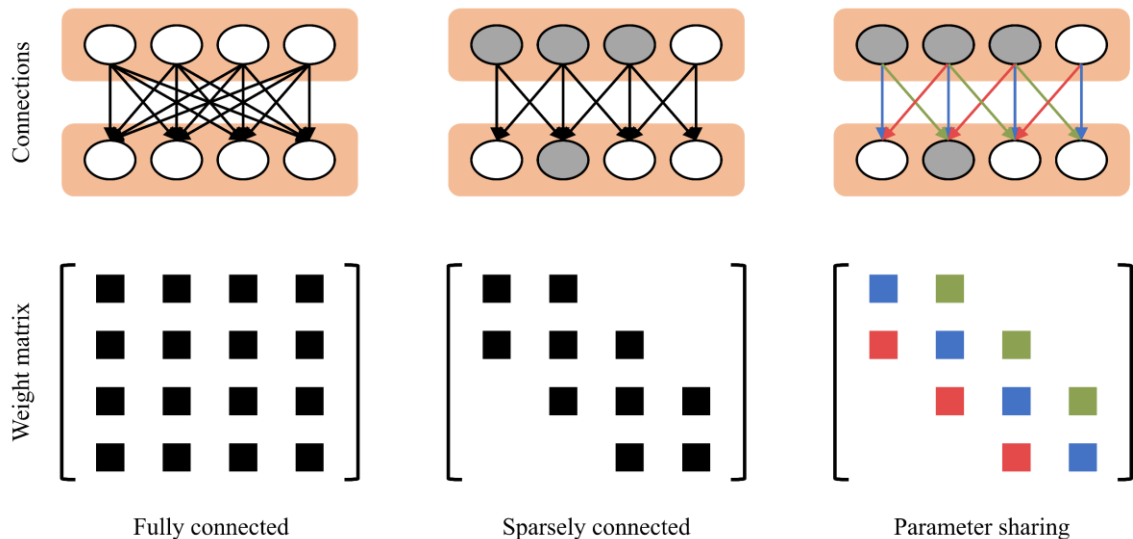


Figure 3.3: Two ideas employed by a convolutional layer to reduce the number of parameters and thus the network complexity: sparse connection and parameter sharing. The first row shows layer interconnection for a fully connected layer as well as the two enhancements by the convolution operation. Their respective weight matrices are presented in the second row.

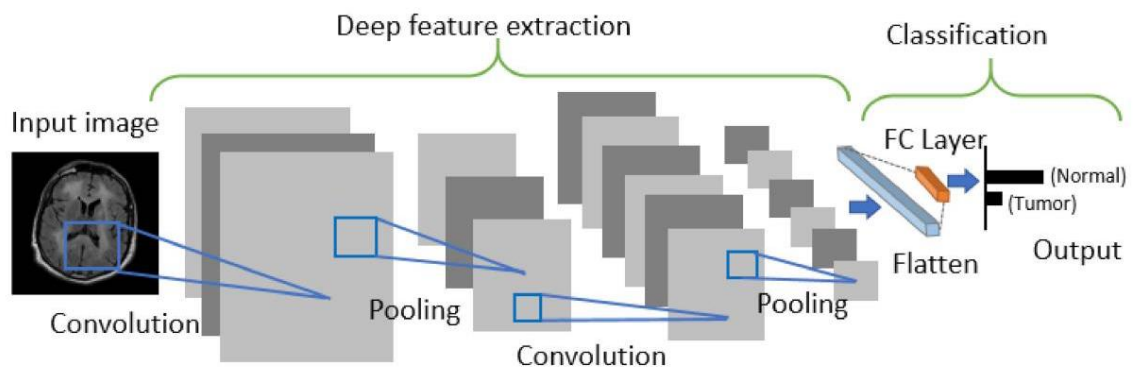


Figure 3.4: Architecture of a convolutional network for image classification (Kang, Ullah, & Gwak, 2021).

### 3.1.3 CNN Architectures

CNNs have proved during the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) (Russakovsky et al., 2015) their ability to accurately detect and localize different types of objects. In 2012, an advanced pre-trained CNN model called AlexNet (Krizhevsky, Sutskever, & Hinton, 2017) showed the best results in the image classification challenge. Then, other CNN models have dominated the ILSVRC competition; namely Visual Geometry Group Network

(VGGNet) (Simonyan & Zisserman, 2014), Residual Neural Network (ResNet) (He, Zhang, Ren, & Sun, 2016), Dense Convolutional Network (DenseNet) (Huang, Liu, Van Der Maaten, & Weinberger, 2017), Xception (Chollet, 2017), MobileNet (Howard et al., 2017), NASNet (Zoph, Vasudevan, Shlens, & Le, 2018), and MobileNetV2 (Sandler, Howard, Zhu, Zhmoginov, & Chen, 2018). Moreover, CNN methods have been applied to perform MRI tumor segmentation (Havaei et al., 2017; naceur, Saouli, Akil, & Kachouri, 2018).

**VGGNet** (Simonyan & Zisserman, 2014) is proposed by the Visual Geometry Group from the University of Oxford and is the winner of the ILSVRC 2014 in the localization task. It is chosen to be the baseline model because of its simplicity, consisting only of small  $3 \times 3$  convolutional layers and max-pooling layers for the downsampling process followed by two fully connected layers for feature extraction.

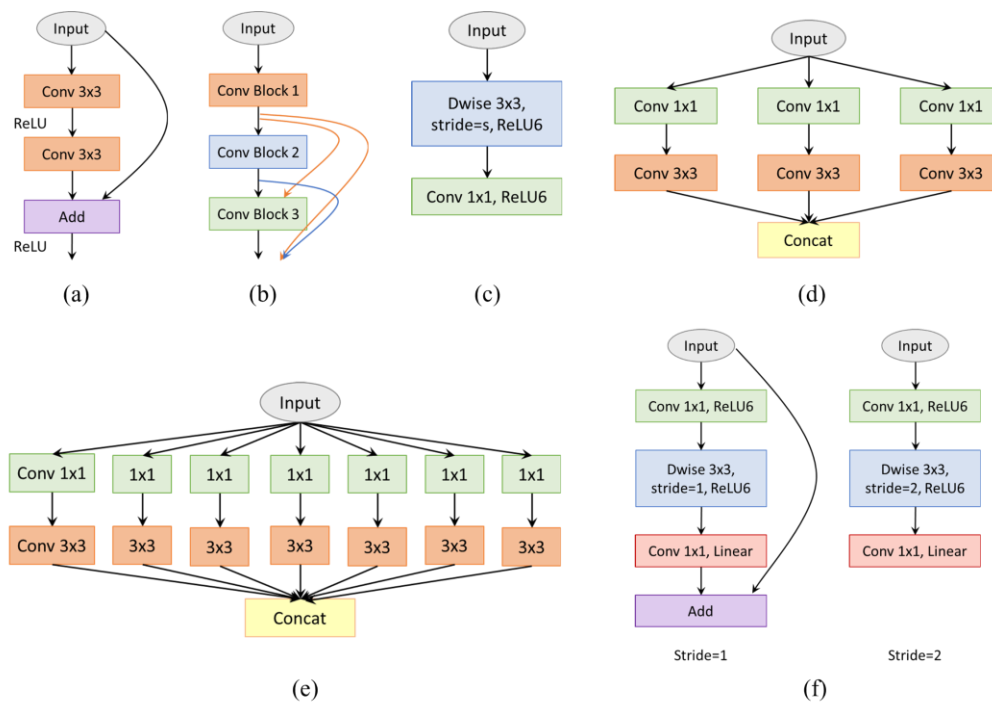


Figure 3.5: Comparison of the basic blocks for different convolutional networks. (a) the residual block of ResNet; (b) a 3-layer dense block, (c) a depth-wise based module of MobileNet, (d) a simplified Inception module, (e) an extreme module of Inception (Xception module), and (f) MobileNetV2 blocks with two stride values.

In fact, increasing the neural network layer would increase the accuracy of the training phase, however, there is a significant problem with this approach; for example, vanishing gradients (Glorot & Bengio, 2010) cause the neural network accuracy to saturate and then degrade rapidly. In ILSVRC 2015, a novel micro-architecture called **ResNet** (He et al., 2016) was introduced to solve this exploding behavior. The ResNet consists of residual blocks as shown in Fig. 3.5 (a), and each block consists of the original two convolutional layers in addition to a shortcut connection from the input of the first layer to the output of the second layer. By employing skip connections to the DNN, neither additional parameters nor computational complexity are added

to the network. Owing to this approach, they are able to train up to 152-layer DNN while maintaining a lower complexity than the above VGG models.

**DenseNet** (Huang et al., 2017) uses the feature map of the preceding layers as inputs into all the following layers, as depicted in Fig. 3.5 (b). This type of DNN model has  $L(L+1)/2$  connections for a CNN with  $L$  layers, whereas traditional networks would have only  $L$  connections. Remarkably, they can achieve additional improvements such as a smaller number of parameters besides the ability to scale the network to hundreds of layers.

**Xception** presents an extreme version of the Inception network (Chollet, 2017). The Inception model aimed at improving the utilization of the computing resources within the neural network through special modules. Each inception module is a multi-level feature extractor by stacking  $1 \times 1$  and  $3 \times 3$  convolutions beside each other in the same layer rather than using only one convolutional layer. The Xception, as shown in Fig. 3.5 (c), achieved a slightly better result than Inception models on ImageNet, however, it showed superior improvement when the used dataset becomes larger.

Google presented **MobileNet** (Howard et al., 2017) as an efficient lightweight network for mobile applications, as presented in Fig. 3.5 (d). Additionally, the BN is applied after each convolution followed by a ReLU activation. Then MobileNetV2 (Sandler et al., 2018) is introduced, which enhanced the state-of-the-art performance of mobile models based on inverted residual blocks as shown in Fig. 3.5 (e). These bottleneck blocks are similar to the residual block of ResNet where the input of the block is added to the output of the narrow layers. ReLU6 is also utilized, because of its robustness in low-precision computation, to remove the nonlinearities in the bottleneck layers. Although **MobileNetV2** shows a similar performance to the previous MobileNet, it uses only 2.5 times fewer operations than the first version.

Google Brain introduced **NASNet** (Zoph et al., 2018) to obtain state-of-the-art segmentation results with relatively smaller model size. The basic architecture of NASNet is made up of two main repeated blocks namely Normal Cell and Reduction Cell. The first type is consisting of convolutional layers with output features of the same dimensions, and the height and width of the other type's output are reduced by a stride of 2. ScheduledDropPath is also presented to make the model generalize well, where each path in the cell can be dropped with an increased probability over the training sequence.

A **U-Net** is a special type of CNNs that contains two main paths: a contracting encoder path and an expansive decoder path to produce full-resolution segmentation maps, as displayed in Fig. 3.6 (Ronneberger, Fischer, & Brox, 2015). U-Net architecture has become one of the most successful and popular CNN architectures for medical semantic segmentation thanks to its precise localization (F. Isensee et al., 2021). The encoder is typically a CNN consisting of consecutive two  $3 \times 3$  convolutional layers, each followed by a ReLU activation and  $2 \times 2$  spatial max pooling. Contrarywise, the decoder aims at upsampling the resultant feature map using deconvolution layers followed by  $2 \times 2$  up-convolution, a concatenation layer with the corresponding down-sampled layer from the encoder, two  $3 \times 3$  convolutions, and a ReLU. Finally, the upsampled features are then directed to a  $1 \times 1$  convolution layer to output the final segmentation map. With the help of data augmentation (Springenberg, Dosovitskiy, Brox, &

Riedmiller), U-Net achieves precise segmentation using only a few training images. Furthermore, the tiling strategy allows the model to employ high-resolution images in the training stage with low GPU memory requirements.

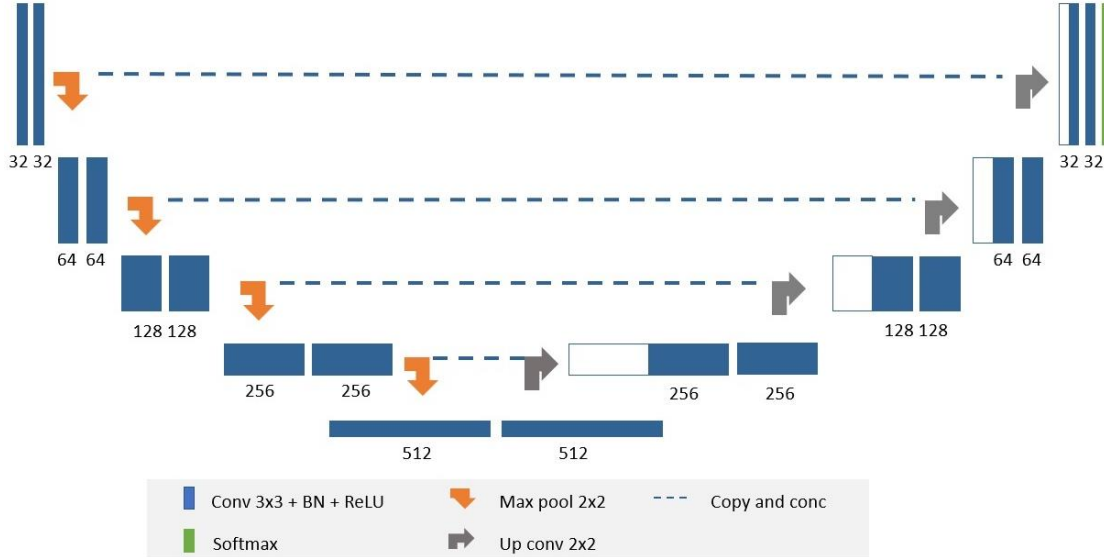


Figure 3.6: U-Net network consists of convolutional blocks (*blue boxes*), maximum pooling (*orange arrows*), upsampling (*grey arrows*), and softmax output (*green block*).

### 3.1.4 Transformers

CNN-based approaches use convolutional operations to capture local texture features by gathering information from neighborhood pixels. Although this strategy has led to performance improvements demonstrating its effectiveness in the medical field, it still exhibits limitations in capturing long-range context. On the other hand, Transformer, the state-of-the-art encoder-decoder architecture, has achieved tremendous success in the natural language processing (NLP) field (Vaswani et al., 2017). The self-attention mechanism in Transformer allows for the model of correlations among all the input tokens and hence is superior to CNN in handling long-range dependencies.

Figure 3.7 shows the model architecture of the Transformer. The Transformer encoder (Vaswani et al., 2017) is composed of a stack of  $N$  identical blocks. Each block consists of two layers; a multi-head self-attention mechanism (MSA) and a multilayer perceptron network (MLP). A LayerNorm (LN) layer is applied before each MSA and MLP layer in addition to employing residual connection around the output of each layer. Formally, the output  $z_\ell$  of the  $\ell$ -th layer in the Transformer encoder can be expressed as:

$$\hat{z}_\ell = \text{MSA}(\text{LN}(z_{\ell-1})) + z_{\ell-1} \quad (3.6)$$

$$z_\ell = \text{MLP}(\text{LN}(\hat{z}_\ell)) + \hat{z}_\ell \quad (3.7)$$

Where  $\ell \in [1, 2, \dots, L]$ ,  $\text{LN}(\cdot)$  is the layer normalization, and  $\hat{z}_\ell$  is the encoded image representation. The main restriction of using Transformers in the medical domain is that they rely heavily on large-scale training datasets since a huge training dataset is not always available (Devlin, Chang, Lee, & Toutanova, 2018; Dosovitskiy et al., 2020; Vaswani et al., 2017).

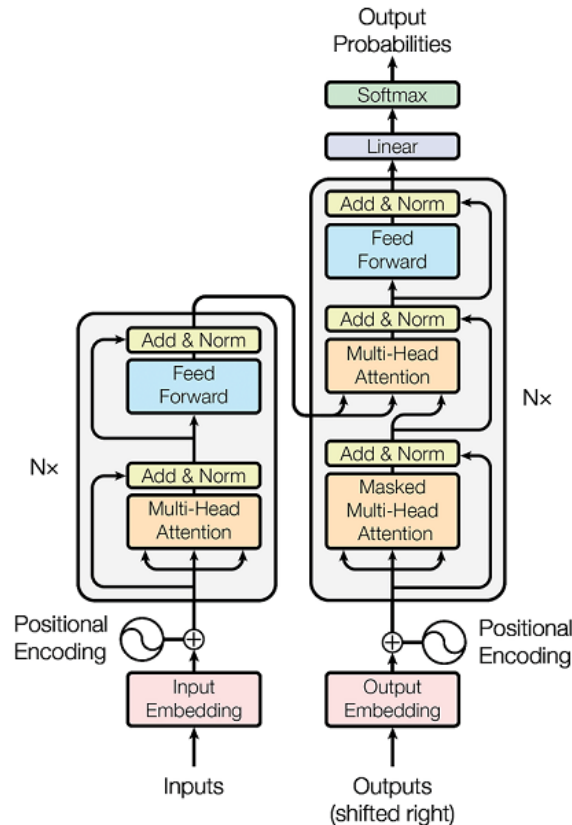


Figure 3.7: Overview of the Transformer architecture (Vaswani et al., 2017).

### 3.1.5 Deep Network Training

The training of a DNN is to find the optimized weights of the connections between the neurons with an objective function termed loss function (LeCun et al., 2015). Supervised learning is the most common type of DNN in which the network receives preassigned labels for the training data. Thereby, the neural network weights  $w$  are adjusted stepwise, and the learning algorithm computes a gradient of the entire training set indicating by what amount the error would increase or decrease accordingly. More formally, the loss function (or the reward function)  $L$  on a set  $D$  of training samples is considered an optimization problem defined as:

$$L = \frac{1}{|D|} \sum_{(x,y) \in D} \mathcal{L}(g(x, w), y) \quad (3.8)$$

where  $g(x, w)$  is the network specified by Equation (3.1) and  $\mathcal{L}(g(x, w), y)$  is the loss defined on a data pair with input  $x$  and desired output  $y$ .

However, for training large datasets, the above gradient descent loss is inefficient and can overfit the training data. Therefore, stochastic gradient descent (SGD) (Amari, 1993) is widely adopted as a common optimization method. Rather than following the gradient of the entire training dataset, SGD calculates the gradient of a randomly selected subset of the training data called minibatch for each iteration and then backpropagates through all layers of the network. Many optimization methods have been proposed to improve SGD such as momentum SGD (Qian, 1999), AdaGrad (Duchi, Hazan, & Singer, 2011), and Adam (Kingma & Ba, 2014) which can be differently chosen depending on the specific task.

## 3.2 Image-guided Neurosurgery

Over the last decades, brain tumor surgery has altered from removing tumors based only on visual inspection to the resection of surrounding malignant cells beyond microscopic navigation with the help of imaging technologies breakthrough (Schipmann-Miletić & Stummer, 2020). IGN is the integration of imaging modalities with brain surgery which provides the ability to accurately visualize surgical targets during planning, operative, and follow-up procedures (Miner, 2017; Schipmann-Miletić & Stummer, 2020).

Thus, diagnostic imaging is critical to IGN to locate the tumor borders precisely as well as the adjacent brain healthy tissue (Miner, 2017). Imaging provides information about possible pathologies, intra-path anatomy, planning, progress monitoring, and assessing the interventional results (Miner, 2017). Further, intra-operative imaging guides the neurosurgeon dynamically in identifying residual tumor mass and real-time placement of surgical instruments allowing for the maximum extent of resection (EOR) as the basis of modern neurosurgery with a strong correlation to survival rate (Hervey-Jumper & Berger, 2016; Sanai, Polley, McDermott, Parsa, & Berger, 2011). Still, the pre- and intra-operative imaging data need advanced image processing and computer vision methods to allow IGN systems to reach their full potential (Schipmann-Miletić & Stummer, 2020).

### 3.2.1 IGN Components

A historical overview of image-guided interventional systems including their primary technological components and clinical applications are presented in (Cleary & Peters, 2010). For instance, **semantic segmentation** is an integral part of IGN that provides information about the location of surgical targets; tumors, and surrounding anatomical structures (Chapter 4). Another essential component of IGN is **registration** to accurately transform certain intra-operative patient data (physical space) into the corresponding imaging coordinates on the pre-operative MRI (image space) (Chapter 5). This can be done through paired point transformation using adhesive masks with multiple fiducials on their surface, anatomic structure identification, or hybrid systems. **Multimodal registration** can be also employed for combining information from different imaging modalities to provide the surgeon with map correspondences between pre-operative, intra-operative, and post-recurrence allowing better guidance (Chapter 5). Furthermore, **tracking devices** are an essential part of IGN systems that are used to track the rela-

tive position of instruments to the patient space (Chapter 7). Finally, **visualization** of data is a particularly important component of IGN in every phase of brain surgery (Chapter 7).

### 3.2.2 IGN Software

In general, IGN platforms can be divided into two main categories: open-source research systems and commercial systems. These software systems are the backbone of IGN which interconnects all IGN components such as segmentation, registration, tracking devices, and visualization. Table 3.1 provides a comparison of features provided by currently available platforms for image-guided interventions. Typical commercially available IGN systems, such as Brainlab AG<sup>1</sup> (Munich, Germany) systems, Medtronic StealthStation<sup>2</sup> (Minneapolis, USA), and ImFusion Suite<sup>3</sup> (Munich, Germany), are routinely used for various neurosurgical interventions. However, these platforms are closed in terms of their imaging data, processing algorithms, tracking, and other modules. One further limitation is that these systems usually do not allow adding other proprietary or open-source components (Shapey et al., 2021).

In contrast to commercial systems, open-source systems are open meaning that their source code is made freely available to the medical and research communities allowing them to integrate new contributions seamlessly. 3D Slicer (Fedorov et al., 2012), the Medical Imaging Interaction Toolkit (MITK) (Galloway et al., 2004), and the Image-Guided Surgery Toolkit (IGSTK) (Enquobahrie et al., 2007) are the three most widely used open-source platforms for the development of general medical image-guided systems (Cleary & Peters, 2010). To provide neuronavigation, these systems can be integrated with additional plugins such as the SlicerIGT extension (Ungi, Lasso, & Fichtinger, 2016), NifTK module (Clarkson et al., 2015), and the PLUS Toolkit (Lasso et al., 2014). IBIS (Drouin et al., 2017) and CustusX (Askeland et al., 2016) are other examples of open research platforms developed as intra-operative navigation systems.

---

<sup>1</sup> Brainlab AG, Munich, Germany; <http://www.brainlab.com/>

<sup>2</sup> Medtronic StealthStation, Minneapolis, USA; <https://www.medtronic.com/>

<sup>3</sup> imFusion Suite, Munich, Germany; <https://www.imfusion.com/>

Table 3.1: Comparison of the available platforms for image-guided interventions regarding visualization, segmentation, registration, acquisition, and iUS navigation.

Features	NifTK	CustusX	IBIS	SlicerIGT	ImFusion	Brainlab	Medtronic
2D/ 3D Visualization	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Customizable User Interface	No	No	No	Yes	No	Yes	Yes
Cranial Planning	No	No	No	No	No	Yes	Yes
Interactive Manual Segmentation	Yes	No	No	No	Yes	Yes	Yes
Patient Registration	No	Yes	Yes	Yes	No	Yes	Yes
MRI-MRI Registration	Yes	Yes	No	No	Yes	No	No
MRI-US Registration	No	Yes	Yes	No	Yes	No	No
iUS Navigation	No	Yes	Yes	Yes	No	Yes	Yes
Brain shift Visualization	No	No	Yes	Yes	Yes	Yes	Yes
Brain shift Correction	No	No	Yes	No	No	No	No
Instrument Tracking	Yes	Yes	Yes	Yes	No	Yes	Yes
Augmented Reality	No	No	Yes	No	No	No	No
Deep learning	Yes	No	No	No	Yes	No	No
FDA Approved	No	No	No	No	No	Yes	Yes
Open source	Yes	Yes	Yes	Yes	No	No	No

### 3.3 Challenges and Limitations

Despite that there is previous research on developing image-guided systems for assisting brain surgery, several challenges and limitations still exist. This section summarizes the most important challenges related to neurosurgical procedures (Section 3.3.1) and challenges related to technology and technical development (Section 3.3.2).

#### 3.3.1 Neurosurgical Challenges

##### 3.3.1.1 Precise Tumor Localization

In neurosurgery, inferring the pathological tissue while avoiding damage to other surrounding anatomical structures is one of the key challenges. This challenge is related to the difficulty of visually delineating these pathological targets from healthy tissue since most primary brain tumors grow by infiltration of healthy parenchyma (Coburger & Wirtz, 2019; Miner, 2017).



This is particularly important in surgery of LGG where these tumors often demonstrate only slight differences in texture and consistency compared to normal brain (Coburger et al., 2016).

Furthermore, 2D iUS systems provide many advantages for intra-operative navigation (refer to Section 2.1.2); however, the main limitation is their low image quality due to poor spatial resolution and small field of view (Unsgaard, Gronningsaeter, Ommedal, & Nagelhus Hernes, 2002). Insufficient imaging quality makes the identification of various brain structures challenging, especially deep structures such as the brainstem (Steno et al., 2021).

### 3.3.1.2 Brain shift

Neuronavigation using 3D pre-operative imaging data has become an indispensable tool for the planning and surgical guidance of brain tumor operations in many centers (Gerard et al., 2020). However, the accuracy of neuronavigation may become unreliable after the shifting of patient anatomy as a result of the surgery, patient positioning, application of mannitol, release of cerebral spinal fluid, and progressive resection of the tumor. This intra-operative brain deformation—called “brain shift”—reduces the utility of pre-operative imaging data for neurosurgical guidance. Brain shift has been observed to be as large as 24 mm at the brain cortex and exceeding 3 mm for the deep tumor margin in glioma operations (Hastreiter et al., 2004).

### 3.3.1.3 Inter-rater Variability

The inter-rater variability for the delineation of gliomas is the difference in tumor volumes, as assessed on imaging, between different interpreters (Kubben, Postma, Kessels, van Overbeeke, & van Santbrink, 2010; Visser et al., 2019). In enhancing tumors in glioblastoma, this is only a minor issue since observer agreements are excellent among experts (range 0.98–1.00). However, in non-enhancing tumor segmentations of glioblastoma, the spatial agreement was invariably highest with poor to fair agreement (range 0.41–0.79) for both pre-operative T2 and FLAIR MRI. The results reveal also that enhancing tumors in non-glioblastoma patients are very challenging to detect even for experienced raters (Visser et al., 2019).

## 3.3.2 Technical Challenges

### 3.3.2.1 Multiple Modules Integration

The software system is the core component of IGN, which contains segmentation, registration, tracking devices, and visualization modules, as discussed in Section 3.2. IGN system development is mostly time-consuming and still needs labor-intensive processes to integrate the above-mentioned elements. Providing that automatic brain tumor localization algorithms, information by tracking systems, and imaging data fusion typically come from different sources that belong to environments, correlating these huge amounts of data with the relevant images with real-time visualization is still a non-trivial task. Further, the software development must be carefully designed to ensure ease of use, robustness, and stability as it is used in life-critical applications (Enquobahrie et al., 2007).

#### 3.3.2.2 Deep Networks Transparency

Most machine learning and/or deep learning techniques are still under development for deployment in the clinical field (Angelov et al., 2021; Xie et al., 2021). This is due to the underlying “black-box” nature of the deep learning methods which are often characterized by the lack of human-like explainable decisions. In addition, these models have a substantial number (within millions) of extracted feature maps in each internal layer which are assumed to contain meaningful information about the input problem and its possible solution. This makes fully understanding deep learning methods highly problematic even for professional experts. Thus, the application of such “black box” models in highly sensitive medical applications is very limited (Angelov et al., 2021; G. Yang et al., 2022).

Recently, there is a growing interest in XAI to address the fundamental question about the justification of the decision-making process made by deep learning models (Du, Liu, & Hu, 2019). Though the explainability does not improve the accuracy of the deep learning model, XAI is important to guarantee safety in use and increase the trust of non-deep learning users, i.e., surgeons and radiologists. XAI provides machine learning methods the ability to describe their “black-box” nature in explainable or interpretable terms to humans (Angelov et al., 2021; G. Yang et al., 2022).

#### 3.3.2.3 Neuronavigational Challenges

Image-guided navigation using MRI imaging acquired pre-operatively provides the surgeon with valuable information to define brain tumor boundaries and their adjacent structures. However, neuronavigation loses its significance as the surgery progresses and brain shift and deformation occur (Bastos et al., 2021; Censi, Mattei, Triventi, Bartolini, & Calcagnini, 2012). Intra-operative navigation using iUS partially handles this problem by providing low-cost real-time imaging with less disruption to the surgical workflow. But interpreting iUS images is still a challenging task and requires neurosurgeons with extensive experience. Further, the reconstruction of 3D iUS volumes may contain some artifacts due to variations in the steadiness of the transducer and the operational speed of the probe translation (Bastos et al., 2021).

# 4 Brain Tumor Segmentation

---

## Contents

<b>4.1</b>	<b>Introduction</b>	<b>34</b>
<b>4.2</b>	<b>Related Work</b>	<b>34</b>
<b>4.3</b>	<b>Methods</b>	<b>36</b>
<b>4.3.1</b>	<b>DeepSeg</b>	<b>36</b>
<b>4.3.1.1</b>	<b>Feature extractor</b>	<b>36</b>
<b>4.3.1.2</b>	<b>Image upscaling</b>	<b>37</b>
<b>4.3.2</b>	<b>3D DeepSeg</b>	<b>37</b>
<b>4.3.3</b>	<b>Ensemble Methods</b>	<b>38</b>
<b>4.3.4</b>	<b>Post-processing</b>	<b>39</b>
<b>4.4</b>	<b>Experiments</b>	<b>39</b>
<b>4.4.1</b>	<b>Data</b>	<b>39</b>
<b>4.4.2</b>	<b>Experimental setup</b>	<b>42</b>
<b>4.4.3</b>	<b>Evaluation Metrics</b>	<b>43</b>
<b>4.4.4</b>	<b>Ablation Study</b>	<b>44</b>
<b>4.5</b>	<b>Results and Discussion</b>	<b>45</b>
<b>4.5.1</b>	<b>Statistical Evaluation</b>	<b>45</b>
<b>4.5.2</b>	<b>Qualitative Output</b>	<b>47</b>
<b>4.5.3</b>	<b>Ensemble Results</b>	<b>49</b>
<b>4.6</b>	<b>Summary</b>	<b>52</b>

Within this chapter, the first module of the proposed IGN pipeline (see Figure 1.2 “DeepIGN”) called *Automatically define the boundary of brain tumors accurately* is described. This chapter is organized as follows: Introduction to glioma detection in Brain MRI and related automatic segmentation methods is given in Section 4.1 and Section 4.2, respectively. Section 4.3 describes the architecture of the proposed framework in addition to the developed ensemble methods. The utilized brain tumor segmentation dataset, experimental setup, and evaluation metrics are presented in Section 4.4. Section 4.5 gives comprehensive experiments and comparisons with the state-of-the-art, and finally, this chapter work is concluded in Section 4.6. Parts of this chapter have previously appeared in (Ramy A. Zeineldin et al., 2023; R. A. Zeineldin et al., 2020; Ramy A. Zeineldin, Mohamed E. Karar, et al., 2022a).

## 4.1 Introduction

Manual segmentation of the brain tumor is a vital procedure and needs a group of clinical experts to accurately define the location and the type of the tumor. Moreover, the process of lesion localization is very labor-based and highly dependent on the physicians' experience, skills, and slice-by-slice decisions. Alternatively, automated computer-based segmentation methods present a suitable solution to save the surgeon's time and to provide reliable and accurate results, while reducing the exerted efforts of experienced physicians to accomplish the procedures of diagnosis or evaluation for every single patient (Karar, Merk, Falk, & Burgert, 2016).

Formerly, numerous machine learning algorithms were developed for the segmentation of normal and abnormal brain tissues using MRI images (W. Wu, Chen, Zhao, & Corso, 2014). However, choosing features that enable this operation to be fully automated is very challenging and requires a combination of computer engineering and medical expertise. Therefore, classical approaches depend heavily on the applied application and do not generalize well. Nevertheless, developing fully automated brain tumor methods is still a challenging task, because malignant areas varied in terms of shape, size, and localization, and they can only be defined through the intensity changes relative to surrounding healthy cells.

Semantic segmentation is currently one of the most important tasks in the field of computer vision toward complete scene understanding. Deep learning has become recently an attractive field of machine learning that outperforms traditional computer vision algorithms in a wide range of applications such as object detection (Ouyang et al., 2017), semantic segmentation (Shelhamer, Long, & Darrell, 2017) as well as other applications such as navigation guidance (Saleh, Zeineldin, Hossny, Nahavandi, & El-Fishawy, 2018).

## 4.2 Related Work

Early approaches to applying semantic segmentation in the medical field use patch-wise image classification (Cireşan, Giusti, Gambardella, & Schmidhuber, 2012). However, it suffers from two main problems: First, the training patches are much larger than the training samples, which require a higher number of computation cycles resulting in a large running time consumption. Second, the segmentation accuracy depends heavily on the appropriate size of patches. Accordingly, new U-shaped encoder-decoder architecture was introduced, refer to Section 3.1.3, which can solve classical semantic segmentation problems by using two main paths: a contracting path (or encoder) and an expansive path (or decoder) (Ronneberger et al., 2015).

Inspired by the recent Densenet architecture (Huang et al., 2017) and U-Net (Ronneberger et al., 2015), DeepSCAN architecture was proposed for semantic segmentation. Figure 4.1 shows the two DeepSCAN architectures introduced for brain tumor segmentation. Instead of using transition layers and pooling operations, dilated convolutions are used to increase the receptive field of the encoder. Similar to Densenet, the output of each layer is concatenated with its input before passing to the next layer. Moreover, label-uncertainty is applied directly to the loss-function, which allows the prediction of the CNN to be involved in evaluating the network

decision. This hybrid 2D/ 3D approach led to more stable results and is one of the winning solutions in the BraTS 2018 challenge.

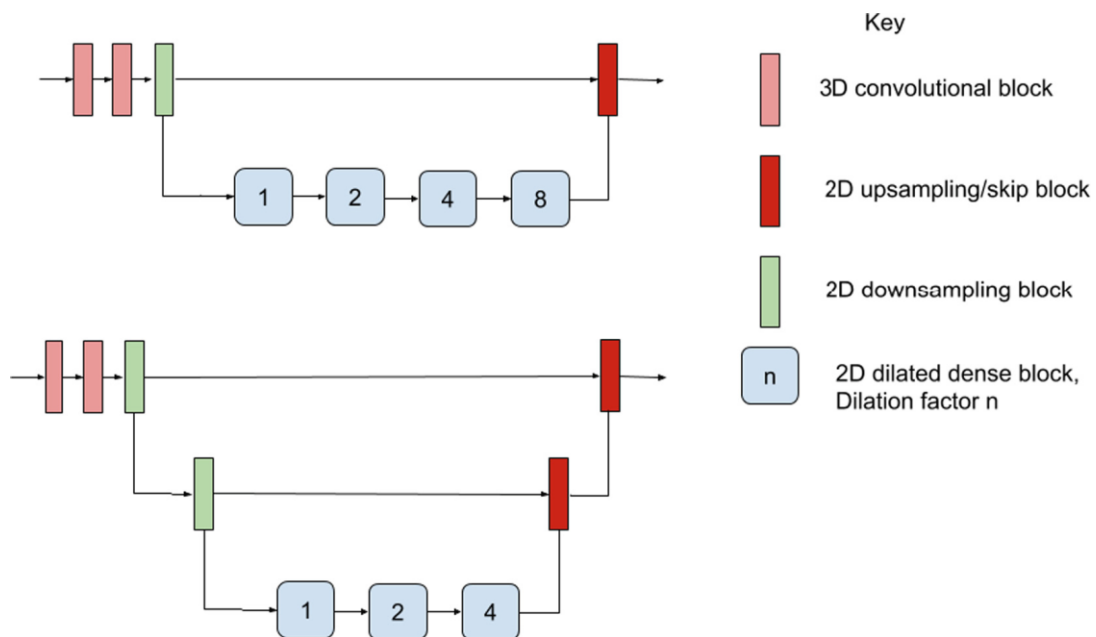


Figure 4.1: DeepSCAN architectures, as applied to brain tumor segmentation in BraTS 2018 challenge (McKinley, Meier, & Wiest, 2019).

The baseline nnU-Net is outlined in Fig. 4.2, which is a self-adaptive deep learning-based framework for 3D semantic biomedical segmentation (F. Isensee et al., 2021). Unlike the proposed DeepSeg, nnU-net does not use any of the recently proposed architectural advances in deep learning and only depends on plain convolutions for feature extraction. nnU-Net used strided convolutions for downsampling and convolution transposed for upsampling. The initial filter size of convolutional kernels is set to 32 and doubled at the following layers with a maximum of 320 in the bottleneck layers.

By modifying the baseline nnU-Net and making BraTS-specific processing, nnU-Net won first place in the segmentation task of the BraTS challenge in 2020 (Fabian Isensee, Jäger, Full, Vollmuth, & Maier-Hein, 2021). The softmax output was replaced with a sigmoid layer to target the three evaluated tumor sub-regions: whole tumor (consisting of all 3 labels), tumor core (label 1 and label 4), and enhancing tumor (label 4). Further, the training loss was changed to a binary CE instead of a categorical CE that optimized each of the sub-regions independently. Also, the batch size was increased to 5 as opposed to 2 in the baseline nnU-Net and more aggressive data augmentations were incorporated. Similar to DeepSeg, nnU-Net utilized BN instead of instance normalization. After all, the sample dice loss function was changed to batch dice by computing over all samples in the batch.

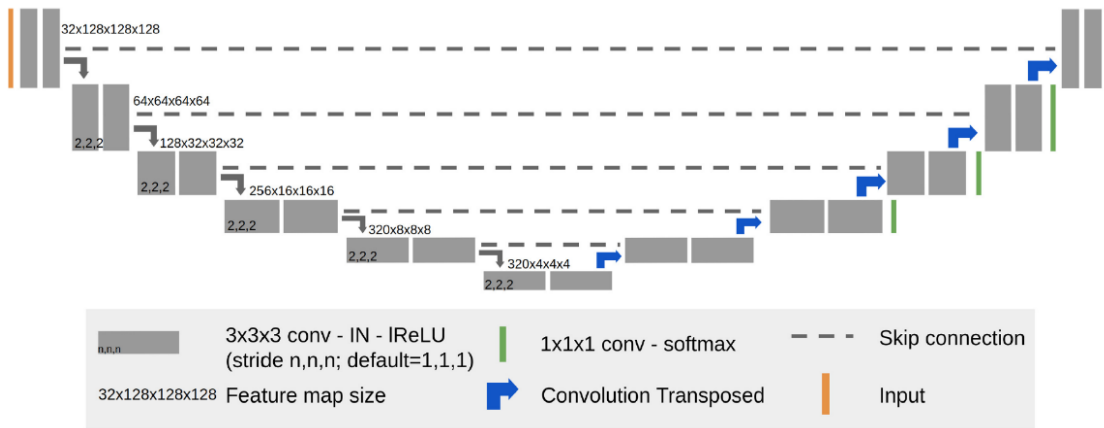


Figure 4.2: nnU-Net network consists of strided convolution blocks (*grey boxes*), and upsampling as convolution transposed (*blue arrows*). The input patch size was set to  $128 \times 128 \times 128$  and the maximum filter size is 320 (Fabian Isensee et al., 2021).

This chapter aims at developing a new fully automated MRI brain tumor segmentation based on modified U-Net models. First, DeepSeg is proposed in Section 4.3.1 as a generic modular architecture of the brain tumor segmentation with two elements: feature extraction and image expanding paths, in order to support applying different deep neural network models successfully. Moreover, two fully automated CNN methods were proposed for glioma segmentation based on an ensemble of three encoder-decoder methods, namely, DeepSeg (R. A. Zeineldin et al., 2020), nnU-Net (Fabian Isensee et al., 2021), and DeepSCAN (McKinley et al., 2019). All networks follow the U-Net pattern (Çiçek, Abdulkadir, Lienkamp, Brox, & Ronneberger, 2016; Ronneberger et al., 2015) and consist of encoder-decoder architecture interconnected by skip connections.

## 4.3 Methods

### 4.3.1 DeepSeg

DeepSeg is a generic decoupled framework for automatic tumor segmentation, as shown in Fig. 4.3. Thanks to the basic U-Net structure (Ronneberger et al., 2015), it consists of two main parts: a feature extractor (or encoder) part and an image upscaling (or decoder) part. This universal design has two main advantages: First, it allows the extensibility of the system, i.e. different encoders and decoders can be added easily. Moreover, a discriminative comparison between the various proposed models can be done straightforwardly. In the following, the proposed architecture is described in detail.

#### 4.3.1.1 Feature extractor

The modified U-Net encoder has been implemented by using advances in CNNs including dropout and batch normalization (BN) (Ioffe & Szegedy, 2015; Srivastava, Hinton, Krizhevsky, & Salakhutdinov, 2014). In addition, state-of-the-art DNN architectures are integrated into the

benchmarking system to extract the feature map. These models are utilized to achieve better performance than the obtained results in the ILSVRC competition until now (Russakovsky et al., 2015). Every proposed model has its own set of parameters and computational resource requirements, as described below.

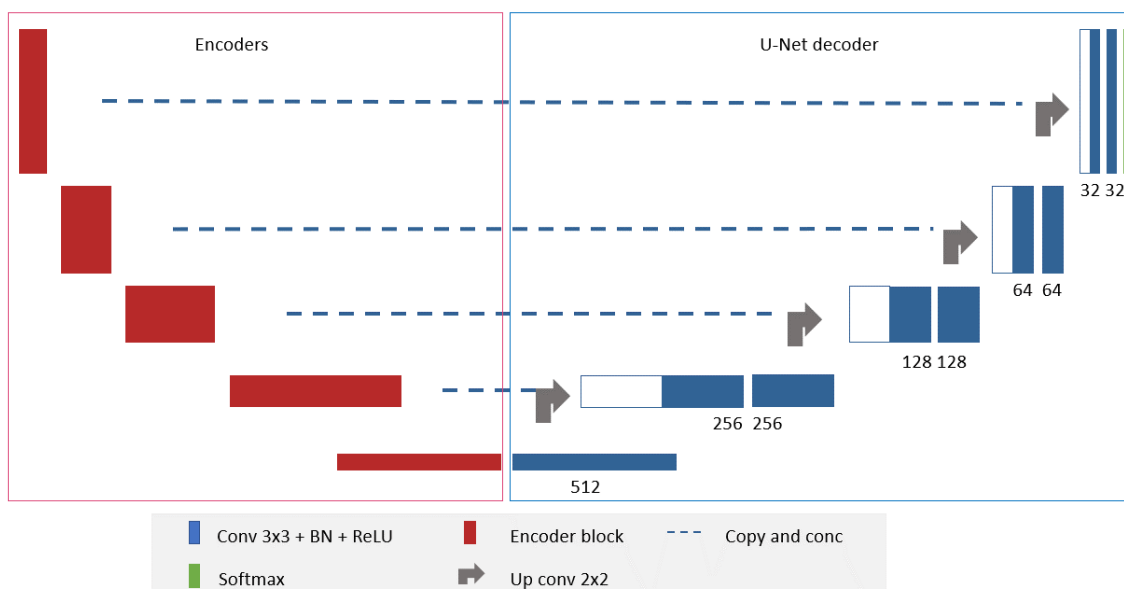


Figure 4.3: DeepSeg architecture for using different feature extractor models of MRI brain tumors.

#### 4.3.1.2 Image upscaling

In semantic segmentation, it is very crucial to use both semantic and spatial information so that the neural network can perform well. Hence, the decoder should recover the missing spatial information to get the full-resolution segmented map from the consequential encoded features. By skip connections (Fig. 4.3), U-Net can obtain the semantic feature map from the bottleneck and recombine it with higher resolution outputs from the encoder respectively.

Unlike the standard U-Net decoder, some modifications were incorporated for further exceptional segmentation results. Firstly, a BN layer is applied between each convolution and ReLU to make each layer learn independently from other layers and thus contribute to faster learning. Additionally, a smaller filter size of 32 as the base filter is selected and doubled at the following layers, in order to apply the full size as input rather than using patches or small regions of the input. Finally, the output of the network is passed into a softmax output layer which converts the output logics into a list of probability distributions.

#### 4.3.2 3D DeepSeg

Figure 4.4 shows a 3D enhanced version of the first model, DeepSeg, which is a modular framework for fully automatic brain tumor detection and segmentation. The proposed network differs from the original network in the following: First, the original DeepSeg network was proposed for 2D tumor segmentation using only FLAIR MRI data, however, 3D convolutions

are applied here over all slices for more robust and accurate results. Second, all the available MRI modalities (T1, T1Gd, T2, and FLAIR) are incorporated so that the GBM sub-regions could be detected in comparison with the whole tumor only in the original DeepSeg version (Section 4.3.1). Third, additional modifications are made such as region-based training, excessive data augmentation, a simple postprocessing technique, and a combination of cross-entropy (CE) and Dice similarity coefficient (DSC) loss functions.

Following the structure of U-Net, DeepSeg consists of two main parts: a feature extractor part and an image upscaling part. Downsampling is performed with  $2 \times 2 \times 2$  max-pooling and upsampling is performed with  $2 \times 2 \times 2$  up convolution. DeepSeg uses the recently proposed advances in CNNs including dropout, BN, and ReLU (Ioffe & Szegedy, 2015; Srivastava et al., 2014). The feature extractor consists of five consecutive convolutional blocks, each containing two  $3 \times 3 \times 3$  convolutional layers, followed by ReLU. In the image upscaling part, the resultant feature map of the feature extractor is upsampled using deconvolutional layers. The final output segmentation is attained using a  $1 \times 1 \times 1$  convolutional layer with a softmax output.

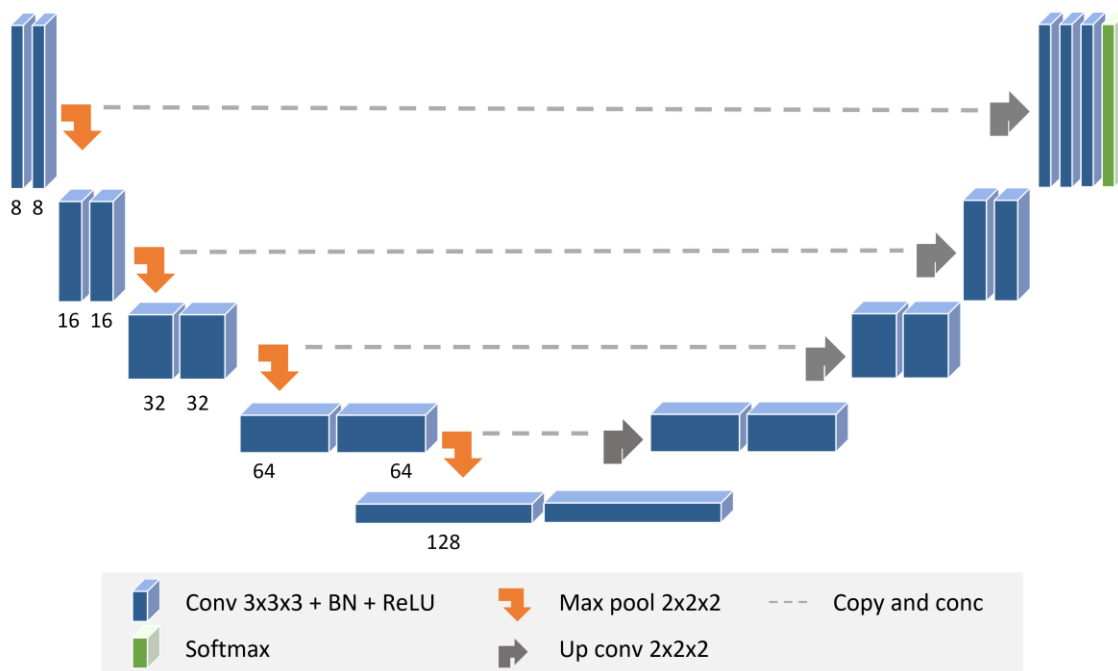


Figure 4.4: 3D DeepSeg network consists of convolution neural blocks (blue boxes), downsampling using maximum pooling (orange arrows), upsampling using up convolution (blue arrows), and softmax output layer (green block). The input patch size was set to  $128 \times 128 \times 128$ .

### 4.3.3 Ensemble Methods

The ensemble is a widely adopted effective paradigm of deep learning, which aims to improve the global performance of deep models (Dietterich, 2000). For the image segmentation problem, a standard ensemble method has been developed by taking the combination of two or more segmentations networks. However, this simple strategy does not take into account potential variability in the quality or performance between models, addressing each network output in the



same way. Furthermore, such voting strategies do not enable prior information on the segmented structure to be incorporated (P. Yang, Hwa Yang, B Zhou, & Y Zomaya, 2010). Another advanced ensemble method called the Simultaneous Truth and Performance Level Estimation (STAPLE) (Warfield, Zou, & Wells, 2004), was proposed as an instance of the expectation-maximization algorithm (Dempster, Laird, & Rubin, 1977). Using a collection of image segmentations, STAPLE calculates a probability estimate of the actual segmentation and simultaneously a measure of the performance level represented by each segmentation.

### 4.3.4 Post-processing

Determining the small blood vessels in the tumor core (necrosis or edema) is one of the most challenging segmentation tasks in the BraTS Challenge. In particular, this is clear in low-grade glioma (LGG) patients where they may not have enhancing tumors and, therefore, the BraTS challenge evaluates the segmentation as binary values of 0 or 1. Although if there are only small false positives in the predicted segmentation map of a patient with no enhancing tumor will result in a dice value of 0. To overcome this problem, all enhancing tumor output were re-labeled with necrotic (label 1) if the total predicted ET regions are less than a threshold. This threshold value was selected based on our analysis of the validation set results so that our model performs better. This strategy has a possible side effect of removing some correct predictions.

## 4.4 Experiments

### 4.4.1 Data

MRI data from the Medical Image Computing and Computer-Assisted Intervention (MICCAI) BraTS challenges 2019 and 2021 (Baid et al., 2021; S. Bakas et al., 2017; Spyridon Bakas et al., 2018; Menze et al., 2015) have been used in this chapter for accomplishing the brain tumor segmentation task.

DeepSeg was performed using the FLAIR MRI data from the BraTS 2019 challenge (Menze et al., 2015). Although T1Gd is the standard imaging for Glioma, FLAIR is becoming increasingly relevant in the case of malignant tumors, since there is a trend to also resect the FLAIR positive areas (Y. M. Li, Suki, Hess, & Sawaya, 2016). Moreover, the advantages of FLAIR images in the brain surgery of LGGs have been investigated (Coburger et al., 2016). BraTS 2019 dataset (Spyridon Bakas et al., 2018) contains multi-institutional pre-operative MRI of 336 heterogeneous (in shape, appearance, size, and texture) Glioma patients (259 HGGs and 76 LGGs). While the BraTS 2021 dataset (Baid et al., 2021; S. Bakas et al., 2017; Menze et al., 2015) provides the largest multi-institutional annotated mpMRI dataset, with more than 2000 cases, aiming at evaluating state-of-the-art automated deterministic solutions for the segmentation of intra-axial brain tumors.

Each patient of BraTS datasets has four multimodal scans: native T1, post-contrasted T1Gd, T2, and FLAIR. MRI data were acquired from multiple institutions using various clinical protocols and different scanners. The manual segmentation of the data was done by experienced neuro-

radiologists, from 1 to 4, following the same annotation procedure. After that, the MRI scans are resampled and interpolated to the same resolution of  $1 \text{ mm}^3$ . Fig. 4.5 displays the provided segmented labels: background (label 0), necrotic and non-enhancing tumor core (NCR/NET) (label 1), ED (label 2), and ET (label 4). These labels are combined to generate the final evaluation of three regions: the tumor core (TC) of labels 1 and 4, enhancing tumor (ET) of label 4, and the whole tumor (WT) of all labels.

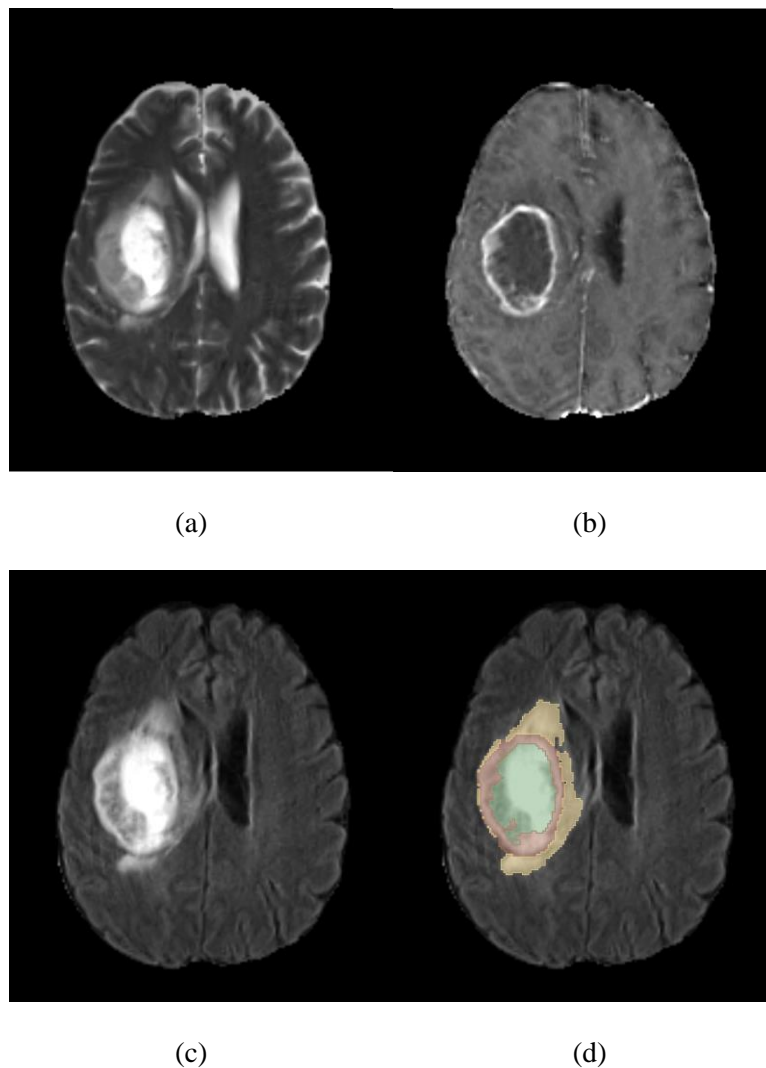


Figure 4.5: A sample of the mpMRI BraTS 2021 training set. Shown are image slices in two different MRI modalities T1Gd (a), T2 (b), FLAIR (c), and the ground truth segmentation (d). The color labels indicate Edema (*yellow*), enhancing solid tumor (*red*), non-enhancing tumor core, and necrotic core (*green*). Images were obtained by using the 3D Slicer software (Fedorov et al., 2012).

The BraTS 2021 data were acquired using different clinical protocols, from different MRI scanners and multiple institutions, therefore, a pre-processing stage is essential. First, standard pre-processing routines have been applied by the BraTS challenge as stated by (Baid et al., 2021). This includes conversion from DICOM into NIFTI file format, re-orientation to the same coordinate system, co-registration of the multiple MRI modalities, resampling to  $1 \times 1 \times 1 \text{ mm}$

isotropic resolution, and brain extraction and skull-stripping. Following these pre-processing steps, the image cropping stage is applied where all brain pixels were cropped, and the resultant image was resized to a spatial resolution of  $192 \times 224 \times 160$ . This method effectively results in a closer field of view (FOV) to the brain with fewer image voxels leading to a smaller resource consumption while training the deep learning models. Finally, z-score normalization was applied by subtracting the mean value and dividing it by the standard deviation individually for each input MRI image.

For the training of large neural networks using limited training data, some precautions should be taken to prevent the problem of overfitting. One of them is data augmentation, which is the process of creating new artificial training data from the original one to improve the model performance by making the model generalize well to the new testing data. In this study, a set of simple on-the-fly data augmentation methods is applied (as listed in Table 4.1) by horizontal and vertical flipping, rotation, scaling, shearing, and shift. Unfortunately, these simple methods are not enough to get sufficient immune training data, therefore more complex methods are also introduced such as elastic distortion corresponding to uncontrolled noise of MRI sensors, where  $\sigma$  is the elasticity coefficient and  $\beta$  is the multiplying factor of the displacement fields which controls the intensity of deformation. Figure 4.6 shows some examples of the applied augmentation techniques.

Table 4.1: List of the applied data augmentation methods.

Methods	Parameters
Flip horizontally	20% of all images
Flip vertically	20% of all images
Scale	$\pm 20\%$ on both horizontal and vertical direction
Translation	$\pm 20\%$ on both horizontal and vertical direction
Rotation	$\pm 25^\circ$
Shear	$\pm 8^\circ$
Elastic transformation	$\beta = 720, \sigma = 24$

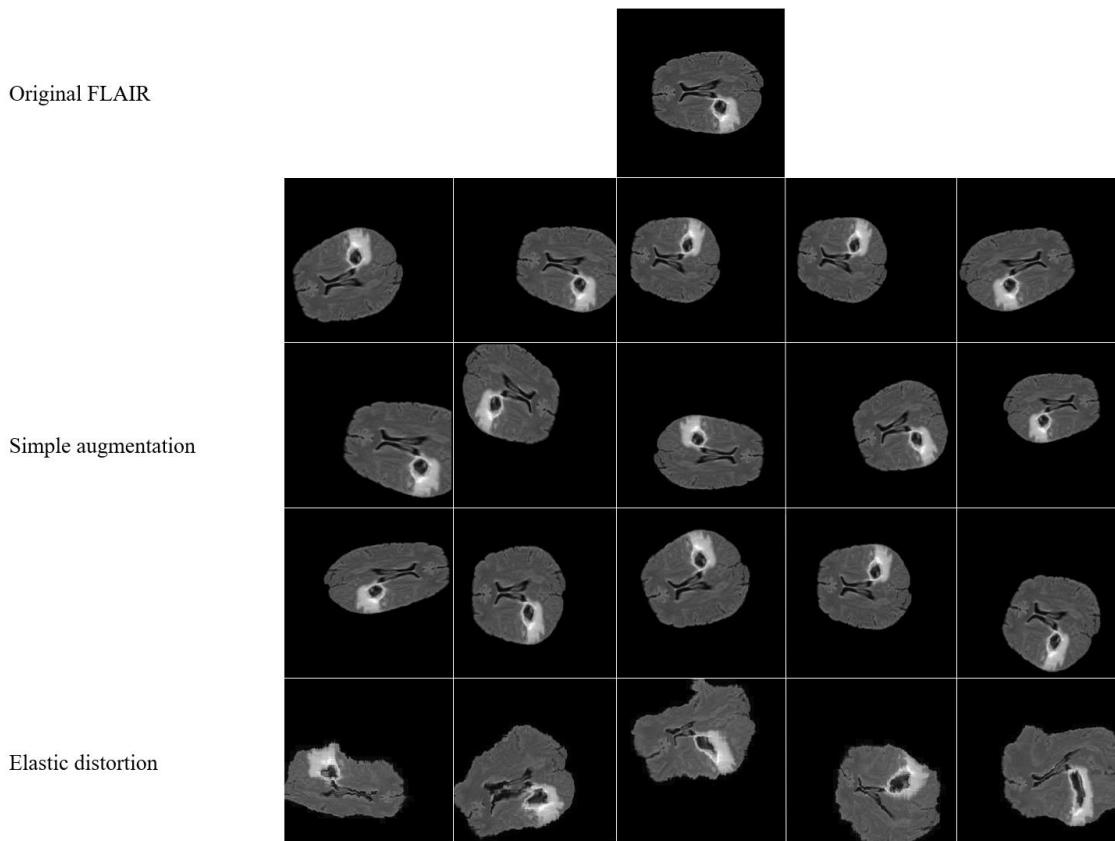


Figure 4.6: Random augmented image transformation. The first row shows the original image. The next three rows present horizontal and vertical flipping, scaling, translation, rotation, and shearing methods. The elastic transformation is presented in the last row.

#### 4.4.2 Experimental setup

The proposed methods of this chapter were run on AMD Ryzen 2920X (32M Cache, 3.50 GHz) CPU with a single Nvidia GPU (RTX 2080 Ti or RTX 3060). Proposed DeepSeg models were implemented using Tensorflow (Abadi et al., 2016) while nnU-Net was implemented using PyTorch (Paszke et al., 2019).

For DeepSeg, experiments are done using FLAIR MRI sequences with a resolution of  $224 \times 224$  in order to use all the proposed feature extractor networks. DeepSeg networks are trained for 35 epochs and a batch size of 16. During the training process, spatial dropout with a rate of 0.5 was used after the feature extractor path. This is a simple type of regularization to ensure that the neural networks generalize well without overfitting the training dataset. Adam optimizer (Kingma & Ba, 2014) has been applied with a learning rate of 0.00001.

For 3D DeepSeg and ensemble methods, the input is set to randomly sampled patches of  $128 \times 128 \times 128$  voxels with varying batch sizes from 2 to 5 and the post-processing threshold is set to 200 voxels. This tiling strategy allows the model to be trained on multi-modal high-resolution MRI images with low GPU memory requirements. Each model was trained as five-fold cross-validation on the 1251 training cases of BraTS 2021 for a maximum of 1200 epochs. Adam

optimizer (Kingma & Ba, 2014) has been applied with an initial learning rate of 1e-4 and a default value of 1e-7 for epsilon

Nevertheless, the BraTS dataset suffers from a data imbalance problem where the tumor pixels are less than 2% and the healthy pixels are mostly 98% of the whole dataset. To solve this problem in two steps, firstly, the models were trained on the brain sequences and ignored the empty slices, secondly, weighted CE loss for each label was used to pay more attention to the malignant labels than the background as defined by:

$$L_{2D} = - \sum_{c=1}^{N_c} y_c \log(p_c) * w_c \quad (4.1)$$

where  $N_c$  is the number of classes including the background and the tumor cells in this study,  $y_c$  represents the true labels for the  $c^{th}$  class,  $p_c$  is the predicted softmax output for those true labels, and  $w_c$  is the proposed weight class map of (0.05, 0.95) to focus on the tumor pixels rather than the background.

For training 3D DeepSeg and ensemble methods, the loss function is a combination of CE and DSC loss functions, which can be calculated as follows:

$$L_{3D} = DSC + CE = \frac{2 * \sum_{c=1}^{N_c} y_c p_c + \varepsilon}{\sum_{c=1}^{N_c} y_c + \sum_{c=1}^{N_c} p_c + \varepsilon} - \sum_{c=1}^{N_c} y_c \cdot \log(p_c) \quad (4.2)$$

where  $\varepsilon$  is the smooth parameter to make the dice function differentiable. To further overcome the effect of class imbalance between tumor labels and the brain healthy tissue, on-the-fly spatial data augmentations were applied during training (random rotation between 0 and 30°, random 3D flipping, power-law gamma intensity transformation, or a combination of them).

### 4.4.3 Evaluation Metrics

For the evaluation of segmentation results, four metrics namely; DSC, sensitivity, specificity, and the Hausdorff distance (HD) are computed. DSC score calculates the overlap of the segmented region and the ground truth  $y$  and is applied to the network softmax predictions  $p$  as follows:

$$DSC = \frac{2 * \sum y p + \varepsilon}{\sum y + \sum p + \varepsilon} \quad (4.3)$$

Note that  $\varepsilon$  is the smooth parameter to make the dice function differentiable. This dice overlap can take values from 0 (represents lower overlap) to 1 (indicates a full overlap). Specificity and sensitivity are given by:

$$Specificity = \frac{TN}{TN+FP} \quad (4.4)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (4.5)$$

where true positives (TP) and false positives (FP) refer to the number of retrieved points that are correct/incorrect, and similarly for true and false negatives, TN and FN, respectively.

Dice, Sensitivity, and Sensitivity metrics are measures of pixel-based overlap between the ground truth and the predicted segmentations. In addition, the HD gives the largest distance of the segmentation set to the nearest point in the truth set, as defined by

$$HD(p, y) = \max\{h(p, y), h(y, p)\} \quad (4.6)$$

with

$$h(p, y) = \max_{i \in p} \left\{ \min_{j \in y} \{D(i, j)\} \right\} \quad (4.7)$$

where the shortest Euclidian distance  $D(i, j)$  is calculated for every point  $i$  of the segmentation set  $p$ , with respect to the ground truth point  $j$  in the image. 95% percentile of the Hausdorff distance (HD95) was utilized to reduce the number of outliers in the local distance error estimation.

#### 4.4.4 Ablation Study

Thanks to the DeepSeg framework, several methods were analyzed and compared simultaneously. Table 4.2 illustrates the different characteristics of these automated methods with the corresponding computational times. Training and prediction times present the average estimated time of applying each algorithm about 35 times during the training and validation respectively. These tests showed that the MobileNet encoder requires the smallest resources with only 22 MB of memory and roughly 5.6 thousand parameters. It is worth mentioning that MobileNet and MobileNetV2 are mainly developed for mobile and embedded applications where the hardware resources are limited. Likewise, U-Net, modified U-Net, and NASNet consumes a small amount of memory of 30 MB, 30 MB, and 37 MB respectively. Apparently from the results, there is a proportional relationship between the number of parameters and the demanded memory. On contrary, the ResNet model consumes the largest amount of memory of 118 MB, which is not considered a problem since modern GPUs possess a memory of several Gigabytes. Other models such as DenseNet, VGGNet, and Xception are located in the middle level of memory consumption of 51 MB, 71 MB, and 103 MB, respectively.

Moreover, the number of layers has a significant influence on both the training and prediction time. For instance, the training time of one epoch using U-Net with the smallest number of layers (39 layers), is 381 seconds and the prediction time is just 1.1 seconds. But the NASNet model with 818 layers requires 684 seconds for one epoch to train and the prediction of one patient took 4.4 seconds. Nevertheless, this is not the general rule since modified U-Net, MobileNet, and MobileNetV2 share the second place with a training time of 385 seconds even though they have various numbers of layers of 74, 129, and 202, respectively. The main reason is the internal building architecture of MobileNet variants which is developed for smartphone devices.

Table 4.2: A comparative performance of the employed models. Average computational times for each encoder of 35 results during training and validation phases.

Encoder	Size (MB)	Training Time (sec)	Prediction Time (sec)	Parameters	Layers
U-NET	30	<b>381</b>	<b>1.1</b>	7760642	<b>39</b>
Modified U-NET	30	385	1.3	7763050	74
VGGNet	71	540	1.6	18540938	56
ResNet	118	446	2.3	30546458	223
DenseNet	51	482	3.2	12947674	474
Xception	103	580	1.9	26769602	184
MobileNet	30	385	1.5	7590746	129
NASNet	37	684	4.4	8652846	818
MobileNetV2	<b>22</b>	386	1.8	<b>5591770</b>	202

## 4.5 Results and Discussion

### 4.5.1 Statistical Evaluation

The DeepSeg framework consists of several automated feature extractors in addition to an image-expanding path. The corresponding evaluation results have been obtained by running two-fold cross-validation on the 336 training cases of the BraTS 2019 dataset divided as follows: 270 cases for training and 66 for validation. Table 4.3 summarizes the comparison and the overall measurement results of all tested methods on the BraTS 2019 training database.

The proposed DeepSeg architectures were able to accurately detect tumor regions in the validation set with mean DSC scores ranging from 0.809 to 0.839, while the mean dice score of the expert's annotation for the whole tumor mass is about 0.85 as reported in (Diba, Sharma, Pazandeh, Pirsiavash, & Gool, 2017). Although statistical analysis of results is relatively close or identical (like Specificity), these results give an important indication that fully automated deep learning models maybe utilized in the task of brain tumor segmentation. As illustrated in Table 4.3, The DenseNet, Xception, VGGNet, and MobileNet encoders achieved the best DSC scores of 0.839, 0.839, 0.837, and 0.835, respectively. Although the Xception encoder showed the best value for the sensitivity of 0.856 approximately 7% better than the original U-Net model, it achieved the same value of specificity. This result confirms that point-based approaches are not enough for evaluating the brain tumor segmentation method. Therefore, the HD95 measurements were applied to verify both the best accuracy and performance among all tested deep encoders. The MobileNet showed the shortest HD95 value of 10.924.

Table 4.3: Mean DSC, Sensitivity, Specificity, and Hausdorff distance scores of testing different encoders on BraTS 2019 training data.

Encoder	DSC $\uparrow$	Sensitivity $\uparrow$	Specificity $\uparrow$	HD95 $\downarrow$
U-NET	0.809	0.799	<b>0.998</b>	12.926
Modified U-NET	0.814	0.783	<b>0.999</b>	13.341
VGGNet	<b>0.837</b>	0.819	<b>0.998</b>	12.633
ResNet	0.811	0.789	<b>0.998</b>	13.652
DenseNet	<b>0.839</b>	0.827	<b>0.998</b>	13.156
XCeption	<b>0.839</b>	<b>0.856</b>	<b>0.998</b>	11.337
MobileNet	<b>0.835</b>	0.843	<b>0.998</b>	<b>10.924</b>
NASNet	0.834	0.826	<b>0.998</b>	12.608
MobileNetV2	0.827	0.822	<b>0.998</b>	12.029

For consistency with other publications, the proposed DeepSeg architectures have been also tested on the validation datasets of BraTS 2019 (125 cases). Table 4.4 presents the compared scores of mean dice similarity coefficient, sensitivity, specificity, and HD95, similar to the online evaluation platform (<https://ipp.cbica.upenn.edu/>). These results showed that the proposed models are robust and able to deal with MRI segmentation tasks. In Table 4.4, the DenseNet architecture outperformed other models for the DSC (0.841) as well as in the training

Table 4.4: Mean DSC, Sensitivity, and Specificity scores of applied models on BraTS 2019 validation data.

Encoder	DSC $\uparrow$	Sensitivity $\uparrow$	Specificity $\uparrow$	HD95 $\downarrow$
U-NET	0.813	0.841	0.987	19.747
Modified U-NET	0.820	0.853	0.987	12.014
VGGNet	0.829	0.837	0.990	9.756
ResNet	0.823	0.832	0.990	10.005
DenseNet	0.841	0.860	0.989	10.595
XCeption	0.834	0.865	0.988	12.571
MobileNet	0.830	0.855	0.989	11.696
NASNet	0.830	0.861	0.988	11.673
MobileNetV2	0.822	0.854	0.988	13.894



set, however, it ranked second with an HD95 (10.595) which is clinically accepted. The dice metrics and HD95 are the most important measurements when evaluating and comparing deep learning models because they show the percentage of the overlapping between ground truth segmentation and predictions. In contrast, the lack of false positives indicated high values of both specificity and sensitivity, which may not precisely reflect the actual performance.

## 4.5.2 Qualitative Output

Figures 4.7 and 4.8 show segmentation results for the proposed DeepSeg architectures generated from the validation set (67 cases). In both figures, the first row indicates the FLAIR images

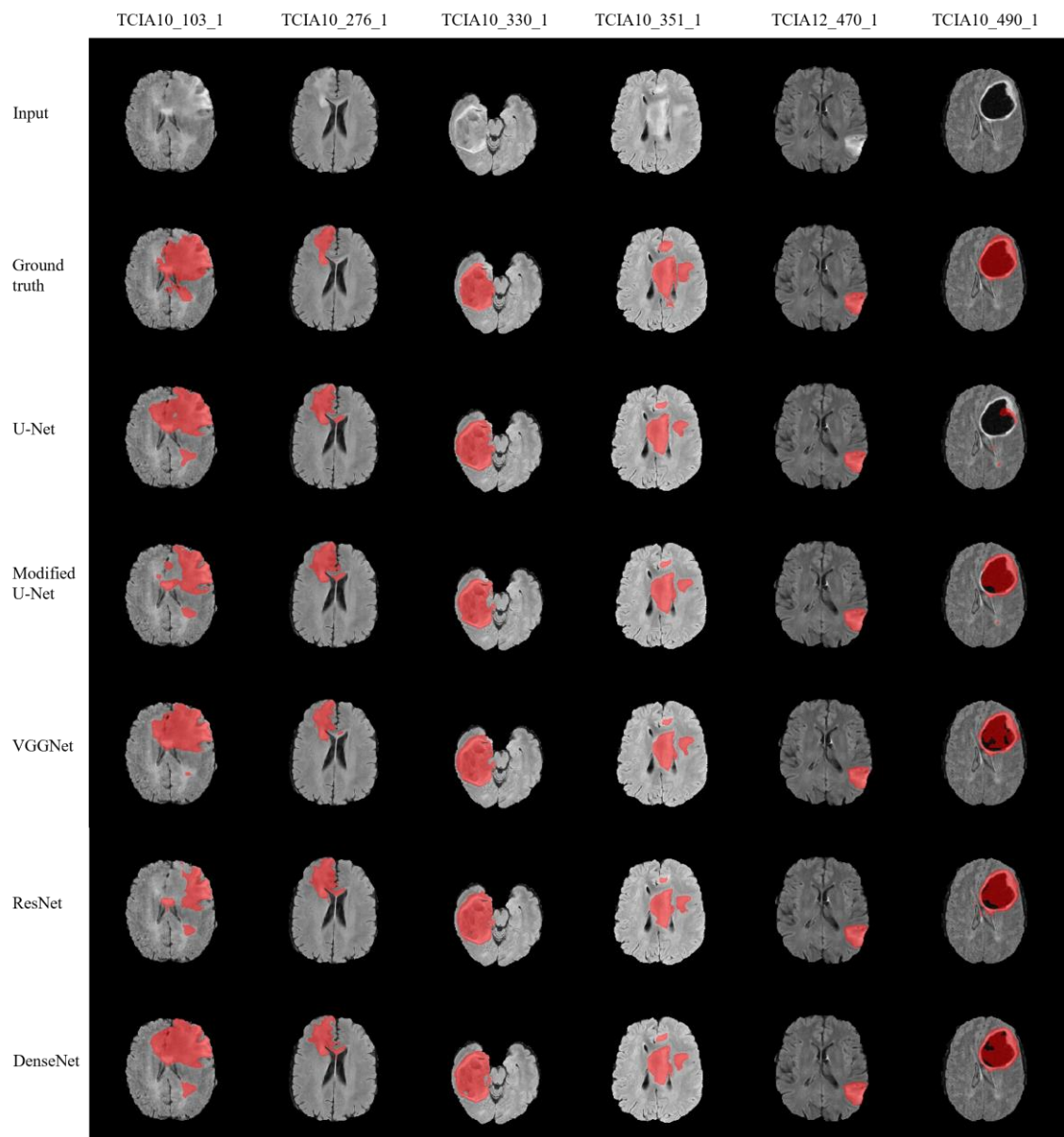


Figure 4.7: Brain tumor segmentation results. FLAIR, ground truth, and output of Original U-Net, Modified U-Net, VGGNet, ResNet, and DenseNet. Tumor regions are indicated in red.

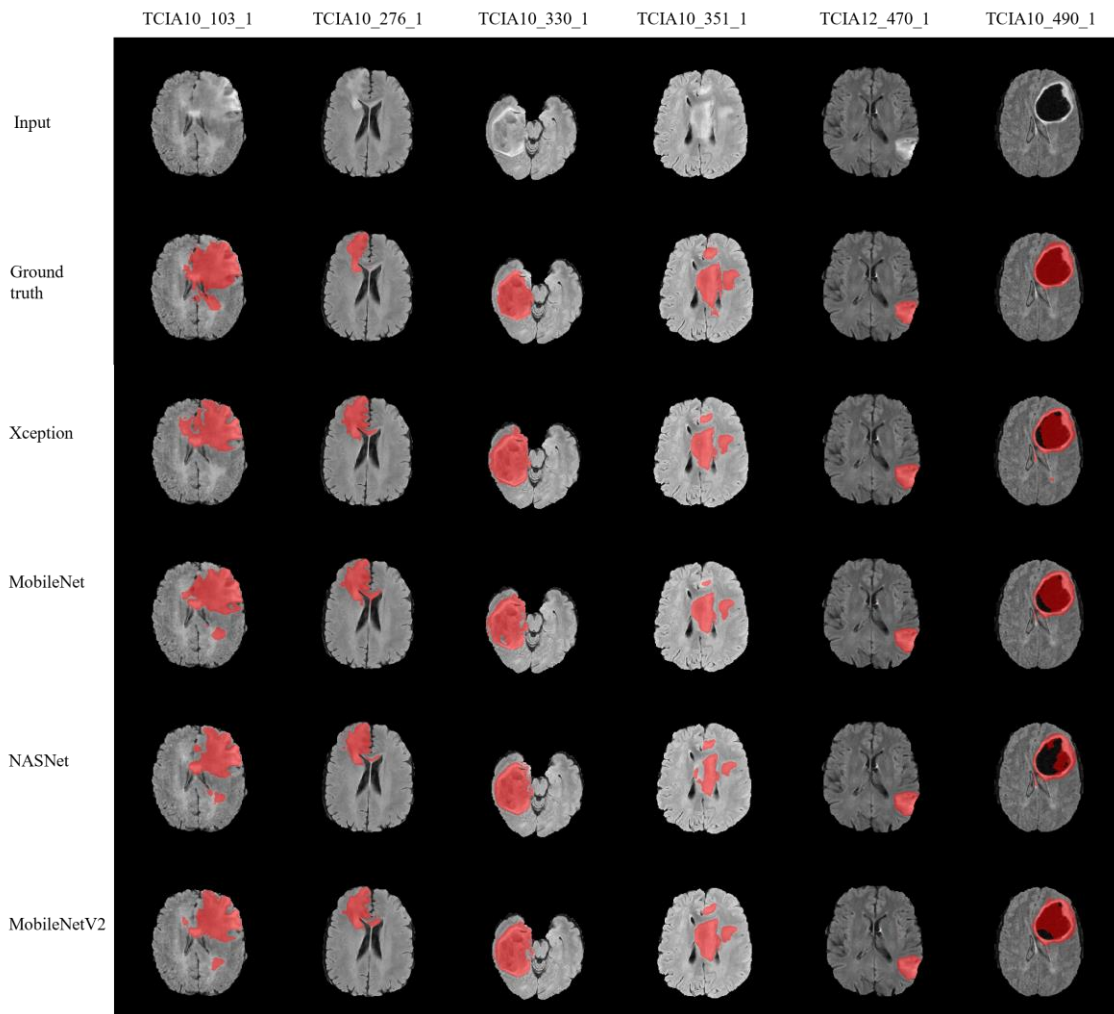


Figure 4.8: Brain tumor segmentation results. FLAIR, ground truth, and output of Xception, MobileNet, NASNet, and MobileNetV2. Tumor regions are indicated in red.

in gray color mode, and the manual ground truth segmentations are shown in the second row. In the following rows, segmentation results of different automated methods are presented. It can be observed that segmented tumor boundaries (indicated in red) from proposed encoders are very similar to the manual annotation even when the lesion region is heterogeneous in shape, volume, and texture. For instance, a small-sized tumor in case TCIA12\_470\_1 was accurately segmented by the proposed methods, however, when the heterogeneity of malignant cells increased, the performance varied remarkably. This is clear in the TCIA10\_103\_1 case since some encoders such as U-Net, VGGNet, and MobileNet tend to over-segment the tumor area, while modified U-Net, ResNet, Xception, NASNet, and MobileNetV2 tend to under-segment. This result showed superior accuracy of the Xception, and DenseNet encoders compared to other tested architectures for the most difficult tumor segmentation case, e.g. the case of TCIA10\_351\_1. Although the DenseNet encoder provided a lower score of tumor segmentation result in the case of TCIA10\_490\_1, it is valid and clinically accepted. However, other encoders such as U-Net and NASNet failed to give accepted segmentation results.

### 4.5.3 Ensemble Results

The results of ensemble models on the BraTS 2021 validation set are summarized in Table 4.5, where the five models for each cross-validation training configuration are averaged as an ensemble. Two evaluation metrics are used for the BraTS 2021 benchmark, computed by the online evaluation platform of Sage Bionetworks Synapse<sup>1</sup>, which is the DSC and the HD95. The averages of DSC scores and HD95 values were computed across the three evaluated tumor sub-regions and then used to rank the methods in the final column.

DeepSeg A refers to the baseline DeepSeg model, which has large input patches of the full pre-processed image, and a smaller batch size of 2. With DSC values of 0.82, 0.84, and 0.90 for the ET, TC, and WT regions, respectively, DeepSeg A model yields good results, especially when compared to the inter-rater agreement range for manual MRI segmentation of GBM (Tacher et al., 2013; Visser et al., 2019). By using a region-based version of DeepSeg with an input patch size of  $128 \times 128 \times 128$  voxels, batch size of 5, applied post-processing stage, and on-the-fly data augmentation, the DeepSeg B model achieved better results of DSC values of 0.84, 0.85, and 0.91 for the ET, TC, and WT regions, respectively.

Additionally, two different configurations of the BraTS 2020 winning approach nnU-Net (Fabian Isensee et al., 2021) and DeepSCAN, one of the BraTS 2018 top-ranking approaches, were used. The first model, nnU-Net A, is a region-based version of the standard nnU-Net with a large batch size of 5 and more aggressive data augmentation as described in (Fabian Isensee et al., 2021), trained using batch Dice loss, and including the postprocessing stage. nnU-Net B model is very similar to the nnU-Net A model with an applied brightness augmentation probability of 0.5 for each input modality compared with 0.3 for model A. nnU-Net models ranks fourth and fifth in our ranking (see Table 4.5) achieving an average DSC and HD95 results of 87.78, 87.87 and 9.60, 10.14 for each model, correspondingly. DeepSCAN model is similar to nnU-Net and DeepSeg model with the output layer as three logits, one for the WT, TC and ET regions rather than using a softmax layer. DeepSCAN model ranks third in the local ranking (see Table 4.5) achieving the best average HD95 of 8.79 while maintaining a good DSC of 0.8739.

In BraTS 2021 Challenge, the Ensemble 2021 model is implemented using DeepSeg B + nnU-Net A + nnU-Net B, while the best ranking HD95, namely the Deep-SCAN model, is included in the Ensemble 2022 model. The final ensemble is implemented by first predicting the validation cases individually with each model configuration, followed by averaging the softmax outputs to obtain the final cross-validation predictions. After that, the STAPLE (Warfield et al., 2004) is applied to aggregate the segmentation produced by each of the individual methods using the probabilistic estimate of the true segmentation. It is worth mentioning that the Ensemble 2021 method is ranked among the ten best teams of the BraTS 2021 challenge, whereas the Ensemble 2022 method is the winner of the BraTS 2022 challenge.

<sup>1</sup> Team name is CaMed; <https://www.synapse.org/#!Synapse:syn25829067/wiki/612712/>

Table 4.5: Results of the five-fold cross-validation models on BraTS 2021 validation cases. All reported values were computed by the online evaluation platform Synapse. The average of DSC and HD95 scores are computed and used for ranking the methods locally.

Model	DSC $\uparrow$				HD95 $\downarrow$				Rank
	ET	TC	WT	Avg	ET	TC	WT	Avg	
DeepSeg A	0.8164	0.8400	0.8998	0.8521	19.77	10.25	5.11	11.71	7
DeepSeg B*	0.8356	0.8508	0.9137	0.8667	17.75	11.56	4.15	11.15	6
nnU-Net A**	0.8402	0.8718	0.9213	0.8778	16.03	8.95	3.82	9.60	4
nnU-Net B***	0.8372	<b>0.8784</b>	0.9205	0.8787	17.73	8.81	3.87	10.14	5
DeepSCAN****	0.8306	0.8683	0.9228	0.8739	<b>14.50</b>	7.91	3.95	<b>8.79</b>	3
Ensemble 2021 (*, **, ***)	0.8410	0.8733	0.9200	0.8781	16.02	8.91	3.81	9.58	2
Ensemble 2022 (*, ***, ****)	<b>0.8423</b>	<b>0.8762</b>	<b>0.9270</b>	<b>0.8819</b>	17.50	<b>7.53</b>	<b>3.60</b>	9.54	1

- Bold values correspond to higher scores

Figure 4.9 shows the qualitative segmentation predictions of the Ensemble 2022 model on the BraTS 2021 validation dataset. These outcomes were generated by applying the ensemble model. The rows show the best, median, and worse segmentations based on their DSC scores, respectively. From this figure, it can be seen that the model achieves very good results with overall high quality. Although the worst case, BraTS2021\_Validation\_01774, has a TC of zero, this finding was not quite surprising as illustrated in Section 4.3.4 as a side effect of applying the postprocessing strategy. Notably, the WT region was detected with a good quality (DSC of 0.9606) which could be already valuable for clinical use.

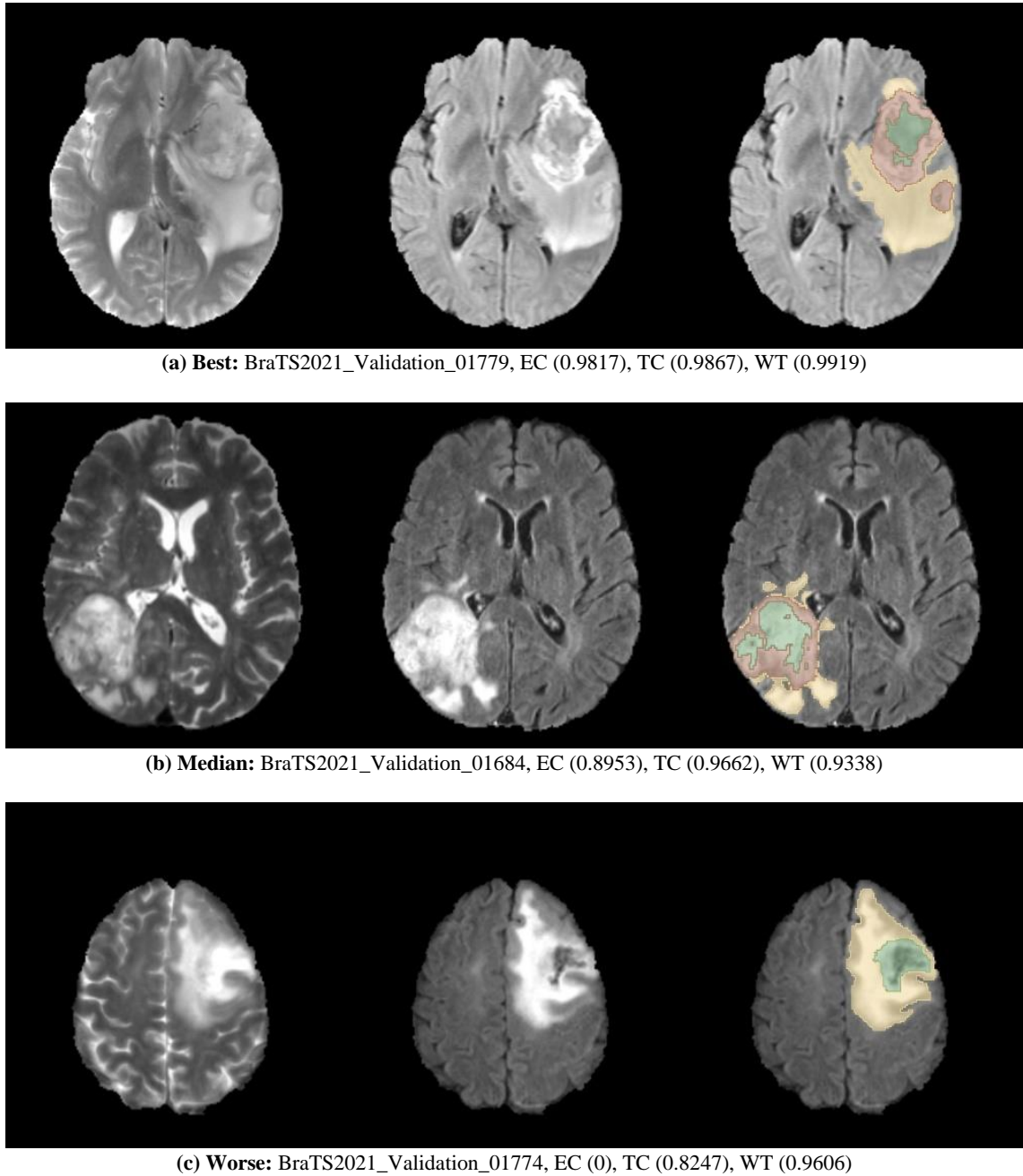


Figure 4.9: Sample qualitative validation set results of the ensemble model. The best, median, and worse cases are shown in the rows. Columns display the T2, FLAIR, and overlay of the predicted segmentation on the FLAIR image. WT includes all visible labels (green, yellow and red labels), TC is a union of green and red, while the ET class is shown in red.

## 4.6 Summary

This chapter demonstrated the feasibility of employing deep learning approaches for assisting the procedures of brain surgery. The DeepSeg framework is developed successfully for fully automated segmenting of brain tumors in MR FLAIR images, based on different architectures of deep CNN models. Moreover, the findings of this comparative study have been validated using the BraTS online evaluation platform, as illustrated in Table 4.4.

Furthermore, two ensemble models of two encoder-decoder-based CNN networks were proposed namely, DeepSeg (R. A. Zeineldin et al., 2020), nnU-Net (Fabian Isensee et al., 2021), and DeepSCAN (McKinley et al., 2019). Table 4.5 list the results of the methods on the BraTS challenge validation set. Remarkably, the method achieved DSC of 92.00, 87.33, and 84.10 as well as HD95 of 3.81, 8.91, and 16.02 for, ET, TC, and WT regions on the validation dataset, respectively. For the testing dataset, the final ensemble yielded DSC of 87.63, 87.49, and 91.87 in addition to HD95 of 12.1343, 14.8915, and 6.2716 for ET, TC, and WT regions, correspondingly. It is worth mentioning that the Ensemble 2021 and Ensemble 2022 methods are ranked among the ten best teams and ranked first in the BraTS 2021 and BraTS 2022 challenges, respectively. The qualitative evaluation also supports the numerical evaluation showing a high-quality segmentation. Clinical partners suggested that this approach can be applied for guiding brain tumor surgery.

# 5 Multimodal Registration

---

## Contents

<b>5.1</b>	<b>Introduction</b>	<b>54</b>
<b>5.2</b>	<b>Related Work</b>	<b>54</b>
<b>5.3</b>	<b>Methods</b>	<b>55</b>
<b>5.3.1</b>	<b>Deformable Image Registration</b>	<b>55</b>
<b>5.3.2</b>	<b>Learning-based Registration Framework</b>	<b>56</b>
<b>5.3.2.1</b>	<b>Deep Neural Network Architecture</b>	<b>56</b>
<b>5.3.2.2</b>	<b>Loss Functions</b>	<b>58</b>
<b>5.3.2.3</b>	<b>Global Optimization</b>	<b>59</b>
<b>5.3.3</b>	<b>Self-supervised iRegNet</b>	<b>59</b>
<b>5.4</b>	<b>Experiments</b>	<b>60</b>
<b>5.4.1</b>	<b>Data</b>	<b>60</b>
<b>5.4.2</b>	<b>Registration Procedure</b>	<b>61</b>
<b>5.4.3</b>	<b>Experimental Setup and Evaluation</b>	<b>62</b>
<b>5.4.3.1</b>	<b>Supervised Approach</b>	<b>62</b>
<b>5.4.3.2</b>	<b>Self-supervised Approach</b>	<b>63</b>
<b>5.5</b>	<b>Results and Discussion</b>	<b>64</b>
<b>5.5.1</b>	<b>Quantitative Registration Results</b>	<b>64</b>
<b>5.5.2</b>	<b>Generality Evaluation</b>	<b>64</b>
<b>5.5.3</b>	<b>Processing Time Analysis</b>	<b>66</b>
<b>5.5.4</b>	<b>Comparison with Other MRI-iUS Studies</b>	<b>68</b>
<b>5.5.5</b>	<b>Qualitative Analysis by Neurosurgeons</b>	<b>70</b>
<b>5.5.6</b>	<b>Self-supervised Results</b>	<b>72</b>
<b>5.6</b>	<b>Summary</b>	<b>75</b>

The chapter describes the second module called *Registration of pre-operative MRI to iUS images for brain shift compensation*, proposed in the IGN pipeline (see Figure 1.2 “DeepIGN”). The next section highlights the limitations of classical registration methods and the need for automatic multimodal registration methods. The state-of-the-art registration methods are mentioned in Section 5.2. Section 5.3 describes briefly medical image registration and the proposed MRI-iUS registration workflow. In Section 5.4, experiments and the employed registration procedure are presented. Finally, experimental results on two datasets in multiple locations are discussed in Section 5.5, and conclusions are summarized in Section 5.6. Some parts of this chapter have already appeared in (Ramy A. Zeineldin et al., 2020; Ramy A. Zeineldin, Karar, Elshaer, et al., 2021; Ramy A. Zeineldin, Karar, Mathis-Ullrich, & Burgert, 2021, 2022b).

## 5.1 Introduction

Imaging has proven to be a valuable tool for assisting neurosurgeons in the planning, interventional, and post-operative clinical phases (Coburger & Wirtz, 2019; Miner, 2017). Yet, achieving accurate lesion localization and differentiation from the surrounding anatomical structures remains a challenging task in neurosurgery. This challenge is related to the difficulty of visually defining these pathologic structures from healthy tissue in addition to the brain movements, known as “brain shift”, due to neurosurgical manipulation, gravity, and anesthesia (De Momi et al., 2016).

Hence, iMRI and iUS images have been used as compensation for brain shifts during surgery (Delorenzo et al., 2010). The iMRI scanner however limits the physician’s access to the operative field and special surgical tools are required, which may be associated with high costs. iUS is portable, inexpensive, requires little preparation, and provides fast data acquisition. Though iUS can visualize interior soft tissue and structures, it has difficulty of imaging through bones, and its high dependency on inter-operator interpretation may result in image inconsistency. Consequently, the fusion of pre-interventional MRI images with the iUS data acquired intra-operatively is proposed to compensate for the brain shift to enable guided surgery.

Therefore, an automatic, fast, robust fusion of 3D-reconstructed iUS data with the pre-operative MRI images becomes highly important to accomplish interventional procedures. However, the registration of misaligned pre-operative MRI images to the iUS is still a complex and challenging problem according to the type of information represented by each modality. Previous studies of medical image registration can be categorized into classical and learning-based methods (Grant Haskins, Kruger, & Yan, 2020; Liu et al., 2019; Sotiras, Davatzikos, & Paragios, 2013).

## 5.2 Related Work

Classical or non-learning MRI to US image registration methods is formulated as an iterative pair-wise optimization problem that requires proper feature extraction, choosing a similarity measurement, defining the used transformation model, and finally an optimization mechanism to investigate the search space. Over time, an extensive literature has developed using diverse combinations of the similarity metrics such as Correlation Coefficient (CC), Correlation Ratio (CR), Mutual Information (MI), Normalized Correlation Coefficient (NCC), Self-Similarity Correlation (SSC), and Linear Correlation of Linear Combination ( $LC^2$ ) (De Nigris, Collins, & Arbel, 2013; Drobny, Vercauteren, Ourselin, & Modat, 2018; Mattias P. Heinrich, 2018; Mattias Paul Heinrich, Jenkinson, Papież, Brady, & Schnabel, 2013; Jiang, Shi, Yao, Wang, & Song, 2016; I. Machado et al., 2019; Inês Machado et al., 2018; Masoumi, Xiao, & Rivaz, 2019; Rivaz, Chen, & Collins, 2015; Rivaz, Karimaghloo, Fonov, & Collins, 2014; Shams, Boucher, & Kadoury, 2018; Wein, 2018; Wein et al., 2013; Zimmer, Ballester, & Piella, 2019). One major drawback of the traditional registration methods is the high computational cost required to align every 3D MRI and iUS pair even with the efficient implementation on modern GPUs.

To overcome the limitations of classical methods, learning-based approaches have been proposed in recent years. Recently, deep learning, a subfield of artificial intelligence, has gained



increasing popularity because of its outstanding performance in various computer vision and image analysis applications including, but not limited to object detection (Krizhevsky et al., 2017), feature extraction (He et al., 2016), image segmentation (Ronneberger et al., 2015; R. A. Zeineldin et al., 2020), image classification (Saleh et al., 2018), and other medical applications (Armanious et al., 2020; Dolz et al., 2020). The learning process is achieved through backpropagation, which is a feedback loop for computing the partial derivative of the cost function with respect to the network weights (LeCun et al., 2015). Initially, supervised deep learning methods were proposed (Cheng, Zhang, & Zheng, 2018; Rohé, Datar, Heimann, Sermesant, & Pennec, 2017; Sun & Zhang, 2018; Zhong et al., 2018) to learn similarity features from the training data using different imaging modalities. Then, deep learning methods have been applied to solve the challenging multi-modal medical image registration problem where images from different image sensors are aligned together (G. Haskins et al., 2019; Hu et al., 2018; Lee, Liu, Cheng, & Fu, 2019; Ma et al., 2017; X. Yang, Kwitt, Styner, & Niethammer, 2017).

Further, unsupervised learning was developed as a demand for faster registration procedures and to eliminate the challenges related to the ground truth data generation and optimization techniques (Balakrishnan, Zhao, Sabuncu, Guttag, & Dalca, 2019; de Vos et al., 2019; H. Li & Fan, 2018). Although the proposed method, iRegNet, uses an encoder-decoder-based method similar to (X. Yang et al., 2017) and (Balakrishnan et al., 2019), this is, to the best of our knowledge, the first study to use the truth warped images as the target of the registration rather than the fixed image which provides more precise information for the training process. Unsupervised approaches show promising registration results; however, it is still difficult to apply them to multi-modal registration of MRI and iUS applications since represented information originates from very distinct physical properties. Lately, CNN learning methods have been introduced as part of the Correction of Brain shift with Intra-Operative Ultrasound challenge (CuRIOUS) (Xiao et al., 2020) in conjunction with the MICCAI (Sun & Zhang, 2018; Zhong et al., 2018). Overall, once the deep learning networks are trained, they can provide a faster registration than classical optimization methods, without the need for fine-tuning parameters at the test time, in addition to being more robust to outliers.

## 5.3 Methods

### 5.3.1 Deformable Image Registration

Image registration is the process of finding spatial correspondences between two or more images (Karar et al., 2016; Kneöaurek, Ivanovic, Machac, & Weber, 2000; Sotiras et al., 2013). Within the medical field, image registration is attractive for providing more information when the imaging data come from different sources and/ or different modalities. The term deformable denotes that the images are related through non-linear spatial deformation and the resultant transformation not only includes rigid operations (such as rotation and translation) but also non-uniform operations like shearing. Consider two sets of images: the source or moving image  $I_M$  and the target or fixed image  $I_F$ . Then, the goal of deformable registration is finding the optimal deformation field  $\phi$  that relates the two images while optimizing the energy function  $\varepsilon$ :

(5.1)

where  $S$  quantifies the level of similarity between the  $I_M$  and  $I_F$  images and the regularization term  $\mathbb{R}(\phi)$  allows a smooth transformation aiming to fulfill any user-specific requirements. In this study, the MRI is used as the moving image and iUS as the fixed image so that the MRI could be updated to reflect the brain shift intra-operatively.

### 5.3.2 Learning-based Registration Framework

In this section, the proposed registration workflow is described (Fig. 5.1). Section 5.3.2.1 presents the developed CNN. The loss functions of the CNN are illustrated in Section 5.3.2.2), and the global optimization of the deformation field is explained in Section 5.3.2.3).

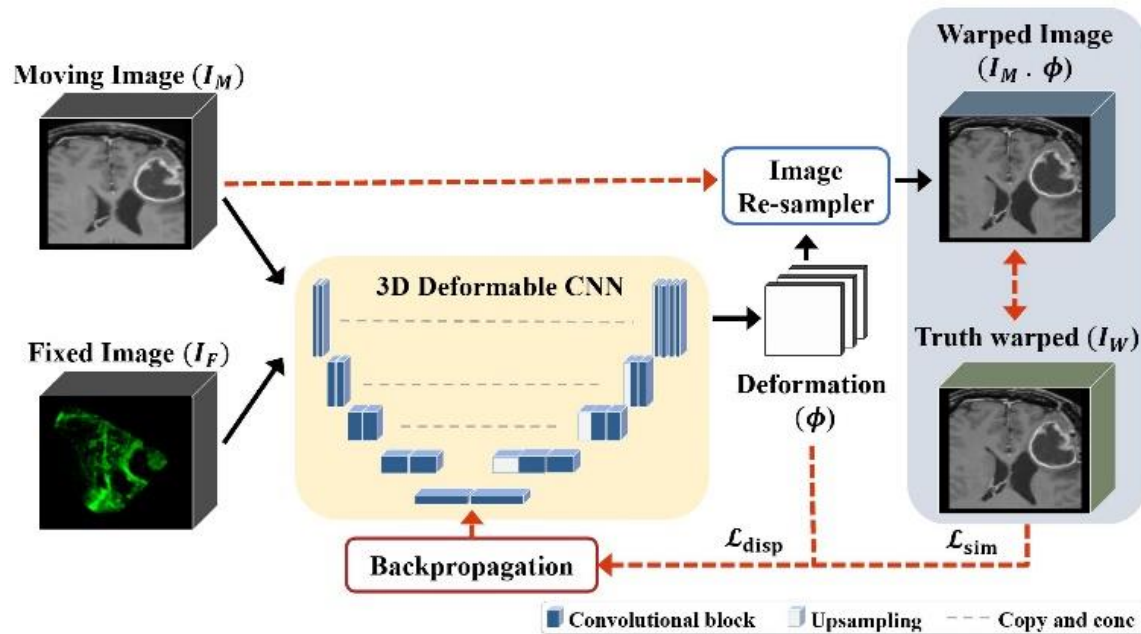


Figure 5.1: A representative workflow of the proposed MRI-iUS deformable registration approach, where dashed red arrows indicate the data flow only required in the training stage.

#### 5.3.2.1 Deep Neural Network Architecture

Fig. 5.2 depicts the overall architecture of the proposed CNN utilized in the experiments. The proposed network is based on the U-Net structure (Ronneberger et al., 2015), which has been widely utilized in various medical applications achieving competitive performance. A 3D version of U-Net is introduced by Çiçek et al (Çiçek et al., 2016), in which 3D operations are applied instead of the standard 2D processes. Besides, several studies have demonstrated enhancements to the original U-Net (de Vos et al., 2019; Grant Haskins et al., 2020; R. A. Zeineldin et al., 2020).

Similar to the standard U-Net, the proposed network has an encoder-decoder architecture with an image analysis path (left side) and a deformation estimation path (right side). As shown in

Fig. 5.2, the feature analysis is designed as a 3D CNN with four repetitive down-sampling blocks. Table 5.1 lists the detailed implementation of each layer in the 3D CNN. The input to the network is  $128 \times 128 \times 128$  voxels with 2 channels of the MRI and iUS images. Each encoder block consists of two consecutive  $3 \times 3 \times 3$  convolutional layers (unpadded convolutions) with a stride of 2, each followed by a ReLU and BN. At each down-sampling step, the number of feature maps is doubled, while the spatial dimension is halved using 3D spatial max pooling.

Initialization of all convolutional kernels is done by using the Glorot or Xavier uniform (Glorot & Bengio, 2010) with the default bias set to zeros. It is worth noting that this contracting architecture is similar to the classical pyramid image registration scheme.

To restore the original image resolution, four up-sampling blocks are adopted in the deformation estimation path. Every step in the up-sampling path is composed of a  $3 \times 3 \times 3$  transposed convolutional layer (up-convolution) with a stride of 2 followed by a ReLU and BN. Dissimilar to the encoder, using up-convolution doubles the input spatial resolution and halves the number of feature maps. The high-level features in the encoding path are concatenated with the corresponding low-level features in the decoding path via skip connections. As the output layer, a  $1 \times 1 \times 1$  convolutional layer is incorporated to get the output deformation field with a dimension of  $128 \times 128 \times 128$  in x, y, and z directions, respectively.

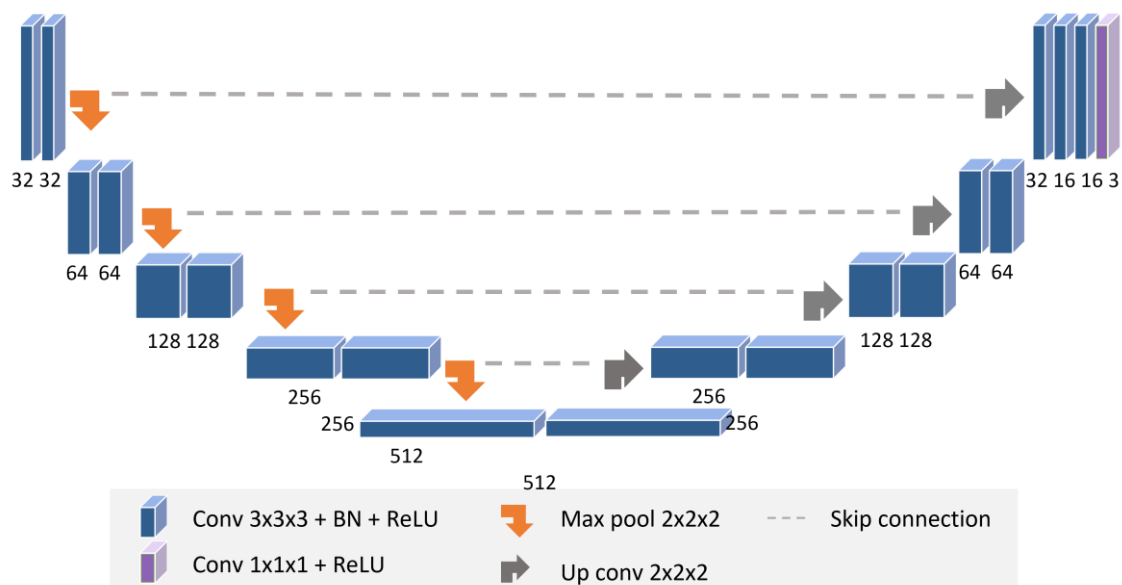


Figure 5.2: The enhanced 3D CNN architecture for predicting the deformation field  $\phi$ . The network consists of 3D convolution with batch normalization layers (blue blocks), maximum pooling (orange arrows), up-sampling (grey arrows), and concatenate connections (dashed grey lines). The number of features is doubled in each step of the encoder part, while halved in the decoder part.

Table 5.1: The deformable CNN architecture details.

#	Operation	Output	#	Operation	Output
0	Input	128×128×128×1		Concatenate	32×32×32×320
	Input	128×128×128×1		Conv3D/ BN/ ReLU	32×32×32×128
	Concatenate	128×128×128×2		Conv3D/ BN/ ReLU	32×32×32×128
1	Conv3D/ BN/ ReLU	64×64×64×32	9	UpSampling3D	64×64×64×128
2	Conv3D/ BN/ ReLU	32×32×32×64		Concatenate	64×64×64×160
3	Conv3D/ BN/ ReLU	16×16×16×128		Conv3D/ BN/ ReLU	64×64×64×64
4	Conv3D/ BN/ ReLU	8×8×8×256		Conv3D/ BN/ ReLU	64×64×64×64
5	Conv3D/ BN/ ReLU	8×8×8×512	10	UpSampling3D	128×128×128×32
	Dropout	8×8×8×512		Concatenate	128×128×128×34
7	UpSampling3D	16×16×16×512		Conv3D/ BN/ ReLU	128×128×128×16
	Concatenate	16×16×16×640		Conv3D/ BN/ ReLU	128×128×128×16
	Conv3D/ BN/ ReLU	16×16×16×256	11	Conv3D/ BN/ ReLU	128×128×128×16
	Conv3D/ BN/ ReLU	16×16×16×256	12	Conv3D	128×128×128×3
8	UpSampling3D	32×32×32×256			

### 5.3.2.2 Loss Functions

The choice of the loss function plays a crucial role in network training and contributes to the overall performance of CNN. As illustrated in Fig. 5.1, the overall loss function  $L_{overall}$  consists of two main elements (refer to Equation (2)). The similarity measurement between the resultant deformed image ( $\phi \cdot I_M$ ) and the ground truth warped image  $I_W$  is denoted by  $L_{sim}$ , while  $L_{disp}$  represents the spatial deformation gradient error.

$$L_{overall} = L_{sim} + L_{disp} \quad (5.2)$$

In the experiments, two distinct similarity metrics of the mean squared error ( $MSE$ ) and the local normalized correlation coefficient ( $NCC$ ) are employed as  $L_{sim}$ . Let  $I_M(P)$  and  $I_W(P)$  represent a corresponding patch  $P$  of  $N_P$  patches in the moving and truth-warped images, respectively.  $MSE$  and  $NCC$  are calculated as follows:

$$MSE(I_W, \phi \cdot I_M) = \frac{1}{|N_P|} \sum_{P \in N_P} ((I_W(P) - \phi \cdot I_M(P)))^2 \quad (5.3)$$

$$NCC(I_W, \phi \cdot I_M) = \frac{1}{N_P} \sum_{P \in N_P} \frac{\sum_i (I_W(P) - \overline{I_W(P)}) \sum_i (\phi \cdot I_M(P) - \overline{\phi \cdot I_M(P)})}{\sqrt{\sum_i (I_W(P) - \overline{I_W(P)})^2} \sqrt{\sum_i (\phi \cdot I_M(P) - \overline{\phi \cdot I_M(P)})^2}} \quad (5.4)$$

where  $I_W(P)$  denote the mean pixel intensities for the warped image. For MRI-iUS registration, it is very important to choose modality-invariant similarity metrics that can evaluate the similarity between MRI and iUS images after the deformable alignment. The *NCC* similarity is invariant to scaling and linear intensity variations, which makes it more preferable in the current application. Similar to traditional registration approaches, the second loss works as a regularization term preventing a non-smooth deformation field. Let  $d_{pred}$  denote the predicted spatial gradient and  $d_{truth}$  denote the ground truth gradient. Then,  $L_{disp}$  can be calculated as follows:

$$L_{disp} = \sum_{p \in X} \|d_{truth}(P) - d_{pred}(P)\| \quad (5.5)$$

### 5.3.2.3 Global Optimization

Traditionally, the optimization of the deformation field has been formulated as an iterative pairwise optimization problem (refer to equation (1)). This is a computationally expensive problem that consumes a large processing time and may last for hours for a single pair depending on the used CPU. In contrast, deep learning methods recast the classical optimization problem into a problem of cost function estimation (LeCun et al., 2015). In other words, this formulates the problem to find a function that takes a pair of MRI-iUS images and directly computes the output deformation field using backpropagation, which makes the proposed method optimize over the whole training set moving away from expensive iterative optimization.

## 5.3.3 Self-supervised iRegNet

In this sub-section, a fully automatic, patient-specific registration approach is proposed for pre- and post-operative brain MRI sequences of only a single modality using iRegNet (Ramy A. Zeineldin, Karar, Elshaer, et al., 2021). In particular, an unsupervised approach of iRegNet is introduced (see Fig. 5.2) in which only moving and fixed MRI pairs are utilized. Then, the proposed method optimizes deformation fields directly from input images using backpropagation.

More formally, self-supervised learning is incorporated to compute the optimal deformation field  $\hat{\phi}$  corresponding to the smoothness regularization. This model uses only the input MRI volume pair, and the registration field is computed accordingly by the CNN network. Formally, this task is defined as:

$$\hat{\phi} = \arg \min_{\phi} \mathcal{L}_{sim}(I_F, \phi \cdot I_M) + R(\phi) \quad (5.6)$$

where  $\mathcal{L}_{sim}$  computes the image similarity between the warped image ( $\phi \cdot I_M$ ) and the fixed image  $I_F$ ,

## 5.4 Experiments

### 5.4.1 Data

In this study, the proposed iRegNet method was tested on two public multi-center databases, which are BITE (Mercier, Del Maestro, et al., 2012) and RESECT (Xiao et al., 2017). BITE is the first online dataset for tracked 3D iUS volumes of the brain alongside pre-operative MRI images. It contains 14 patients with either LGG or HGG from the Montreal Neurological Institute, Canada. Nevertheless, the technology used to collect the iUS in the BITE dataset is no longer up-to-date and recent US scanners provide improved quality and higher-resolution images. Consequently, RESECT was proposed to overcome this problem and help, therefore, develop image registration techniques for brain shift compensation. The dataset contains pre-operative MRI and iUS images from 22 patients with LGG who have received surgeries at St. Olavs University Hospital, Norway. Table 5.2 gives detailed information on the applied datasets with a wide variety of (a) data-acquiring locations, (b) patient and tumor details, and (c) applied MRI and US protocols.

Expert-labeled anatomical markers were provided for both databases, to facilitate the baseline evaluation of MRI to iUS registration. The procedure of generating these landmarks highly depends on the employed datasets. For the BITE dataset, homologous landmark points were chosen manually by at least two experts. An average of nine landmarks was provided in the BITE database for each patient. For the RESECT dataset, more landmarks were provided which may lead to better validation of registration accuracy. In both datasets, the package named ‘register’, included in the MINC toolkit (Vincent et al., 2016), was used to visualize the 3D MRI-US pairs and produce the homologous landmarks. It should be noted that this process was performed only in the training stage, however, during the inference, no landmarks are utilized to get the corrected MRI volumes. For each pair in BITE, the homologous landmarks from the first two experts (D.L.C. and L.M.) were used without the third expert’s labels since they tagged only up to patient #6 (Mercier, Fonov, et al., 2012).

Furthermore, the Brain Tumor Sequence Registration Challenge 2022 (BraTS-Reg 2022) dataset was utilized to evaluate the proposed self-supervised network. The BraTS-Reg 2022 dataset (Baheti et al., 2021) comprises 250 patient-specific pairs of pre-operative and follow-up brain multi-institutional MRI scans. For each patient, i) native T1-weighted (T1), ii) contrast-enhanced T1 (T1ce), iii) T2-weighted (T2), and iv) T2 Fluid Attenuated Inversion Recovery (FLAIR) sequences are provided for the pre-operative and follow-up with a time-window in the range of 27 days to 37 months. Reference landmark annotations for the validation set are not made available to the participants. Instead, participants can use the online evaluation platform<sup>1</sup> to evaluate their models and compare their results with other teams on the online leaderboard<sup>2</sup>.

---

<sup>1</sup> BraTS Evaluation Plattform; <https://ipp.cbica.upenn.edu/>

<sup>2</sup> BraTS Leaderboard; <https://www.cbica.upenn.edu/BraTSReg2022/lboardValidation.html/>

Standard pre-processing techniques were applied such as rigid registration to the same anatomical template, resampling to the same isotropic resolution ( $1\text{mm}^3$ ), skull removal, and brain extraction. Following these pre-processing steps, the image cropping stage is applied where all brain pixels were cropped. Afterward, z-score normalization was applied by subtracting the mean value and dividing it by the standard deviation individually for each input MRI image.

## 5.4.2 Registration Procedure

As listed in Table 5.2, it is evident that there are several dissimilarities between the two datasets in terms of imaging locations, study characteristics, followed MRI and iUS protocols, and, therefore, a preprocessing step is essential before performing the MRI-iUS registration. First, the ultrasound images are resampled to the isotropic  $1 \times 1 \times 1 \text{ mm}^3$  voxel size, the same as the MRI spatial resolution.

Table 5.2: A detailed description of the two databases used in this study.

		BITE	RESECT
(a) Imaging Site		Montreal Neurological Institute, Montreal, Canada	St. Olavs University Hospital, Trondheim, Norway
(b) Study Characteristics	N. of Patients	14	22
	Tumor Type	LGG (4) HGG (9)	LGG (22)
	N. of Landmarks	355	338
(c) MRI Protocol	MRI Scanner	1.5T General Electric Signa EXCITE	3T Siemens Magnetom Skyra and 1.5T Siemens Magnetom Avanto
	Date of Acquisition	Avg. 17 days before surgery (1–72 days)	1 day before surgery
	MRI Modalities	T1w Gd-enhanced	T1w Gd-enhanced and T2w fluid-attenuated inversion recovery (FLAIR)
	MRI Resolution	$256 \times 256 \times 256$	$256 \times 256 \times 192$
	Voxel Size	$1 \times 1 \times 1 \text{ mm}^3$	$1 \times 1 \times 1 \text{ mm}^3$
(d) iUS Protocol	iUS Probe	Phased-array transducer	12FLA-L linear probe
	iUS Frequencies	7-4 MHz	6-12 MHz
	iUS Resolution	$0.3 \times 0.3 \times 0.3 \text{ mm}^3$	$0.14 \times 0.14 \times 0.14 \text{ mm}^3$ to $0.24 \times 0.24 \times 0.24 \text{ mm}^3$

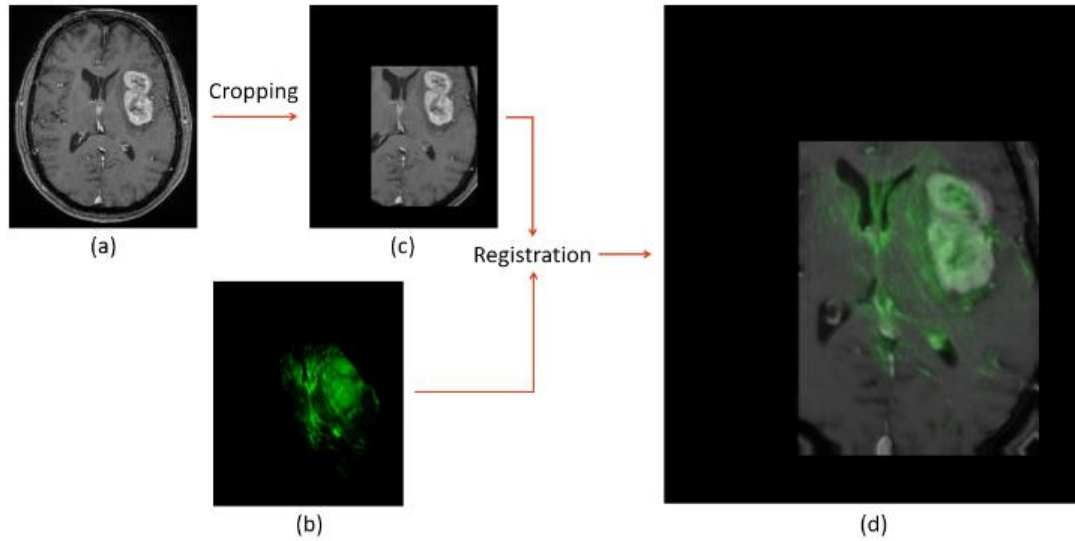


Figure 5.3: The MRI-iUS registration procedure (a) Input MRI; (b) Input iUS; (c) Cropped MRI; (d) An overlap between iUS and corrected MRI.

Second, iUS images are pad to  $128 \times 128 \times 128$  voxels to make them suitable for the DNN input. Third, the initial alignment of MRI images to iUS data is used, and then the MRI is cropped to match the field of view of the iUS, as shown in Fig. 5.3. Fourth, z-score normalization is applied by subtracting the mean value and dividing by the standard deviation individually for each input volume. Fifth, an affine MRI to iUS alignment is achieved using the MINC toolkit (Vincent et al., 2016) to focus on the non-linear misalignment. Finally, truth deformation fields for all patients were computed using the software named ‘register’, included in the MINC toolkit so that the deformation field gradient error  $L_{disp}$  could be estimated (refer to Section 5.3.2.1).

## 5.4.3 Experimental Setup and Evaluation

### 5.4.3.1 Supervised Approach

As the number of cases is rather limited, intensive data augmentation methods were applied to help prevent the model from overfitting and improve the registration results. This involves random 3D flipping, 3D rotations [0-30 degrees], random gamma intensity transformation [0.8-1.2], and elastic deformation. For the experiments, the model was implemented in Python using the TensorFlow library (Abadi et al., 2016). The experiments were conducted on an AMD Ryzen 2920X (32M Cache, 3.50 GHz) CPU with 32 GB RAM and a single NVIDIA RTX 2080Ti GPU with 11GB. For training the network, the cases are divided into two sets 78% for the training set and 22% for the validation set, the Adam optimizer (Kingma & Ba, 2014) with an initial learning rate of 0.0001, and a batch size of 2 was used. All iRegNet models were trained for 500 epochs, with one epoch being defined as an iteration over 500 mini-batches.

The proposed approach was evaluated using three configurations: The first configuration, referred to as Model BITE or *Model B*, involved training on the BITE dataset only. The second



configuration, referred to as Model RESECT or *Model R*, involved training on the RESECT dataset only. The last configuration, referred to as Model Combined or *Model C*, involved training on both BITE and RESECT datasets. It is important to note that a total of six experiments, which use affinely aligned MRI-iUS images, were conducted for the registration of MRI-iUS based on alternating NCC and MSE as a similarity measurement for training the CNN. For the training phase, 11, 17, and 28 pairs of MRI and iUS images are used for *Model B*, *Model R*, and *Model C*, respectively. Whereas 3, 5, and 8 pairs of MRI and iUS images are used for validation in the same order.

The two datasets, BITE and RESECT, provide expert-annotated landmarks for each corresponding MRI-iUS pair (as summarized in Table 5.2). In line with previous studies (De Nigris et al., 2013; Drobny et al., 2018; Mattias P. Heinrich, 2018; Hong & Park, 2018; Jiang et al., 2016; Inês Machado et al., 2018; Masoumi et al., 2019; Rivaz et al., 2015; Rivaz et al., 2014; Shams et al., 2018; Sun & Zhang, 2018; Wein, 2018; Wein et al., 2013; Zhong et al., 2018), the *mTRE* is used, which represents the average pair-wise distance between the corresponding points in MRI and iUS volumes after registration. Let  $m$  and  $u$  denote the expert-labeled corresponding annotations in the MRI and iUS volumes, respectively. The *mTRE* of  $N_l$  corresponding landmarks, following the registration, is calculated as follows:

$$mTRE = \frac{1}{N_l} \sum_{i=1}^{N_l} \|\phi(m_i) - u_i\| \quad (5.7)$$

where  $\|r\|$  is the L2 norm of the vector  $r$ . The evaluation of the experiments was performed using the same approach reported in (Mattias P. Heinrich, 2018).

#### 5.4.3.2 Self-supervised Approach

To ensure computational efficiency for the GPU, MRI scans for each patient were center cropped to  $160 \times 192 \times 160$  pixels. For training and validation, sets, respectively, the BraTS-Reg dataset was randomly split into 112 (80%) and 28 (20%) volumes. Another 20 MRI volumes were provided by the BraTS-Reg organizers as online validation set with landmarks provided only for the fixed MRI scans. Finally, an affine alignment was performed on moving and fixed MRI volumes using the BRAINSFit toolkit (Johnson, Harris, & Williams, 2007) to focus on the non-linear misalignment between volumes. For the experiments, the model was implemented in Python 3.7 using the TensorFlow 2.4 library. The experiments were run on an AMD Ryzen 2920X (32M Cache, 3.50 GHz) CPU with 64 GB RAM and a single NVIDIA GPU (RTX 3060 12 GB or RTX 2080 Ti 11 GB). The Adam optimizer (Kingma & Ba, 2014) with an initial learning rate of  $1e-4$  and a batch size of 2 was used.

To compare with other studies, the mean *mTRE* (Equation (5.7)), which represents the average distance between the corresponding landmarks in each pre-post MRI pair before and after registration, was used. In addition, the proposed method was evaluated by the online submission platform using the following metrics, namely Median Absolute Error (MAE), robustness, and smoothness of the displacement field.

## 5.5 Results and Discussion

### 5.5.1 Quantitative Registration Results

Tables 5.3 and 5.4 summarize the *mTREs* of pre-and post-registration of the three proposed methods for all the trained 14 BITE and 22 RESECT cases, individually. The last row of each table denotes a summary of the results of the listed cases. The results show that the proposed methods provide a major improvement over the initial alignment. For the BITE database, the developed algorithms reduced the initial *mTRE* from  $(4.18 \pm 1.84 \text{ mm})$  to a range from  $(1.47 \pm 0.61 \text{ mm})$  to  $(2.00 \pm 0.45 \text{ mm})$  based on the applied configuration. Similarly, average *mTRE* from  $(0.84 \pm 0.16 \text{ mm})$  to  $(2.50 \pm 0.66 \text{ mm})$  was achieved on the RESECT database starting with an initial *mTRE* value of  $(5.35 \pm 4.19 \text{ mm})$ . Notably, the use of NCC as a similarity metric had a great impact on registration accuracy as all NCC-based models show an improvement over the standard MSE-based methods. This proves the effectiveness of using NCC for multi-modal registration as it is intensity and scaling invariant.

From the BITE results, it is important to note that the most accurate results were generated by *Model C NCC*. This result emphasizes that utilizing more training data in deep learning often leads to performance enhancement. However, this did not have the same effect on the RESECT dataset wherein *Model C NCC* ranked third after *Model R* versions. This might be due to the differences in the two databases, refer to Table 5.2, resulting in a model struggling to extract common features. Still, *Model C NCC* improved significantly the initial *mTRE* from  $(5.35 \pm 4.19 \text{ mm})$  to  $(0.84 \pm 0.16 \text{ mm})$ . The results indicated that the deep learning approach can perform automatic accurate deformable MRI to iUS image registration, and thus could be used in image-guided neurosurgical interventions.

### 5.5.2 Generality Evaluation

One of the main goals is to build an automated learning model that can be applied to multi-site data without database-specific model parameters fine-tuning. To achieve this goal, the three proposed models, *Model B*, *Model R*, and *Model C*, were evaluated on both utilized databases. For instance, *Model B* was trained using the BITE dataset only, therefore, testing this model on the other dataset of RESECT would give us a general idea about how the developed method may generalize with other unseen datasets.

Remarkably, *Model B* and *Model B NCC* decreased the initial *mTRE* from  $(5.35 \pm 4.19 \text{ mm})$  to only  $(2.50 \pm 0.66 \text{ mm})$  and  $(1.85 \pm 0.40 \text{ mm})$  experienced on the untrained RESECT dataset. Likewise, the two variations of *Model R*, trained on the RESECT dataset only, delivered astonishing results on the BITE dataset with average *mTRE* from  $(2.00 \pm 0.45 \text{ mm})$  to  $(1.98 \pm 0.44 \text{ mm})$  over the initial  $(4.18 \pm 1.84 \text{ mm})$ . These findings further strengthened the conviction that deep learning-based models can deliver competitive MRI to iUS registration results even if they are not trained on the evaluated data, which shows therefore a promise for use during brain surgery.

Table 5.3: Pre- and post-registration TREs for the proposed models in the BITE dataset. *Model B*, *Model R*, and *Model C* represent the proposed models trained on the BITE, RESECT, and both datasets, correspondingly. Test cases are shown in *bold underlined* and the lowest error in each row is highlighted in *bold*.

Case	Initial	Model B MSE	Model B NCC	Model R MSE	Model R NCC	Model C MSE	Model C NCC
Case01	5.88(2.31)	1.42(0.79)	1.17(0.55)	1.82(0.96)	1.83(0.98)	1.82(0.97)	<b>1.04(0.58)</b>
Case02	6.06(1.61)	1.14(0.67)	1.27(0.77)	1.46(0.83)	1.54(0.82)	1.39(0.80)	<b>0.97(0.45)</b>
Case03	8.91(2.02)	1.57(1.12)	1.39(0.87)	2.24(1.23)	2.19(1.28)	2.11(1.26)	<b>1.22(0.87)</b>
Case04	3.87(1.19)	0.91(0.41)	1.01(0.51)	1.57(0.86)	1.50(0.72)	1.45(0.69)	<b>0.90(0.47)</b>
Case05	2.57(1.61)	1.43(1.09)	1.50(1.00)	2.02(1.51)	2.04(1.50)	2.00(1.51)	<b>1.21(0.80)</b>
Case06	2.24(1.05)	1.21(0.57)	1.24(0.57)	1.57(0.68)	1.52(0.63)	1.50(0.65)	<b>1.08(0.50)</b>
<b>Case07</b>	3.02(1.58)	2.15(0.95)	2.16(0.77)	2.15(0.79)	<b>1.99(0.81)</b>	<b>1.99(0.76)</b>	2.08(0.80)
Case08	3.75(1.97)	<b>1.36(0.67)</b>	1.61(0.81)	2.31(1.11)	2.22(1.06)	2.17(1.03)	<b>1.37(0.69)</b>
Case09	5.08(1.33)	1.60(0.71)	1.87(1.16)	2.43(1.14)	2.48(1.12)	2.39(1.07)	<b>1.49(0.81)</b>
<b>Case10</b>	2.99(1.34)	1.80(0.82)	1.59(0.79)	<b>1.41(0.69)</b>	1.45(0.73)	1.44(0.73)	1.75(0.88)
<b>Case11</b>	1.51(0.73)	1.60(0.84)	1.47(0.73)	1.34(0.64)	<b>1.32(0.68)</b>	<b>1.32(0.68)</b>	1.48(0.69)
Case12	3.68(1.85)	3.14(2.60)	2.91(2.45)	2.74(1.75)	2.70(1.77)	<b>2.66(1.74)</b>	3.38(3.28)
Case13	5.13(2.73)	1.81(1.35)	1.67(1.08)	2.51(1.30)	2.59(1.23)	2.40(1.30)	<b>1.40(0.87)</b>
Case14	3.78(1.23)	1.34(0.64)	1.42(0.78)	2.44(0.91)	2.31(0.85)	2.24(0.87)	<b>1.18(0.43)</b>
mTRE	4.18(1.84)	1.61(0.52)	1.59(0.46)	2.00(0.45)	1.98(0.44)	1.92(0.42)	<b>1.47(0.61)</b>

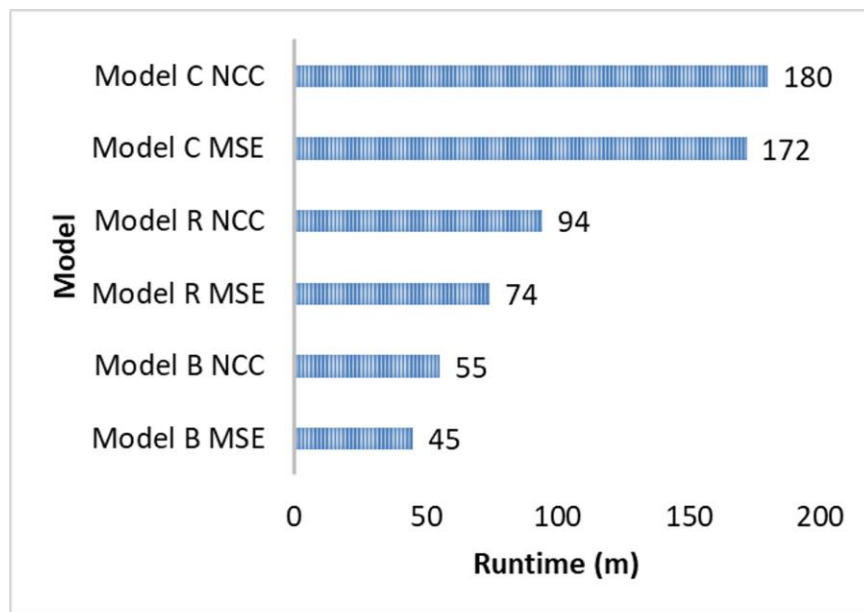
Table 5.4: Pre- and post-registration TREs for the proposed models in the RESECT dataset. *Model B*, *Model R*, and *Model C* represent the proposed models trained on the BITE, RESECT, and both datasets, correspondingly. Test cases are shown in *bold underlined* and the lowest error in each row is highlighted in *bold*.

Case	Initial	Model B MSE	Model B NCC	Model R MSE	Model R NCC	Model C MSE	Model C NCC
<b>Case01</b>	1.81(0.84)	3.71(1.30)	2.07(0.87)	1.17(0.74)	<b>1.13(0.58)</b>	<b>1.13(0.58)</b>	1.42(0.76)
Case02	5.70(1.39)	2.79(0.83)	2.67(0.85)	1.05(0.43)	<b>0.72(0.37)</b>	1.15(0.57)	0.86(0.37)
Case03	9.56(0.52)	3.30(0.97)	1.96(0.68)	0.91(0.38)	<b>0.86(0.34)</b>	0.94(0.37)	0.94(0.37)
<b>Case04</b>	2.45(0.67)	<b>0.89(0.39)</b>	1.04(0.40)	1.26(0.46)	1.04(0.40)	1.04(0.40)	1.72(0.61)
Case05	12.03(1.05)	2.73(0.97)	1.77(0.61)	0.79(0.32)	<b>0.73(0.37)</b>	1.00(0.47)	0.82(0.38)

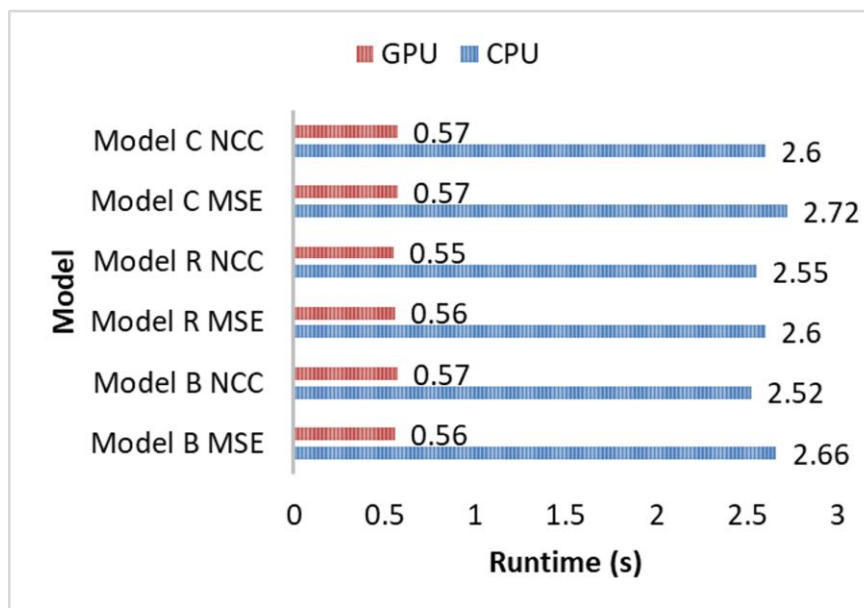
Case06	3.25(0.63)	2.75(0.93)	1.70(0.57)	<b>0.85(0.32)</b>	<b>0.85(0.29)</b>	0.91(0.32)	0.89(0.30)
Case07	1.86(1.06)	2.47(0.90)	2.18(0.66)	0.97(0.41)	<b>0.73(0.32)</b>	1.24(0.56)	0.81(0.37)
<b>Case08</b>	2.65(0.86)	3.14(0.91)	2.10(0.78)	1.29(0.49)	1.28(0.43)	<b>1.21(0.45)</b>	1.60(0.60)
Case12	19.71(0.72)	3.02(1.14)	1.85(0.78)	0.91(0.30)	<b>0.82(0.29)</b>	0.96(0.30)	0.83(0.35)
Case13	4.56(1.29)	2.13(0.65)	1.64(0.48)	0.93(0.39)	<b>0.85(0.30)</b>	1.01(0.40)	0.89(0.34)
<b>Case14</b>	3.02(0.61)	2.63(0.88)	2.55(0.87)	1.11(0.42)	<b>1.05(0.39)</b>	<b>1.05(0.39)</b>	1.17(0.36)
Case15	3.23(1.28)	2.70(0.76)	2.09(0.61)	0.95(0.51)	<b>0.85(0.45)</b>	1.30(0.55)	1.00(0.46)
Case16	3.39(0.83)	2.32(0.72)	1.66(0.42)	<b>0.68(0.20)</b>	0.71(0.24)	0.90(0.34)	0.74(0.24)
Case17	6.37(0.75)	2.48(1.07)	1.63(0.80)	0.73(0.27)	<b>0.69(0.29)</b>	1.02(0.42)	0.73(0.25)
Case18	3.57(0.93)	2.27(0.90)	1.32(0.40)	<b>0.68(0.25)</b>	<b>0.69(0.28)</b>	0.81(0.31)	0.72(0.30)
Case19	3.29(1.25)	1.93(1.09)	1.65(0.45)	0.78(0.39)	<b>0.64(0.26)</b>	0.80(0.38)	0.68(0.29)
Case21	4.56(0.71)	2.92(0.90)	1.84(0.62)	0.82(0.28)	<b>0.78(0.20)</b>	1.05(0.47)	0.82(0.27)
Case23	7.02(1.02)	1.75(0.95)	1.68(0.58)	<b>0.65(0.21)</b>	0.70(0.26)	0.71(0.26)	0.70(0.24)
Case24	1.09(0.40)	1.87(0.95)	1.68(0.46)	0.68(0.26)	<b>0.60(0.28)</b>	0.71(0.26)	0.64(0.29)
<b>Case25</b>	10.06(2.27)	1.52(0.72)	1.41(0.53)	1.11(0.43)	<b>0.94(0.45)</b>	<b>0.94(0.45)</b>	1.60(0.54)
Case26	2.82(0.81)	3.58(1.31)	2.68(0.90)	0.77(0.29)	0.84(0.27)	0.97(0.42)	<b>0.83(0.31)</b>
Case27	5.77(0.66)	2.05(0.91)	1.63(0.74)	1.00(0.38)	<b>0.89(0.33)</b>	1.15(0.42)	1.00(0.35)
mTRE	5.35(4.19)	2.50(0.66)	1.85(0.40)	0.91(0.19)	<b>0.84(0.16)</b>	1.00(0.16)	0.97(0.32)

### 5.5.3 Processing Time Analysis

Fig. 5.4 shows the computation time for the training stage (Fig. 5.4 (a)) and the test stage (Fig. 5.4 (b)). The most remarkable result is that the first configuration, *Model B* and *Model B NCC*, has the lowest training time of 45 and 55 minutes, respectively, because they are trained on the BITE dataset containing only 14 patients. On the other hand, the largest training times of 172 and 180 minutes are obtained by the third setup, *Model C* and *Model C NCC*, which use a total of 36 patients from both datasets. It becomes notable that incorporating the NCC as the similarity measurement leads



(a)



(b)

Figure 5.4: Processing time analysis for the proposed approaches on two different multiple sites datasets. (a) Training time in minutes; (b) Test time in seconds using CPU (*in Blue*) and GPU (*in Red*).

to increase the training time by 5 to 26% over the MSE versions, but the test time remains approximately constant. Overall, the three approaches provide similar processing test times of approximately 0.5 seconds on the GPU and 2.6 seconds on the CPU which provides additional support for using iRegNet in time-critical image-guided interventions

### 5.5.4 Comparison with Other MRI-iUS Studies

Fig. 5.5 shows the initial and final landmarks errors for the proposed iRegNet methods and methodologies found in the literature for MRI-iUS registration, performed on the BITE database. To assess the proposed methods, a comparison was made against other MRI-iUS registration algorithms proposed for brain shift compensation: LC2 (Wein et al., 2013), SSC (Mattias Paul Heinrich et al., 2013), SeSaMI (Rivaz et al., 2014), CoCoMI (Rivaz et al., 2014), RaPTOR (Rivaz et al., 2015), miLBP (Jiang et al., 2016), Laplacian Comm (Zimmer et al., 2019), cDRAMMS (I. Machado et al., 2019), and Arena (Masoumi et al., 2019). The results obtained indicate that the methods, highlighted in orange, outperform other evaluated competing techniques. In particular, the configuration *Model C NCC* ranked first for the BITE with an *mTRE* ( $1.47 \pm 0.61$  mm) with a 0.61 mm margin smaller than the best-performing method cDRAMMS (I. Machado et al., 2019).

Additional comparison of the registration methods against other approaches using pre-operative MRI and pre-resection US images from the RESECT database are presented in Fig. 5.6. iRegNet methods are compared to conventional studies: LC2 (Wein, 2018), SSC (Mattias P. Heinrich, 2018), NiftyReg (Drobny et al., 2018), cDRAMMS (Inês Machado et al., 2018), MedICAL (Shams et al., 2018), Structural Skeleton (Hong & Park, 2018), ARENA (Masoumi et al., 2019) as well as learning studies: FAX (Zhong et al., 2018), CNN + STN (Sun & Zhang, 2018). It is important to note that the LC2 algorithms applied to BITE utilized different configurations than the one applied to RESECT. In (Wein et al., 2013), the algorithm aligns 2D US to 3D MRI volumes initialized with a rigid registration and smaller patch sizes of 2 to 24, while in (Wein, 2018), a non-linear optimization algorithm was initialized with a translation before the rigid registration and the patch size is larger with 73 voxels. Similarly, the SSC methods in (Mattias Paul Heinrich et al., 2013) and (Mattias P. Heinrich, 2018) are separate from each other. In (Mattias P. Heinrich, 2018), the authors set the parameters for the discrete optimization with a complex 107 degrees of freedom. On the other hand, the graph is simplified in (Mattias Paul Heinrich et al., 2013) and contains no loops leading to a faster and smoother transformation.

As can be seen from Fig. 5.6, *Model R* variants rank first on the RESECT dataset with average *mTRE* ( $0.84 \pm 0.16$  to  $0.91 \pm 0.19$  mm) followed by the learning-based method FAX with *mTRE* of ( $1.21 \pm 0.55$  mm). Although team FAX reported comparable results, this method failed to obtain similar results on the test dataset of the CuRIOUS challenge (Xiao et al., 2020), which presumably is due to an overfit over the training images. In contrast, *Model B* and *Model B NCC* enhanced the initial *mTRE* of RESECT by 2.85 and 3.50 mm in turn, however, they failed to provide competitive outcomes. This implies that the accuracy of deep learning approaches does seem to depend on the size of the available training data.

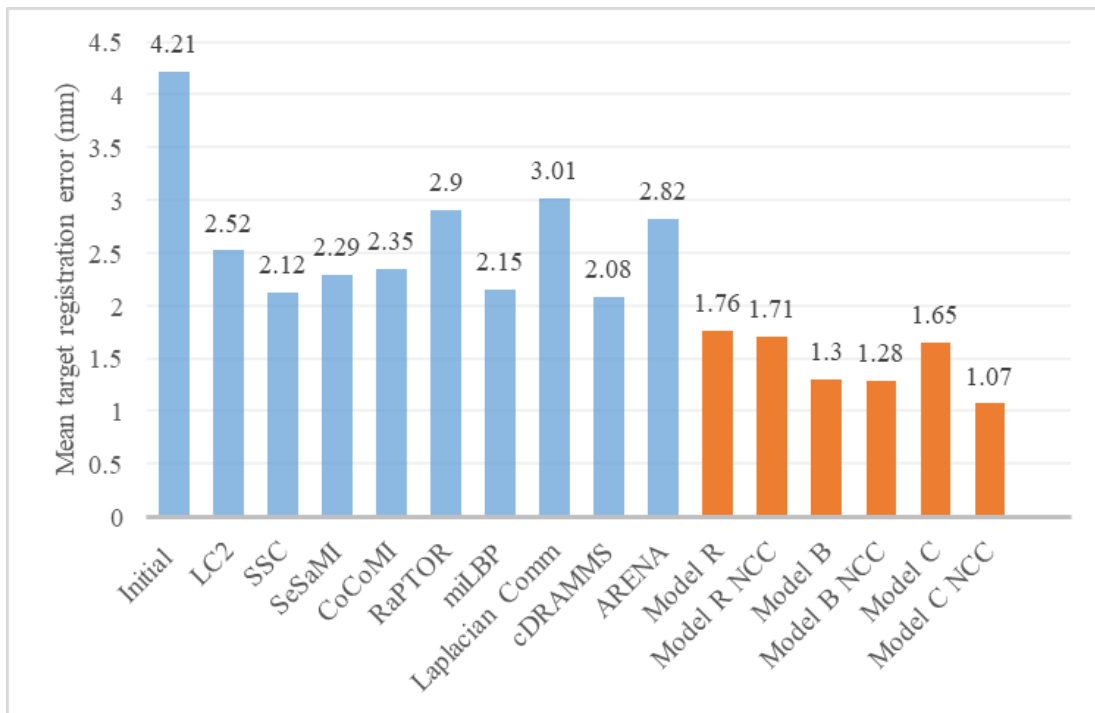


Figure 5.5: A comparison of the registration error ( $mTRE$ ) for the proposed iRegNet methods and the state-of-the-art methods on the BITE dataset.

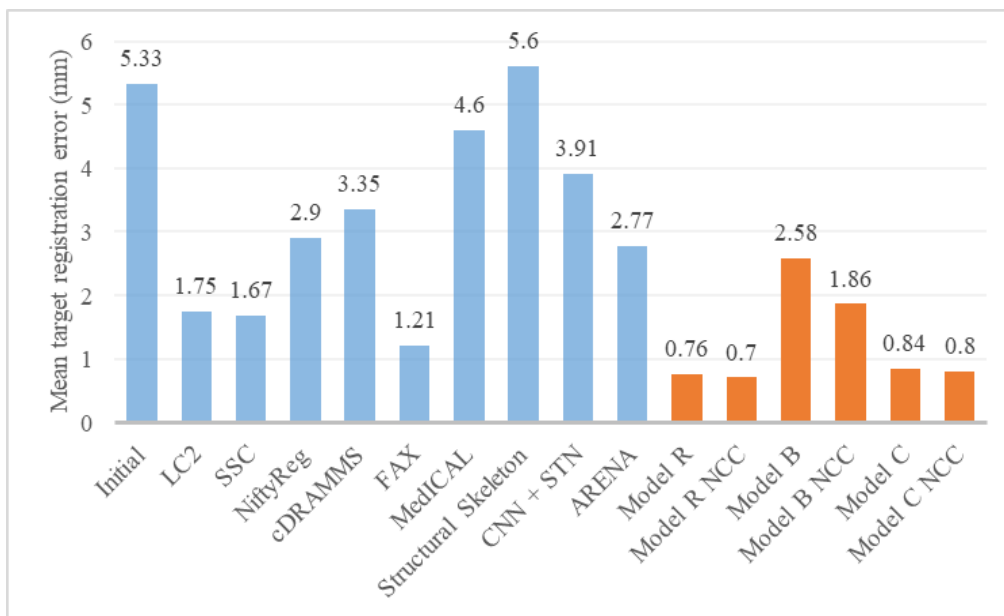


Figure 5.6: A comparison of the registration error ( $mTRE$ ) for the proposed iRegNet methods and the state-of-the-art methods on the RESECT dataset.

### 5.5.5 Qualitative Analysis by Neurosurgeons

Two experienced neurosurgeons (Z. A.) and (M. S.) with eight and five years of clinical practice, respectively, visually inspected the MRI-iUS registration results individually and rated the results based on the tumor boundaries and other brain structures, such as sulci and falx. Table 5.5 summarizes their qualitative analysis. It is worth mentioning that only 11% of alignments were classified as “limited” (no improvements over the initial registration), and 8% or 17%, respectively, of aligned results, were “fair” (minor improvements), 36% or 39%, respectively, of alignments, were “good” (major improvements), and 44% or 33%, respectively, of alignments, were “excellent” (little or no visible misalignment).

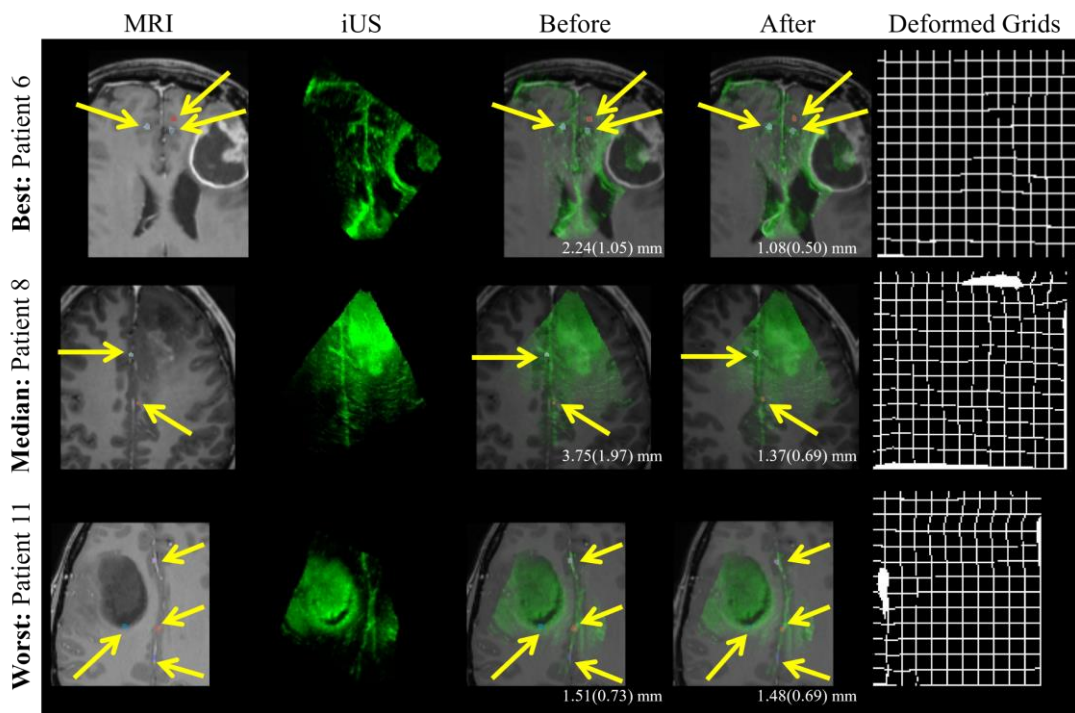
Figure 5.7 displays the results of aligning MRI to iUS using the best method, namely *Model C NCC*, for six different patients: BITE (cases #6, #8, and #11) and RESECT (cases #23, #14, and #21), correspondingly. The cases are selected as best, median, and worst according to their qualitative evaluation as given in Table 5.5. In each figure, columns provide pre-operative MRI, interventional US, initial alignment of both images before registration, and overlap of corrected MRI over iUS after applying iRegNet, correspondingly. Similar to other studies, the raters confirm that the quality of the used US images in the BITE dataset is rather limited as shown in case #8 in Fig. 5.7 (a).

The results of this study demonstrated that iRegNet, *Model C NCC*, yields better registration results (last row) than the initial misaligned MRI and iUS pairs (third row). Although widely accepted, it suffers from a comparable larger training time of 180 minutes, however, this is only apparent in the training stage while obtaining a very competitive inference time of about half-second on GPU-based implementation. Another limitation of this implementation is that this improvement may be unclear in a few cases, such as Case12 from the BITE dataset. A popular explanation is that the initial alignment has a small brain shift that makes it difficult to observe. Overall, this analysis yields an overview of the potential clinical applicability of the method regarding the accuracy and quality of the registration outputs.

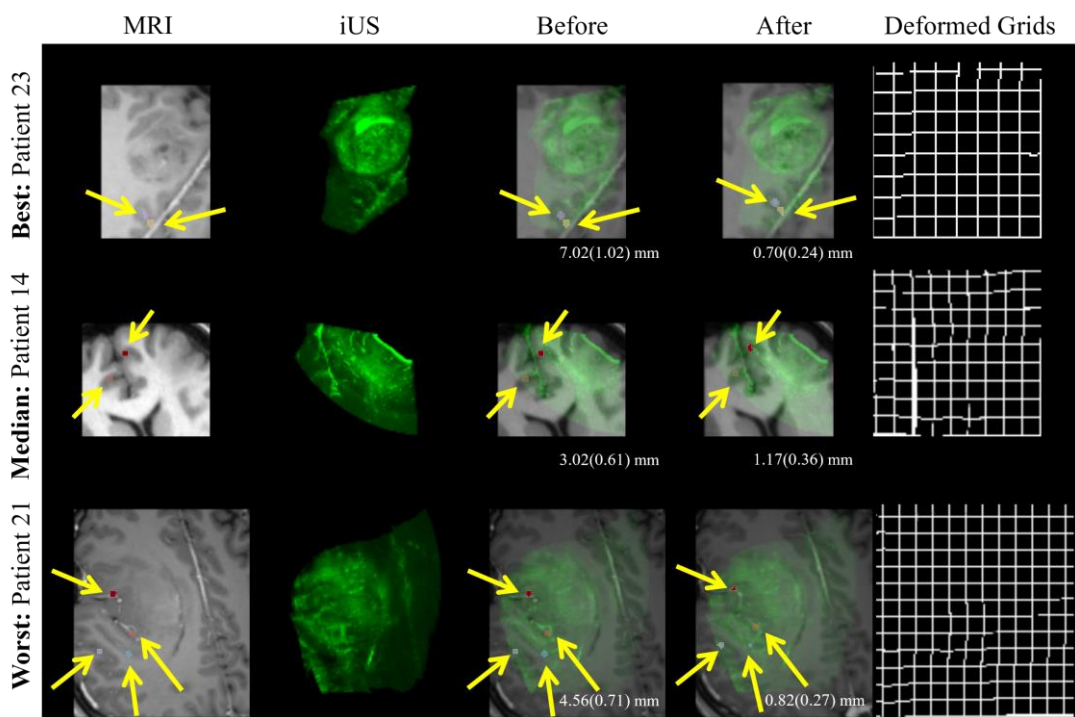
Table 5.5: Statistical assessment of the MRI-iUS alignment of 36 cases by two expert neurosurgeons.

Neurosurgeon	Assessment			
	Limited	Fair	Good	Excellent
#1	4	3	13	16
#2	4	6	14	12
Total	8	9	27	28





(a)



(b)

Figure 5.7: Alignment of pre-operative MRI (*gray color mode*) to iUS (*green color mode*) in six different cases from (a) BITE (cases #6, #8, and #11) and (b) RESECT (cases #23, #14, and #21). Column designations: pre-operative FLAIR MRI, intra-operative US, initial misaligned MRI over iUS before registration, and the final aligned MRI over iUS after registration, and the deformed grids. Yellow arrows indicate expert-labeled landmarks while *mTRE* values are shown (*bottom right*).

### 5.5.6 Self-supervised Results

To explore the MRI modality which achieves the best performance for the task of longitudinal registration, an ablation study has been carried out. The BRAINSFit toolkit was utilized to perform affine alignments on moving and fixed MRI volumes. As listed in Table 5.6, T1ce has obtained the overall best results on the validation dataset in terms of the mean and median MAE scores while FLAIR achieved the best robustness. Therefore, in the experiments, the T1ce volumes from each patient were only used. Theoretically, using multiple modalities could increase the accuracy of the image registration, and this would be further investigated in future work.

Table 5.6: The ablation study of MRI modalities on the BraTS-Reg 2022 validation cases. Bold highlights the best scores.

Modality	MAE <sub>median</sub>	MAE <sub>mean</sub>	Robustness
Initial	8.20	8.65	-
T1	4.74	5.65	0.66
T1ce	<b>4.35</b>	<b>5.23</b>	0.62
T2	4.85	5.60	0.63
FLAIR	4.64	5.40	<b>0.67</b>

Figure 5.8 shows example results from three patients, where the registration of post- to pre-operative MRI scans is achieved using the self-supervised iRegNet method and the comparing baseline method. From the visual results, it can be seen that warped MRI scans are significantly improved after applying iRegNet. Note that, Fig. 5.8 (c) shows the FLAIR scans for the follow-up MRI images, only for visualization purposes, to better depict the surgically imposed cavities of these illustrated examples. All the applied registration methods use only the T1ce modality as discussed in Section 3.2.

Moreover, Table 5.7 reports the registration performance of the proposed method as well as the baseline on the BraTS-Reg challenge validation database. The baseline denotes the BRAINSFIT affine transformation between the full-resolution images of pre-operative and follow-up MRI. Compared with the affine method, the proposed self-supervised method effectively improves registration performance. It is notable that the average runtime of the proposed method is 1 second and does not require any manual interaction or supervision. Besides, only one sequence (T1ce) is required in this application.

The statistics of the paired landmark errors before and after the registration are displayed in Fig. 5.9. For the training database, the model reduced the initial mean MAE (computed by the evaluation platform) from  $8.20 \pm 7.62$  mm to  $3.51 \pm 3.50$  mm. Similarly, an MAE of  $2.93 \pm 1.63$  mm was achieved on the validation database which has an initial  $7.80 \pm 5.62$  mm. This result high-

lights that the method delivers significantly better results than both initial alignment and affine registration.

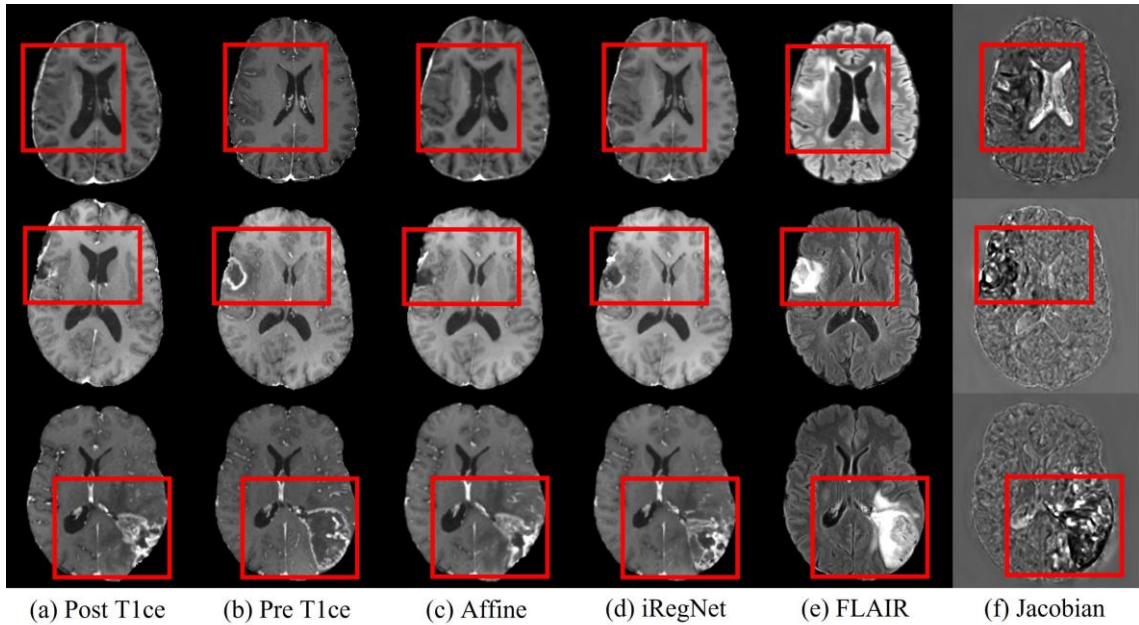


Figure 5.8: Example registration results from three validation cases (patients 141, 148, and 152). From left to right: (a) and (b) the post- and pre-operative MRI T1ce, (c) the follow-up to pre-operative affine registration of BRAINSFit, (d) the follow-up to pre-operative deformable registration of the iRegNet, (e) the pre-operative FLAIR scans, only for visualization purposes, and (f) determinant of the Jacobian of the displacement field are shown, respectively. The red box highlights regions of major differences.

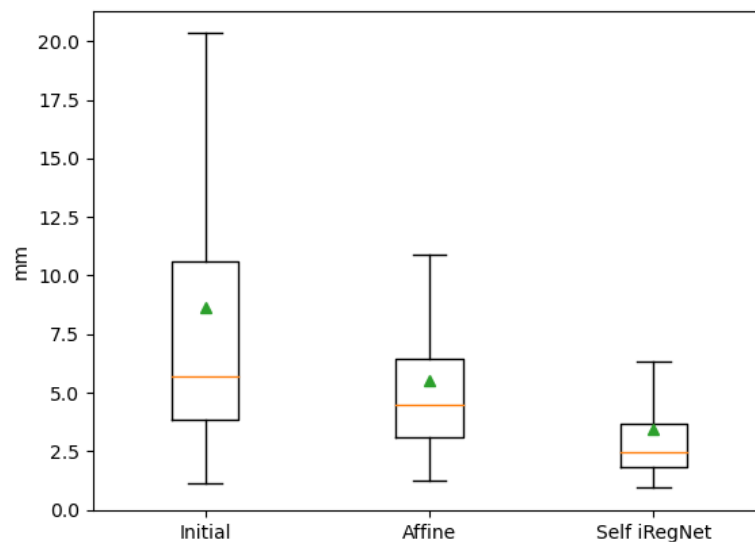


Figure 5.9: Boxplots of the mean landmark errors. For each method, the landmark errors are computed against the fixed landmarks of the BraTS-Reg dataset. From left to right, mean absolute registration errors are shown for the initial dataset, affine, and the enhanced iRegNet, respectively. On each box, the red line is the median and the green triangle is the mean.

Table 5.7: Quantitative results of the proposed method and the baseline affine method on the BraTS-Reg challenge validation set. MAE denotes the average of median absolute error between the predicted coordinates and the ground truth coordinates, whereas Robustness represents the successful rate of measuring how many landmarks have improved MAE after the registration.

Case	Initial	Affine		Enhanced iRegNet	
	MAE	MAE	Robustness	MAE	Robustness
BraTSReg_141	13.50	3.64	1.00	2.18	1.00
BraTSReg_142	14.00	5.98	0.88	7.18	0.75
BraTSReg_143	16.00	8.85	0.88	4.89	1.00
BraTSReg_144	15.00	9.44	0.88	5.64	1.00
BraTSReg_145	17.00	5.36	1.00	4.71	1.00
BraTSReg_146	17.00	7.13	1.00	2.62	1.00
BraTSReg_147	1.50	2.50	0.00	2.53	0.50
BraTSReg_148	3.50	3.06	0.30	2.61	0.75
BraTSReg_149	9.00	2.18	1.00	1.38	1.00
BraTSReg_150	4.00	4.00	0.11	2.20	0.74
BraTSReg_151	3.00	2.00	0.45	1.47	0.75
BraTSReg_152	5.00	2.00	0.95	1.61	0.95
BraTSReg_153	2.00	2.00	0.33	1.68	0.75
BraTSReg_154	2.00	2.10	0.15	1.83	0.55
BraTSReg_155	2.00	2.63	0.21	2.10	0.53
BraTSReg_156	7.00	3.30	1.00	1.62	1.00
BraTSReg_157	10.00	6.52	0.90	4.58	1.00
BraTSReg_158	4.50	3.75	0.40	1.65	1.00
BraTSReg_159	6.00	8.00	0.36	3.58	1.00
BraTSReg_160	4.00	2.50	0.70	2.47	0.60
Mean	7.80	4.35	0.62	2.93	0.84
StdDev	5.62	2.46	0.36	1.63	0.19
Median	5.50	3.47	0.79	2.33	0.97

## 5.6 Summary

iRegNet was presented as an automated, fast, and robust deformable approach for pre-operative MRI to pre-resection iUS registration for compensating brain shift phenomenon. In six experiments, the proposed method has been successfully tested and evaluated on 36 cases from two multi-location datasets, validating the registration performance qualitatively and quantitatively. Notably, iRegNet achieved considerable performance and computational efficiency even with untrained cases, demonstrating the generality of the proposed method. Compared with other registration methods, iRegNet achieved the best accuracy results in terms of the mean TRE with values of  $(1.47 \pm 0.61$  and  $0.84 \pm 0.16$  mm) for the utilized BITE and RESECT datasets, respectively, as illustrated in Tables 5.3 and 5.4. Additional qualitative results indicate that the registered MRI-iUS pairs have a significant improvement over their initial alignment. Also, the proposed iRegNet achieved significant performance on multi-center data and is, therefore, a potentially promising automatic registration algorithm for use with IGN systems. Moreover, iRegNet is flexible, modality-, and anatomy-invariant, and therefore could be used in a wide range of medical image analysis and processing surgical procedures.

Furthermore, a patient-specific registration framework is proposed based on iRegNet, which aligns pre-operative and post-recurrence MRI T1ce sequences. The enhanced iRegNet framework uses deep unsupervised learning for deformable image registration driven by the regularization hyperparameter. The proposed method is evaluated on the BraTS-Reg challenge dataset brain MR images comprising 140, 20, and 40 divided into training, validation, and testing cohorts. Therefore, the results of the proposed registration method are promising and can be applied for clinical use during future work.



# 6 Explainability of Deep Neural Networks

---

## Contents

<b>6.1</b>	<b>Introduction</b>	<b>78</b>
<b>6.2</b>	<b>Related Work</b>	<b>78</b>
<b>6.3</b>	<b>Methods</b>	<b>80</b>
<b>6.3.1</b>	<b>NeuroXAI</b>	<b>80</b>
<b>6.3.1.1</b>	<b>Vanilla Gradient</b>	<b>81</b>
<b>6.3.1.2</b>	<b>Guided Backpropagation</b>	<b>81</b>
<b>6.3.1.3</b>	<b>Integrated Gradients</b>	<b>82</b>
<b>6.3.1.4</b>	<b>Guided Integrated Gradients</b>	<b>82</b>
<b>6.3.1.5</b>	<b>SmoothGrad</b>	<b>82</b>
<b>6.3.1.6</b>	<b>Grad-CAM</b>	<b>82</b>
<b>6.3.1.7</b>	<b>Guided Grad-CAM</b>	<b>83</b>
<b>6.3.2</b>	<b>TransXAI</b>	<b>84</b>
<b>6.3.2.1</b>	<b>CNN-Transformer Hybrid Architecture</b>	<b>85</b>
<b>6.3.2.2</b>	<b>Explainable CNN Generator</b>	<b>86</b>
<b>6.4</b>	<b>Experiments</b>	<b>87</b>
<b>6.4.1</b>	<b>Data</b>	<b>87</b>
<b>6.4.2</b>	<b>Experimental Setup</b>	<b>88</b>
<b>6.5</b>	<b>Results and Discussion</b>	<b>89</b>
<b>6.5.1</b>	<b>Showcase I: Application to Classification</b>	<b>89</b>
<b>6.5.2</b>	<b>Showcase II: Application to Segmentation</b>	<b>92</b>
<b>6.5.2.1</b>	<b>Comparing Different XAI Methods</b>	<b>93</b>
<b>6.5.2.2</b>	<b>Role of MRI in Tumor Detection</b>	<b>96</b>
<b>6.5.2.3</b>	<b>GCAM for CNN Layers</b>	<b>97</b>
<b>6.5.2.4</b>	<b>CNN Node Failure Detection</b>	<b>99</b>
<b>6.6</b>	<b>Clinical Relevance</b>	<b>101</b>
<b>6.7</b>	<b>Summary</b>	<b>102</b>

Within this chapter, the third module of the proposed IGN pipeline (see Figure 1.2 “DeepIGN”) called *Interpreting the developed DNNs using Explainable AI* is described. The remainder of this chapter is organized as follows. First, introduction about making deep learning models transparent is given in Section 6.1. Then, the state-of-the-art explanation methods are given in Section 6.2. Section 6.3 gives an overview of the proposed explainable machine learning system, the hybrid CNN Transformer network for brain tumor segmentation, and relevant interpretability methodologies. The experimental setup and the utilized data are described in Section 6.4. After that, Section 6.5 describes and discusses the results of the extensive experiments. The clinical relevance is clarified in Section 6.6. Finally, the concluding remarks and prospects of this study are given in Section 6.6. Parts of this chapter have previously been published in (R. A. Zeineldin et al., 2022).

## 6.1 Introduction

Recent developments in DNNs have demonstrated their effectiveness for processing and analyzing medical images, including those associated with brain tumor segmentation [10-13], image registration [14, 15], and image classification [16-18]. However, the introduction of deep learning techniques in the clinical environment is still limited due to some restrictions (G. Yang et al., 2022). The most significant one is that deep learning strategies consider only the input images and the output results, without any transparency of the underlying information flow in the network's internal layers. In sensitive applications such as brain imaging applications, it is crucial to understand the reason behind the network prediction to ensure that the model provides the correct estimation.

Accordingly, XAI has gained a substantial interest to explore the “black box” deep learning networks in the medical field (Gulum, Trombley, & Kantardzic, 2021; G. Yang et al., 2022). XAI methods allow researchers, developers, and end-users to obtain transparent deep learning models that can describe their decisions to humans in an understandable manner. For medical end-users, the demand for explainability is increasing to create their trust in deep learning techniques and to encourage them to utilize these systems for assisting clinical procedures. Moreover, the European Union data protection law, titled General Data Protection Regulation (GDPR), imposes the explanation as a requirement for automated learning systems before being used with patients clinically (Temme, 2017).

## 6.2 Related Work

Generally, XAI techniques in medical imaging can be grouped into perturbation-based or gradient-based approaches. Perturbation-based methods investigate the network by changing the input features and measuring the impact on the output estimations by a forward training of the model. Some examples include LIME (Ribeiro, Singh, & Guestrin, 2016), Shapely additive explanations (SHAP) (Ribeiro et al., 2016), Deconvolution (Zeiler & Fergus, 2014), and Occlusion (Zeiler & Fergus, 2014). Gradient-based XAI methods have been widely adopted to provide feature attribution maps by calculating the partial derivative of the output predictions through every layer of the neural network with respect to (w.r.t) the input images. These techniques have the advantage of being post-hoc, meaning that they are applied after the training phase of the deep learning model avoiding the accuracy vs explainability trade-off. In addition, they are usually fast compared with perturbation approaches since their runtime does not depend on the number of input features. A number of publications have been reported for back-propagating approaches such as Vanilla Gradient (VG) (Simonyan, Vedaldi, & Zisserman, 2014), Guided Backpropagation (GBP) (Springenberg et al., 2015), Integrated Gradients (IG) (Sundararajan, Taly, & Yan, 2017), and Guided Integrated Gradients (GIG) (Kapishnikov et al., 2021), SmoothGrad (Smilkov, Thorat, Kim, Viégas, & Wattenberg, 2017), Grad-CAM (GCAM) (Selvaraju et al., 2017), and Guided Grad-CAM (GGCAM) (Selvaraju et al., 2017). Several XAI methods have been previously proposed for natural image tasks, while little attention has been paid to explaining brain imaging applications (Gulum et al., 2021). For brain cancer classification, Windisch et al. (Windisch et al., 2020) applied 2D GCAM to generate



heatmaps indicating which areas of the input MRI made the classifier decide on the category of the existence of a brain tumor. Similarly, 2D GCAM was used in (Esmaeili, Vettukattil, Banitalebi, Krogh, & Geitung, 2021) to evaluate the performance of three deep learning models in brain tumor classification. The key limitation of these studies is that experiments were concluded on 2D MRI slices without investigating the model on 3D medical applications.

In previous studies, several interpretability approaches have been proposed to explain the behavior of machine learning models in medical applications such as COVID-19 diagnosis (Singh, Pandey, & Babu, 2021; Y. H. Wu et al., 2021), retinal imaging (Maloca et al., 2021; Sayres et al., 2019), skin cancer (Datta, Shaikh, Srihari, & Gao, 2021; Shorfuzzaman, 2021; Young, Booth, Simpson, Dutton, & Shrapnel, 2019), colonoscopy (Wickstrom, Kampffmeyer, & Jenssen, 2020), and brain imaging (Eitel & Ritter, 2019; Natekar, Kori, & Krishnamurthi, 2020; Pereira, Meier, Alves, Reyes, & Silva, 2018; S. Pereira et al., 2018; Saleem, Shahid, & Raza, 2021; B. Zhou, Khosla, Lapedriza, Oliva, & Torralba, 2016).

Also, some research works have been conducted to generate explainable results for brain tumor segmentation networks. In (S. Pereira et al., 2018), Pereira et al. employed a joint Restricted Boltzmann Machine system (RBM) and a Random Forest (RF) classifier to enhance the interpretability of a machine learning system. Inspired by (Ribeiro et al., 2016), they provided two levels of interpretation, i.e. local and global, allowing for an evaluation of the extracted task-specific features and the voxel-level predictions, respectively. A key limitation of their mutual RBM-RF feature selection strategy is the requirement for random perturbation of input feature vectors which can be computationally expensive for medical imaging tasks and, therefore, very slow.

Natekar et al. generated visual explanations of three DNNs for brain tumor segmentation (Natekar et al., 2020). They applied 2D GCAM (Selvaraju et al., 2017) to explain the contribution of the internal layers of those segmentation networks helping to understand “why” DNN achieved quantitatively high accurate tumor segmentations. The experiments indicated that DNN follows a human-like hierarchical approach for localizing different parts of the brain tumor. However, the three explained models are implemented in 2D, limiting the applicability of this approach to 3D explainable medical applications.

In (Saleem et al., 2021), a method has been developed for 3D visual explanations towards explaining the “black-box” nature of CNNs. This method extended Class Activation Mapping (CAM) (B. Zhou et al., 2016) to extract 3D explanations to interpret a brain tumor segmentation network. Moreover, they investigated how the input MRI modality perturbation affects the prediction strategy of different brain tumor sub-regions. However, altering the deep model architecture is a requirement to make it interpretable, which trades off the model complexity and accuracy of the network. Furthermore, CAM-based approaches are restricted to a certain type of CNNs without including any multi-modal input or fully connected layers CNNs.

Li et al. developed an explainable ensemble Gaussian kernel (XEGK) to substitute for CNN in feature extraction, in which they used an explainable Gaussian kernel to capture characteristic features from relevant regions of the input (J. Li et al., 2021). They applied their method to mono-channel input and multi-channel inputs by using by leveraging the Gaussian mixture

model (GMM) and fusion of multiple GMMs, respectively. To interpret the experimental results, they used SHAP (Lundberg & Lee, 2017) to reflect the features' importance. SHAP is a perturbation-based approach from the coalitional game theory which assigns a feature importance value for each class prediction. It is therefore inefficient in critical medical applications since the network must be run for the number of samples  $\times$  the number of features times.

## 6.3 Methods

### 6.3.1 NeuroXAI

While the state-of-the-art explanation methods were primarily proposed for interpreting deep image classification, the proposed framework provides an adaptive approach to medical image segmentation as well. Moreover, NeuroXAI converts the segmentation task into a multi-label classification task. This is achieved through global average pooling for each class on the output prediction layer. Therefore, the NeuroXAI offers state-of-the-art XAI methods for the classification and segmentation of both 2D and 3D medical image data. The overall pipeline of NeuroXAI is depicted in Fig. 6.1. It is a two-step approach, a neural network for tumor segmentation, and an explainability generator. The first step is to perform any brain imaging task data using CNN-based architecture. The second step is a justification generator that is employed to provide 3D visual feature explanations. The following subsections describe the detailed structure of the CNN model and the justification generator.

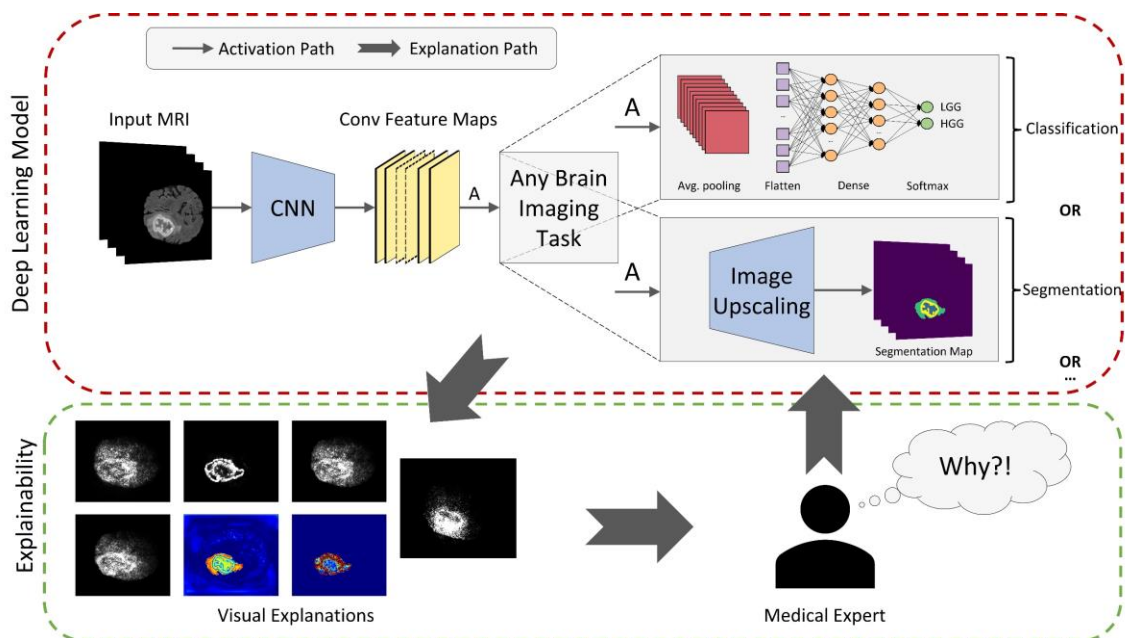


Figure 6.1: Pipeline of the proposed NeuroXAI framework.

### 6.3.1.1 Vanilla Gradient

Vanilla Gradient (VG) (Simonyan et al., 2014) is the simplest form of visualizing regions of the image that contributes most to the classification output of the neural network. This computes the saliency map by making a single backward pass of the activation of the output class after a forward pass over the network, which can be defined as computing the VG of the output activation w.r.t the input image. Let  $P_c(x^I)$  be the prediction of class  $c$ , computed by the classification layer of the CNN for an input image  $x^I$ . The objective of Vanilla Gradient is to find the L2-regularised image, which has the maximum  $P_c$ , while  $\lambda$  is the regularization term:

$$VG = \operatorname{argmax}_c P_c(x^I) - \lambda \|x^I\|_2^2 \quad (6.1)$$

### 6.3.1.2 Guided Backpropagation

GBP is used to highlight features learned by CNNs that contribute most to the predicted result (Springenberg et al., 2015). GBP is a new variant of the deconvolution approach introduced in (Zeiler & Fergus, 2014), especially for visualizing the concepts learned by higher network layers toward CNN analysis. During backpropagation, GBP reverses the data flow of a CNN, making a backward pass of the network from the high-level output layer to the input image. By using the ReLU activation function, GBP masks out the values with at least one negative value which acts as an additional guidance signal to typical backpropagation. Figure 6.2 demonstrates the computation of guided backpropagation in a sample CNN network.

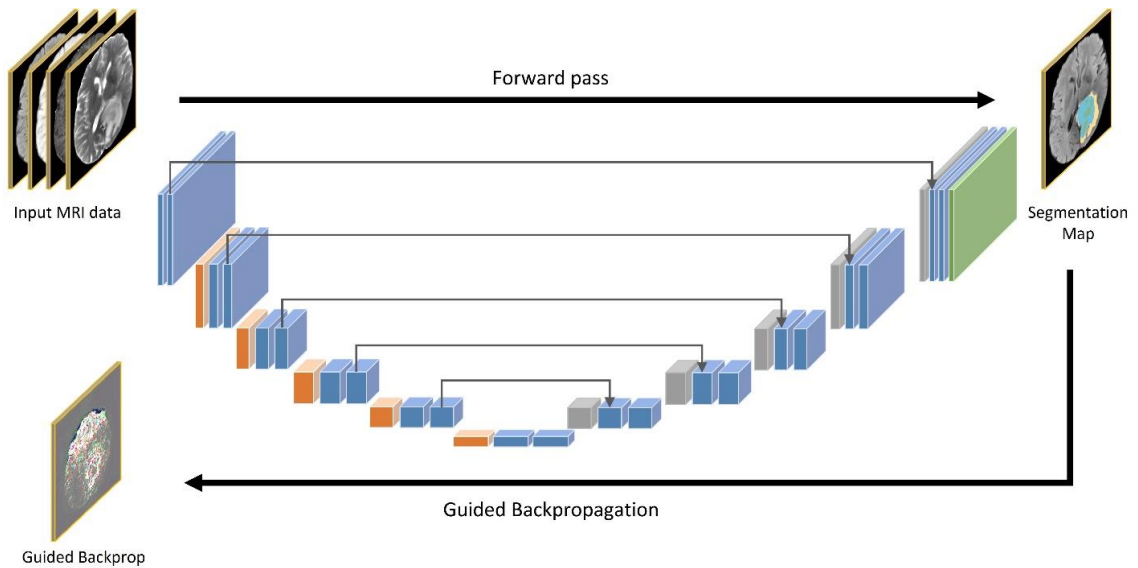


Figure 6.2: Applying Guided Backpropagation to a sample brain tumor segmentation CNN model.

Let  $f$  represent the input feature map calculated by a layer  $\ell$  from a CNN,  $R$  denotes the resultant reconstructed image, and then the GBP can be calculated as follows (Springenberg et al., 2015):

$$R_i^\ell = (f_i^\ell > 0) \cdot (R_i^{\ell+1} > 0) \cdot R_i^{\ell+1} \quad (6.2)$$

$$R_i^{\ell+1} = \frac{\partial f_i^L}{\partial f_i^{\ell+1}} \quad (6.3)$$

Where  $L$  is the total number of layers in the CNN.

### 6.3.1.3 Integrated Gradients

Sundararajan et al. (Sundararajan et al., 2017) introduced IG to mitigate the saturation problem of gradient-based methods. Let a function  $G: \mathbb{R}^n \rightarrow [0, 1]$  denote a DNN which has  $X^I = \gamma (\alpha = 1) \in \mathbb{R}^n$  as the input image, while  $X^B = \gamma (\alpha = 0) \in \mathbb{R}^n$  represents the baseline. The baseline is simply a black image with all values set to zeros. The IG can be computed by accumulating the gradients at all points on the straight-line path from the baseline  $X^B$  to the input image  $X^I$ :

$$IG_i(x) = \int_{\alpha=0}^1 \frac{\partial F(\gamma(\alpha))}{\partial \gamma_i(\alpha)} \frac{\partial \gamma_i(\alpha)}{\partial \alpha} d\alpha \quad (6.4)$$

Here,  $i$  is the feature for the input image whereas  $\alpha$  represents the interpolation constant to perturb image features.

### 6.3.1.4 Guided Integrated Gradients

Kapishnikov et al (Kapishnikov et al., 2021) proposed GIG as an adaption of the attribution path based on the input image, baseline, and the deep model to be explained. Similar to IG, the GIG calculates the gradients on the path which starts at the baseline ( $X^B$ ) and ends at the input being explained ( $X^I$ ). However, the GIG path is determined at every step as opposed to the fixed direction of the IG. This means that GIG finds a subset of features ( $\mathbb{S}$ ) that have the least importance among all features toward the input image. Mathematically,

$$GIG_i(X^B, X^I, G) = \frac{\partial \gamma_i^F(\alpha)}{\partial \alpha} = \begin{cases} x_i^I - x_i^B, & \text{if } i \in \mathbb{S}, \\ 0, & \text{otherwise.} \end{cases} \quad (6.5)$$

$$\mathbb{S} = \arg \min_i(\mathcal{Y}) \quad (6.6)$$

### 6.3.1.5 SmoothGrad

Smilkov et al. (Smilkov et al., 2017) presented an improvement for the common problem of gradient-based methods. SmoothGrad (Smilkov et al., 2017) solved this problem by providing visually sharpened sensitivity maps. It computes the gradient over multiple samples surrounding the input  $X^I$ , and the average is calculated after adding Gaussian noise. More formally,

$$\overline{M}_c(X^I) = \frac{1}{N_s} \sum_1^{N_s} M_c(X^I + \mathcal{g}(0, \sigma^2)) \quad (6.7)$$

Where  $M_c(X^I)$  is the original sensitivity map,  $N_s$  is the number of samples, and  $\mathcal{g}(0, \sigma^2)$  denotes Gaussian noise with variance  $\sigma^2$ . In General,  $M_c(X^I)$  can be any gradient-based visualization method, such as explanation methods in the previous sub-sections.

### 6.3.1.6 Grad-CAM

GCAM is a generalization of the local visualization approach CAM (B. Zhou et al., 2016) for identifying discriminative features and addressing its shortcomings. Figure 6.3 shows an appli-

cation of GCAM to segmentation CNN, which can be applied without any architectural modifications while the model's output layer is differentiable with respect to its input feature neurons. By using the gradient information from the last convolutional layers of the CNN, GCAM can highlight the regions responsible for a particular class of interest.

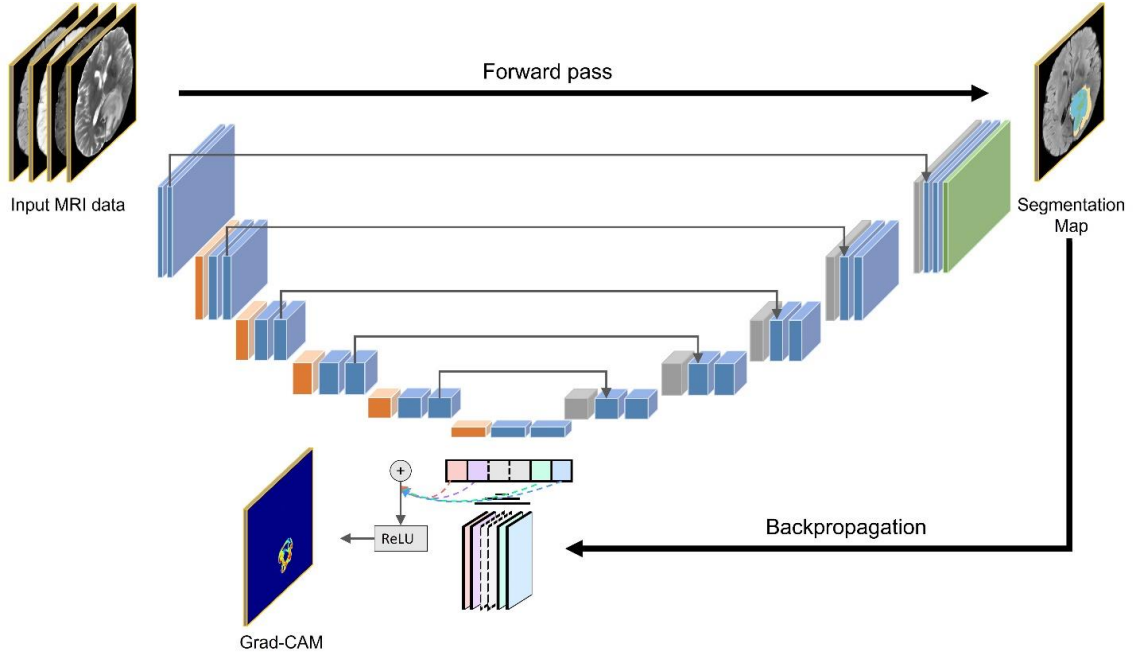


Figure 6.3: Applying Grad-CAM to a sample brain tumor segmentation CNN model.

Let us define the 3D GCAM heatmap as  $L_{GCAM}^c$  which captures the important localization feature map  $k$  for a target class  $c$  with respect to all  $N$  pixels (indexed by  $x, y, z$ ).  $L_{GCAM}^c$  is the linear combination of the forward pass activation map  $A^k$  and the backpropagated gradient  $B_k^c$  for the input activations followed by a ReLU activation function.

$$L_{GCAM}^c = ReLU(\sum_l B_l^c A^l) \quad (6.8)$$

$$B_l^c = \frac{1}{N} \sum_x \sum_y \sum_z \frac{\partial y^c}{\partial A_{x,y,z}^l} \quad (6.9)$$

### 6.3.1.7 Guided Grad-CAM

GBP localizes important features in the input image that contribute more to the network prediction with high-resolution visualization, but the results are not class-distinct. In contrast, GCAM coarse heatmaps are class-discriminative producing smoother regions due to mapping from lower-resolution convolutional layers. To combine the advantages of both methods, GBP and GCAM are combined via pixel-wise products resulting in GGCAM (Selvaraju et al., 2017). This results in generating high-resolution explanations that correspond to the specific class output estimates with respect to any region of interest. The adoption of GGCAM for brain tumor segmentation CNN is shown in Fig. 6.4.

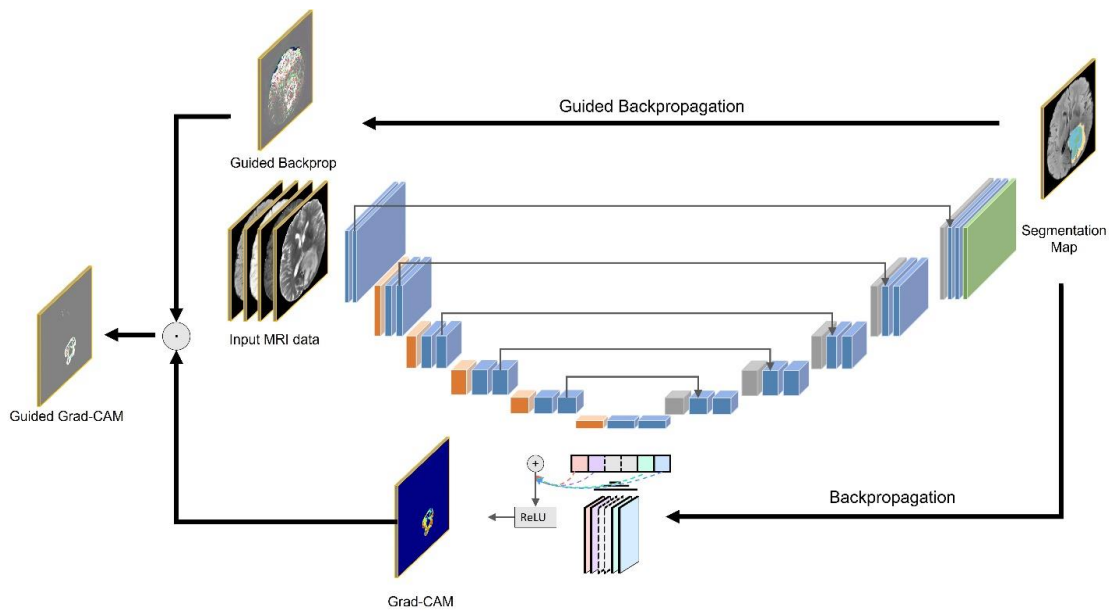


Figure 6.4: Applying Guided Grad-CAM to a sample brain tumor segmentation CNN model.

### 6.3.2 TransXAI

The overall proposed gradient-based justification hybrid CNN-Transformer architecture for explainable brain lesion segmentation is depicted in Fig. 6.5. It is a two-step approach, which

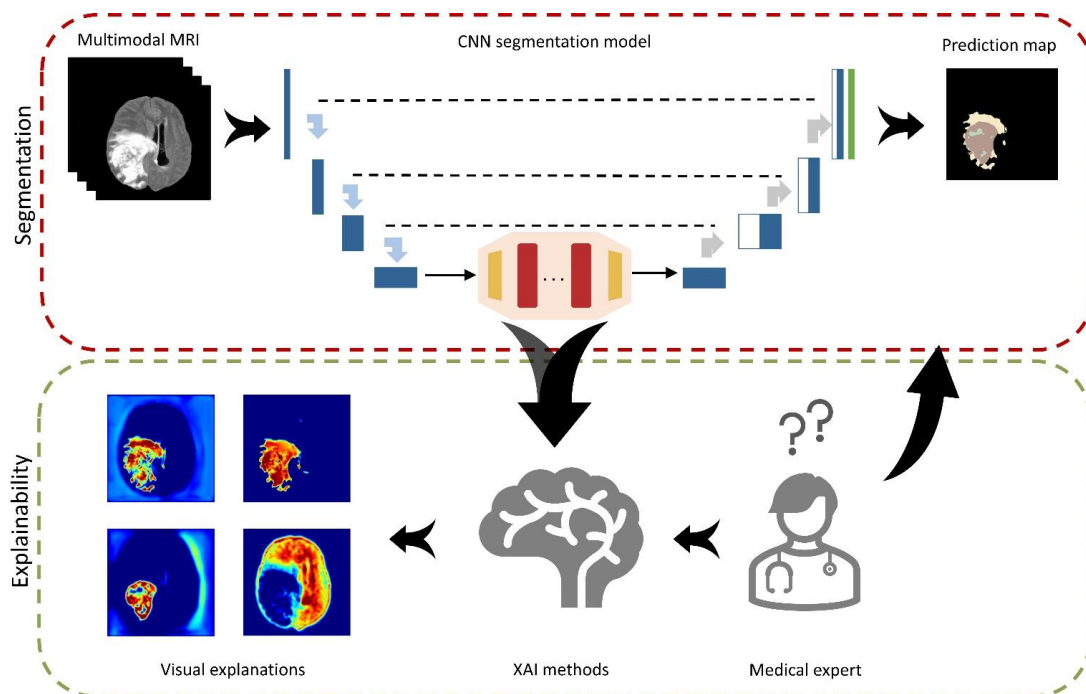


Figure 6.5: Overall proposed TransXAI pipeline for visual justification of glioma segmentation in brain MRI using a hybrid CNN-Transformer architecture.

combines a deep network for tumor segmentation, and an explainability generator. The first step is to segment the brain tumor boundaries from multimodal MRI data using a combined neural network with Transformer. The second step is a justification generator that is employed to provide 2D visual feature explanations. The following subsections describe the detailed structure of the deep model and the justification generator.

### 6.3.2.1 CNN-Transformer Hybrid Architecture

A detailed pipeline of the proposed TransXAI approach is given in Fig. 6.5. Given an input MRI volume  $x \in \mathbb{R}^{H \times W \times N_c}$  where  $H \times W$  is the spatial resolution and  $N_c$  number of channels (# of modalities), a modified 2D CNN is first utilized, based on the widely used U-shaped encoder-decoder architecture (Çiçek et al., 2016; Ronneberger et al., 2015; R. A. Zeineldin et al., 2020), as shown in Fig. 6.6, to extract high-level feature representations capturing local spatial features. The CNN-based encoder blocks contain 2D  $3 \times 3$  convolutional blocks to capture the spatial and depth information. Besides, every CNN block has a BN layer between the convolution layers and ReLU activation (Ioffe & Szegedy, 2015; Srivastava et al., 2014). For downsampling,  $2 \times 2$  max-pooling is used to gradually extract spatial feature maps  $f \in \mathbb{R}^{K \times \frac{H}{8} \times \frac{W}{8} \times \frac{N_c}{8}}$  ( $K = 24$ ), which is 1/8 of input dimensions of H and W. Then, the Transformer encoder blocks leverage to extract the long-distance dependencies through the self-attention mechanism. The decoder is composed of  $2 \times 2$  up convolutional layers that are applied to up-scale the resultant encoded feature representation into the full-resolution segmentation maps of

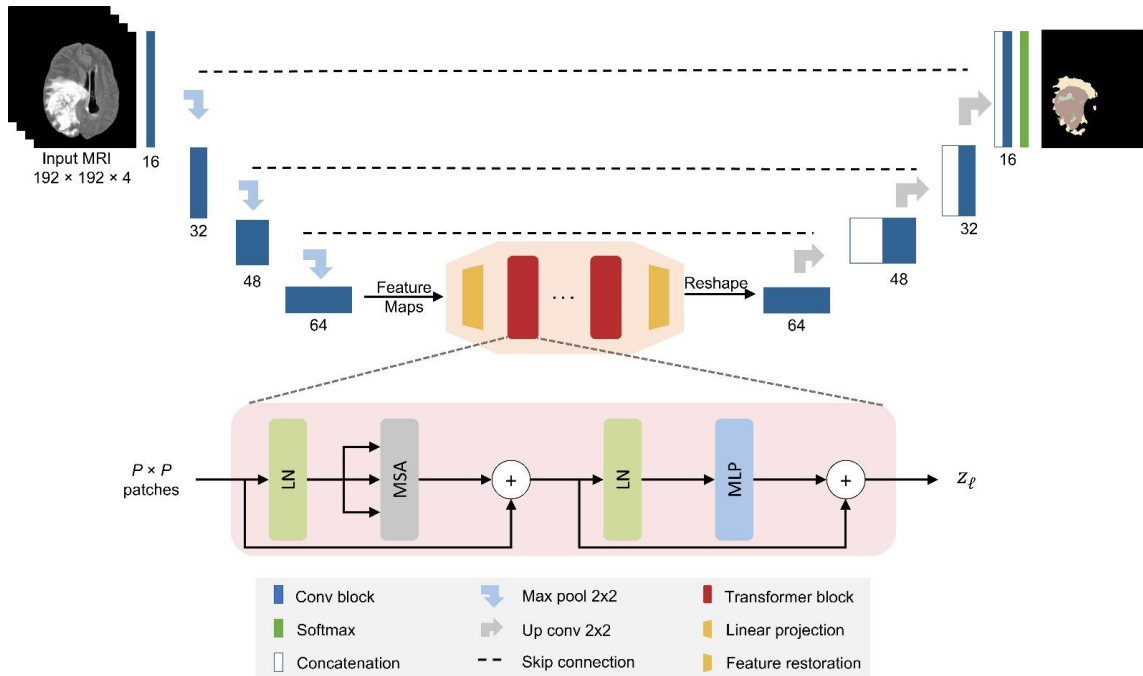


Figure 6.6: The architecture of the hybrid CNN-Transformer brain segmentation network from mpMRI volumes. The input is a 2D multi-modal MRI of T1, T1Gd, T2, and FLAIR with a patch spatial resolution of  $192 \times 192 \times 4$ . The network has 8 convolution neural blocks (blue boxes), each consisting of two successive convolutional layers  $3 \times 3$ , BN layer, and ReLU activation.

$H \times W$ . This hybrid CNN-Transformer strategy allows to model local context information across spatial dimensions as well as global context for volumetric segmentation.

### Transformer Blocks

Fig. 6.6 shows a number of Transformer blocks embedded in the bottleneck of the TransXAI network. Each Transformer block (Vaswani et al., 2017) consists of two layers; MSA and MLP as discussed in Section 3.1.4.

However, using a pure Transformer as an encoder would be impractical due to its computational complexity proportional to the number of input sequences. Therefore, the Vision Transformer (ViT) approach (Dosovitskiy et al., 2020) is incorporated by splitting the  $x$  into fixed-size ( $P \times P$ ) patches image  $x_p \in \mathbb{R}^{P^2 \cdot N_c}$  and then reshaping each patch into a token. Note that the input to the ViT blocks is the extracted image representations by the convolutional neural encoder blocks instead of raw input images.

### Feature Restoration

To match the spatial resolution of the TransXAI decoder, a feature restoration module is introduced to decode the resultant features. Specifically, the Transformer's output sequence  $z_\ell \in \mathbb{R}^{\frac{HW}{P^2} \times N_c}$  is initially reshaped to  $\frac{H}{P} \times \frac{W}{P} \times N_c$ , but the direct usage of the low-resolution Transformer encoded data (compared with the original resolution  $H \times W$ ) may cause loss of low-level tumor region details. To compensate for such information loss, a  $1 \times 1$  convolutional layer is utilized which reduces the number of feature maps.

### Upsampling Path

To gradually recover the abstract features and output the full-resolution segmentation map of  $H \times W$ , progressive upsampling is performed using  $2 \times 2$  up convolutional operations. Inspired by U-Net (Ronneberger et al., 2015), low-level encoder details are fused with high-level decoder counterparts for finer semantic information with spatial details. Finally, a multi-label softmax layer is used to estimate the final probability distribution for the output predictions.

#### 6.3.2.2 Explainable CNN Generator

Since the main goal of this study is to investigate the hybrid CNN-based and Transformer model for brain segmentation, an efficient post-hoc XAI technique is integrated. This means that all experiments are carried out after the inference of the model, i.e., at prediction time. Principally, GCAM is applied from NeuroXAI (Section 6.3.1.6) to explore the spatial attention of the network predictions over internal input features based on our trials with neurosurgeons at Ulm University Hospital.



## 6.4 Experiments

### 6.4.1 Data

MRI data from the BraTS challenges 2019 and 2021 (Baid et al., 2021; S. Bakas et al., 2017; Spyridon Bakas et al., 2018; Menze et al., 2015) have been used in this study for accomplishing the classification and segmentation tasks. Each subject has four MRI sequences including pre-operative multimodal MRI scans of native T1W, Gadolinium T1Gd, T2W, and FLAIR, acquired from multiple different institutions, as shown in Fig. 6.7. Although the main aim of the challenge is to compare the best algorithms for segmenting the ET, the TC, and the WT regions, the BraTS 2019 dataset also provides classification labels for gliomas. The BraTS 2019 database comprises 259 cases of HGGs and 76 cases of LGGs, which were used for the first showcase. The second showcase applies to the BraTS 2021 database, which contains 1251 MRI images with ground truth annotations without any explicit glioma classification.

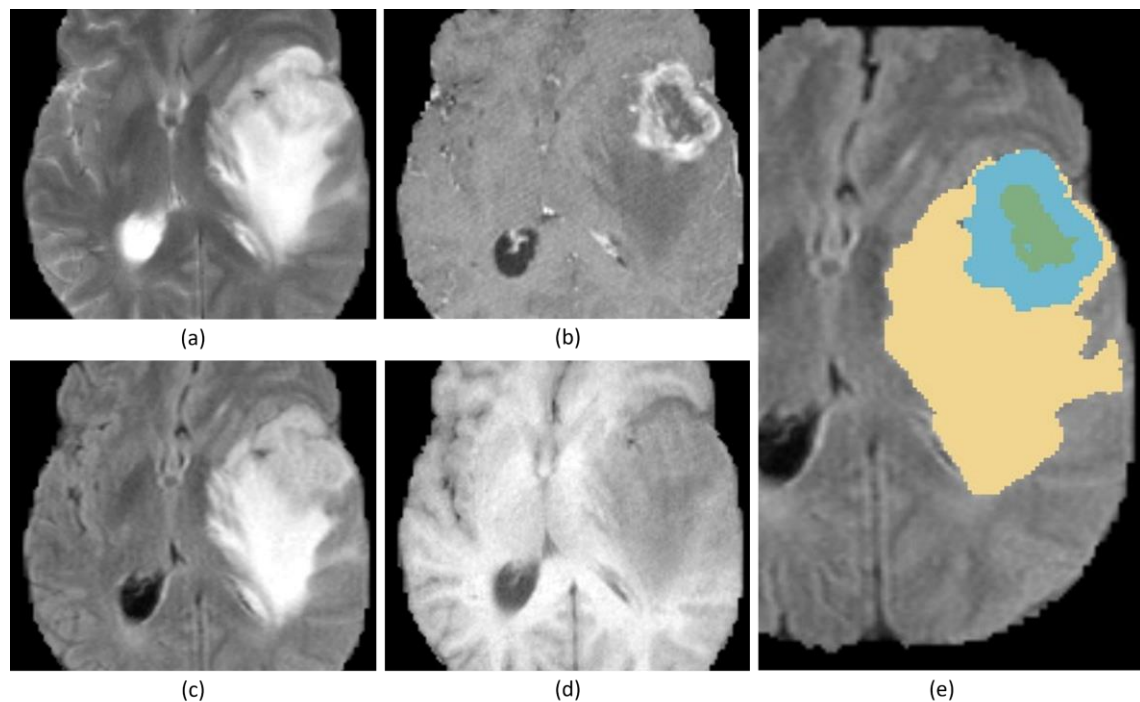


Figure 6.7: Glioma sub-regions in a sample scan from the BraTS 2019 challenge database. Image patches show the different modalities of T1 (a), T1Gd (b), T2 (c), FLAIR (d), and annotated expert-labeled tumor segmentation (e). Ground truth segmentation is provided for the enhancing tumor (*blue*) surrounding the non-enhancing necrotic tumor core (*green*) visible in T1Gd, and (b) the peritumoral Edema (*yellow*) visible in the FLAIR, respectively.

Since MRI sequences were acquired using multi-parametric instruments in multi-location centers, input images needed to be standardized. A pre-processing stage has been applied to all MRI scans, specifically min-max scaling of each MRI modality using z-score normalization, and image cropping to a spatial resolution of  $192 \times 224 \times 160$ . During the training, data augmenta-

tion was applied including random flipping, random rotations, intensity transformation, as well as dynamic patch augmentation cropping size of  $128 \times 128 \times 128$  to avoid overfitting problems.

### 6.4.2 Experimental Setup

The proposed approach NeuroXAI is post-hoc in the sense that it provides explanations after obtaining the model predictions, instead of being inserted into the CNN architecture itself. Therefore, all explainability experiments have been done after the training of the deep learning networks. Pre-trained weights for the segmentation and classification tasks were used for generating the heatmaps of the seven used XAI methods, i.e., VG, GBP, IG, GIG, SmoothGrad, Grad-CAM, and GGCAM. It is notable that our BrainXAI generator can be applied to any CNN-based network, and 3D DeepSeg (Ramy A. Zeineldin, Mohamed E. Karar, et al., 2022a) is given here as an example. In this study, the interpretability of the brain tumor segmentation system is the main focus at hand, thus a thorough accuracy evaluation is out of scope.

For the classification task, a simple classifier was employed based on a pre-trained ResNet (He et al., 2016) because of its accurate classification results. Deep transfer learning was then adopted to make the model capable of extracting features from brain MR images. Table 6.1 summarizes the added top layers to the ResNet-50 in the experiment. This model is not meant to obtain the best performance since the main goal is to illustrate the explainability capabilities of the proposed NeuroXAI framework on a set of different neural networks for two typical applications in medical image analysis.

For the segmentation problem, an encoder-decoder neural network was utilized, namely 3D DeepSeg (refer to Section 4.3.2). The structure of the proposed network is presented in Figure 6.8. The encoder, or feature extractor, is typically a CNN comprising successive blocks each consisting of  $3 \times 3 \times 3$  convolutional layers, BN, a ReLU activation function, and a  $2 \times 2 \times 2$  max pooling operation. On the other hand, the decoder, or segmentation estimator, aims at upscaling the resultant feature maps by consecutive blocks, each consisting of two deconvolutions,  $2 \times 2 \times 2$  up-convolution, and a ReLU activation function. Skip connections connect the two parts of the neural network to combine high-resolution feature maps from the encoder and the corresponding semantic features from the decoder.

Table 6.1: List of the added top layers to the standard ResNet-50.

Type	Output	Feature Maps
Average Pooling 2D	$2 \times 2$	512
Flatten	2048	1
Dense	256	1
Dropout	256	1
Dense	2	1

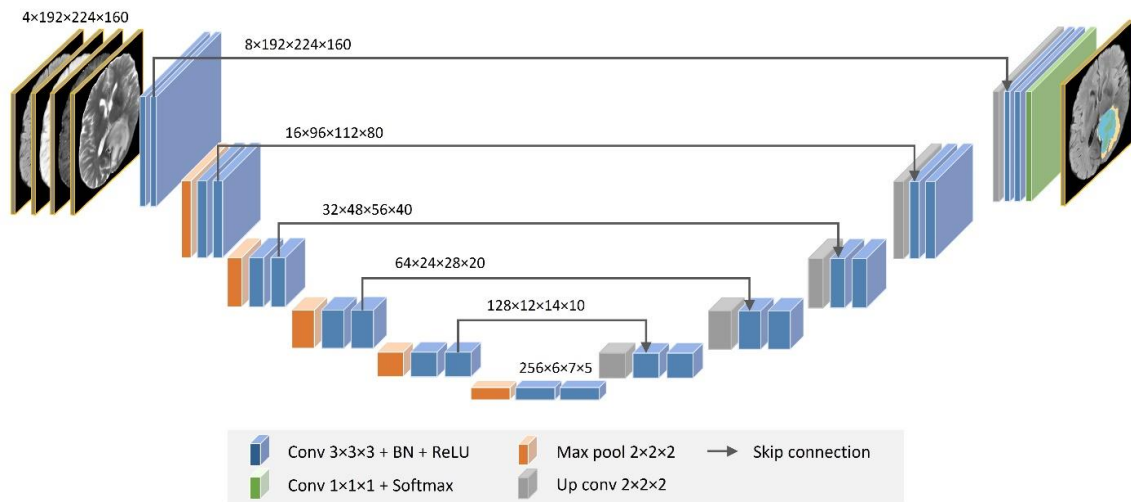


Figure 6.8: The architecture of the enhanced 3D brain segmentation network (DeepSeg) for brain tumor segmentation from mpMRI volumes. The input is a 3D multi-modal MRI of T1, T1Gd, T2, and FLAIR with a patch spatial resolution of  $192 \times 224 \times 160$ . The CNN network has 24 convolution neural blocks (blue boxes), four downsampling blocks (orange boxes), four upsampling blocks (grey boxes), and a final softmax output layer (green box).

Both deep learning models were implemented using the TensorFlow library (Abadi et al., 2016) version 2.4. Adam optimizer (Kingma & Ba, 2014) was used to update the weights of the network, with an initial learning rate of  $1e - 3$  and  $1e - 4$  at the very beginning, the maximum number of training epochs is set to 150 and 1000, and batch size of 64 and 5 for the classification and segmentation networks, respectively. Training the networks was performed on a single NVIDIA graphic card (RTX 2080Ti with 11 GB RAM or RTX 3060 with 12 GB RAM). Explainability experiments were carried out after the training of the original neural network because of using post-hoc XAI methods without network re-training or architecture modifications. The final sensitivity maps were generated by the proposed NeuroXAI framework with the pre-trained saved weights for both deep learning models.

## 6.5 Results and Discussion

### 6.5.1 Showcase I: Application to Classification

Here, the application of NeuroXAI to generate visual explanations for automatic brain glioma grading using deep learning is introduced. To better understand the deep model's prediction, the trained model is used to visualize various gradient-based sensitivity maps (refer to Fig. 6.9 and Fig 6.10). Notably, the classifier model achieved a superior accuracy of 98.62 and specificity of 98.98 for the unseen validation set which is comparable with the state-of-the-art (Ge, Gu, Jakola, & Yang, 2018; Muhammad, Khan, Ser, & Albuquerque, 2021).

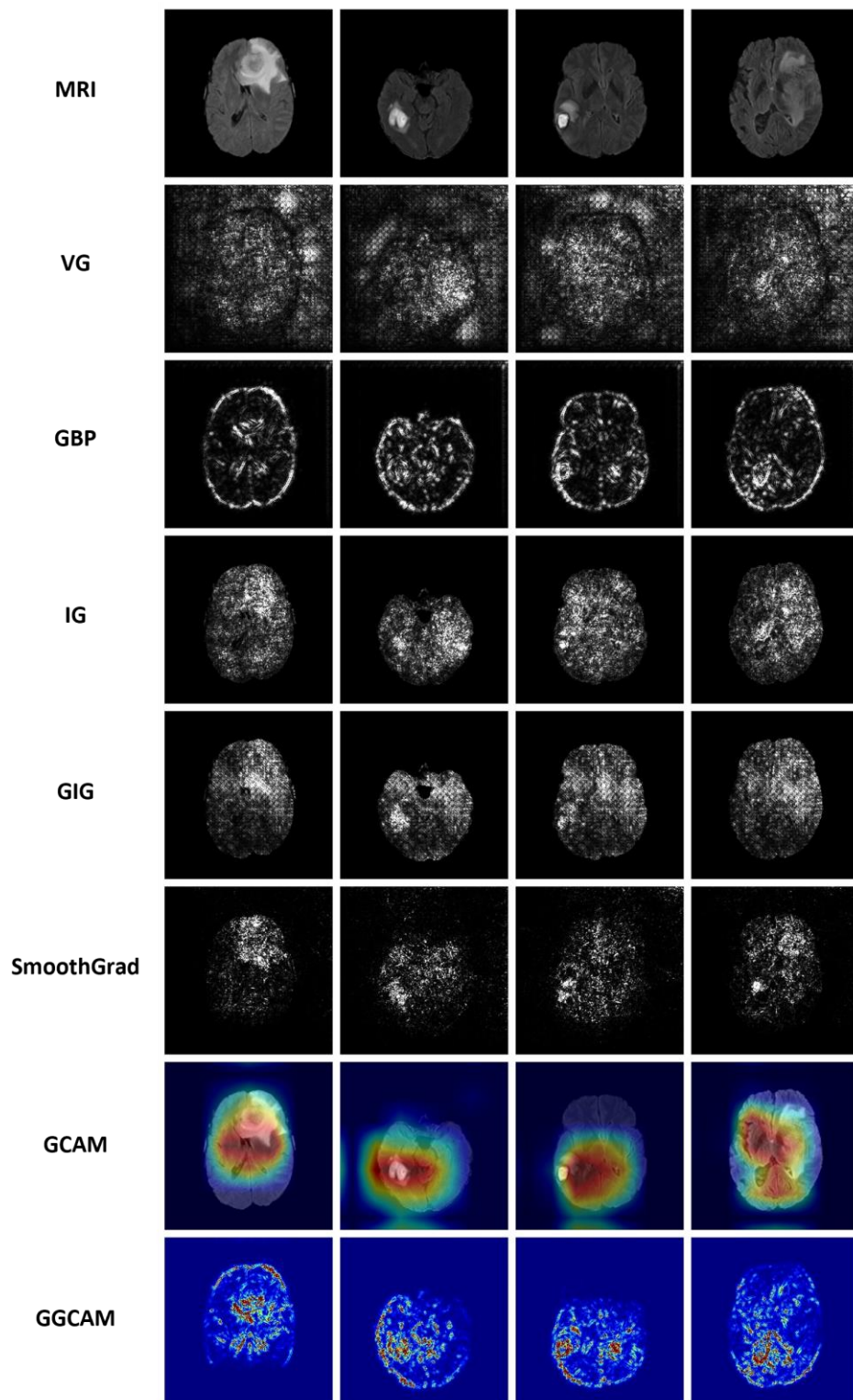


Figure 6.9: Comparing different XAI visualization methods for brain glioma classification. Sensitivity maps are presented for four HGG cases.

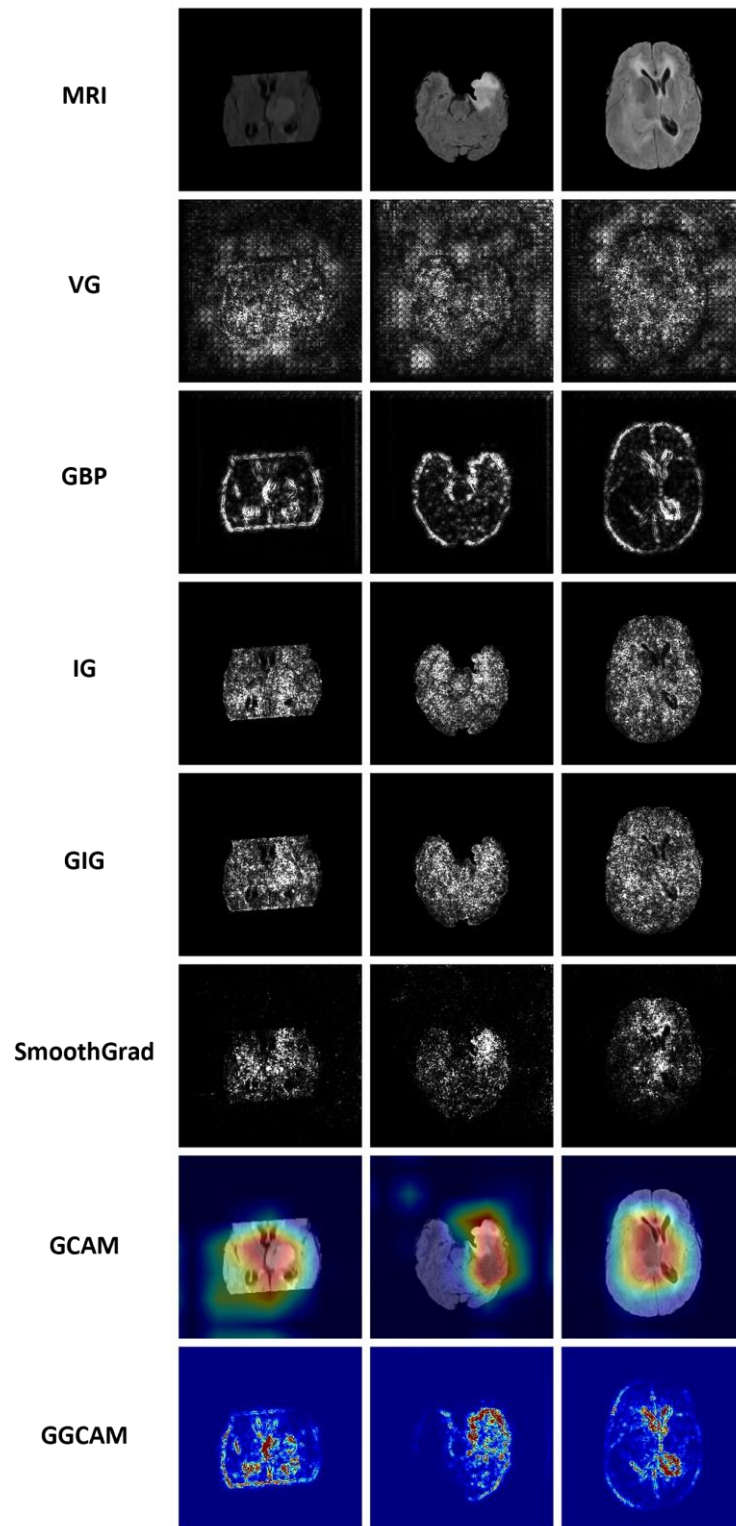


Figure 6.10: Comparing different XAI visualization methods for brain glioma classification. Sensitivity maps are presented for three LGG cases.

Each XAI method is unique and can be helpful in a different scenario with its inherent advantages and limitations. For example, VG is simple with the advantage of being supported by conventional machine learning frameworks such as TensorFlow (Abadi et al., 2016) and PyTorch (Paszke et al., 2019). This makes VG applicable to any DNN without architectural modifications. On the other hand, the saliency maps generated by VG are noisy as well as they suffer from declining influences of features due to gradient saturation as reported in previous work (Shrikumar, Greenside, Shcherbina, & Kundaje, 2016). GBP is efficient in terms of implementation; however, it is limited to CNN models with ReLU activations and does not provide class-distinctive visualization maps.

Recently, IG has become popular thanks to the ease of implementation, no requirement for instrumentation of the network, and the fixed number of calls to the gradient. GIG is an enhancement to eliminate the false perturbations problem of IG, but a choice has to be made at every step of the path from baseline to input, and thus the direction of the path is not fixed. Although SmoothGrad can help improve visualizations of the overall true signal the major drawback of being non-class-discriminative. Conversely, GCAM allows interpreting any convolutional layer of the CNN by highlighting the discriminative region and thus can help in understanding the internal functionality. To eliminate the lower resolution heat maps problem of GCAM, GGCAM was implemented as the combination of GBP and GCAM advantages.

From Fig. 6.9 and Fig. 6.10, it can be seen that visualization maps by pixel-space XAI methods, such as GBP, IG, and GIG, underlined fine-grained details in the input MRI image, but are not class-distinctive. In contrast, localization approaches like GCAM, are highly class-distinctive providing a smooth activation map. Notably, combining GBP with GCAM yielded better-localized visualizations with high resolution. SmoothGrad provided the best overall feature maps highlighting the main discriminative parts of the input FLAIR image so as to make the glioma grading. In contrast, VG provided noisy visualization maps compared with other methods due to the gradient saturation as reported by (Shrikumar et al., 2016), making it less reliable for this application.

## 6.5.2 Showcase II: Application to Segmentation

In this subsection, a feasible application of NeuroXAI is provided to interpret deep brain glioma sub-region segmentation using multimodal MRIs. In encoder-decoder networks, like DeepSeg, generated saliency maps for one of the last encoder layers are smooth and do not capture feasible information in the segmentation problem. This is due to the fact that these layers generate the smallest feature dimensions in the network and intensive upscaling is required to match the output prediction map. In contrast, selecting one of the last layers (e. g. the output layer) from the decoder network provides a higher-resolution feature map showing detailed features of the segmentation process since these layers are combined with the encoder layers through concatenation. Moreover, by incorporating the output layer into the explanation generation process, the limitation of GCAM for generating low-resolution heatmaps is resolved.

Since segmentation networks pinpoint the localized region of brain tumor regions, providing visual saliency maps of the output layer alone does not help in making the network transparent. Therefore, to better investigate the behavior of the deep model and to determine how spatial information flows inside the internal layers, four main experiments were conducted to extend the explainability approach as follows:

- Quantitative evaluation of the BraTS validation database and comparison with SOTA 2D and 3D methods.
- Identifying the contribution of each MRI input modality in the final predicted tumor sub-components.
- Interpreting the CNN layers using the best XAI method to reveal how the network represents information in the internal filters.
- Clinical feedback on the proposed method from the clinical collaborators.

In the following subsections, the above experiments are described, and their results are discussed.

### 6.5.2.1 Comparing Different XAI Methods

To be able to compare different XAI methods' explanations, the BraTS segmentation task is converted into a multi-label classification problem by splitting the final convolutional layer class-wisely. Figure 6.11 shows the qualitative results from different XAI methods for explaining the glioma segmentation network as well as the dice coefficient score for each case as a quantitative metric. Further, Table 6.2 lists the average DSC and HD95 for the proposed model as reported by the BraTS 2021 evaluation platform (Baid et al., 2021).

Table 6.2: Quantitative results of CNN models on the BraTS 2021 Challenge database.

Tumor Component	DSC	HD95
Enhancing Tumor	84.10	16.0179
Tumor Core	87.33	8.9077
Whole Tumor	92.00	3.8097
Average	87.81	9.5784

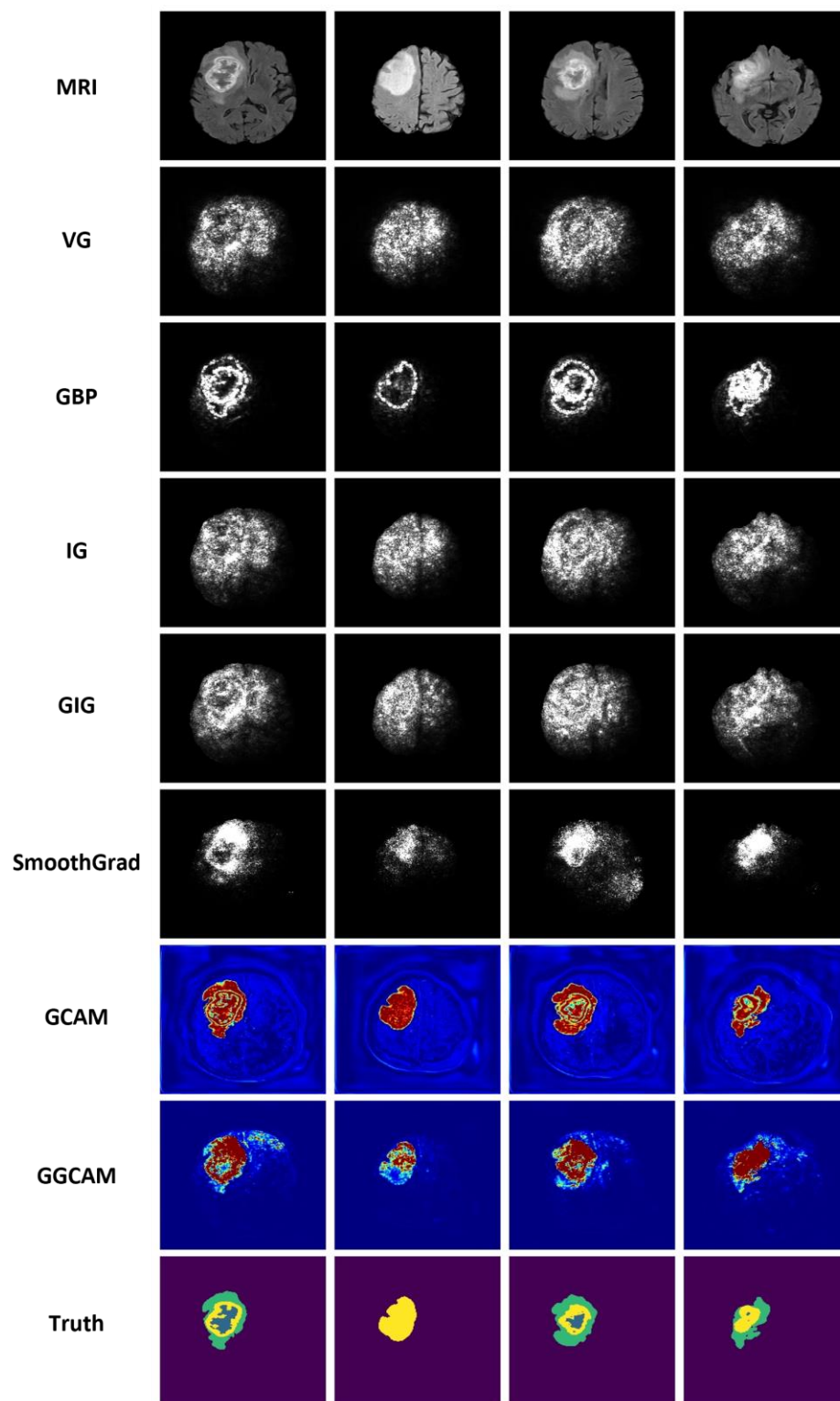


Figure 6.11: Comparing different XAI visualization methods for brain glioma segmentation.



It can be seen that the employed visualization methods generally clustered their attributions around the segmented brain tumor. In particular, GCAM, GGCAM, and SmoothGrad provided the least noisy visualization maps with the advantage of GCAM of being class-discriminative. GBP generated high-resolution saliency maps in which the edges of the tumor sub-regions are highlighted instead of the tumor itself.

Besides, each layer output was analyzed using the GCAM method towards the transparency of the black-box segmentation model. Figure 6.12 provides these explanation maps following the layers from the input MRI scans to the predicted segmentation map. It can be seen that the DNN can learn some explicit concepts, which the CNN was not originally trained on, as well as implicit concepts from the underlying dataset. For instance, layer 22 in Fig. 6.12 (g) seems to be learning the whole tumor region, as an explicit concept from the ground truth labeling data. Another example is shown in Fig. 6.12 (c) for layer 3 learning the Grey and White Matter as an implicit concept that is not included in the training annotations.

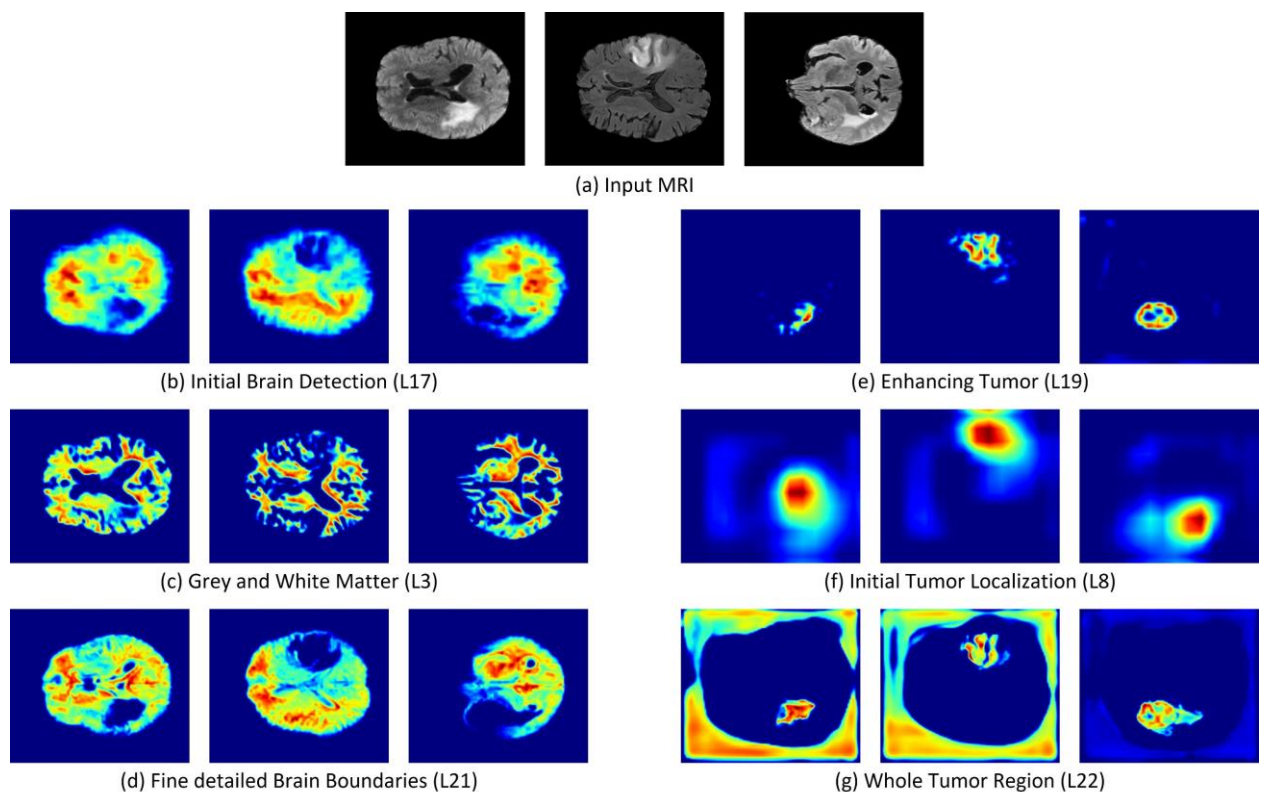


Figure 6.12: Visualization of the information flow in the segmentation CNN internal layers. The input MRI sequences are shown in (a). (b) – (d) show implicit concepts for which no ground truth labels are available in addition to explicit concepts (e) – (g) with trained labels. L stands for convolutional layer.

Recently, Adebayo et al. (Adebayo et al., 2018) observed that the dependence on visual assessment alone can be misleading proposing an extensive evaluation strategy for interpreting the adequacy of several explanation approaches. First, they carried out model parameter randomization tests where the output of a visualization method on a pre-trained network is compared against the same network output with randomly initialized weights. Second, data randomization

tests were performed to evaluate the behavior of saliency maps on randomly permuted labeled datasets against the original labeled one. The results highlight that GCAM passes the sanity checks, while both GBP and GGCAM fail these two sanity checks and remain almost unchanged across all architectures and datasets perturbations regardless of network performance degradation. In particular, the output of explanation maps of GBP and GGCAM would not be reliable for CNN debugging tasks like this application. Therefore, the next experiments were performed by using only GCAM for generating trustworthy heatmaps.

### 6.5.2.2 Role of MRI in Tumor Detection

To better interpret the behavior of the CNN model, a further experiment was performed for generating visual explanations of every tumor class using GCAM. TransXAI was experimented with to infer with a specified MRI modality without involving other MRI sequences. This led to understanding the importance of each MRI input, namely, T1, T1Gd, T2, and FLAIR in the process of different tumor label localization.

Fig. 6.13 outlines the visual representation captured by the output convolutional layer of the TransXAI model with respect to the input MRI modality. The results demonstrate that the

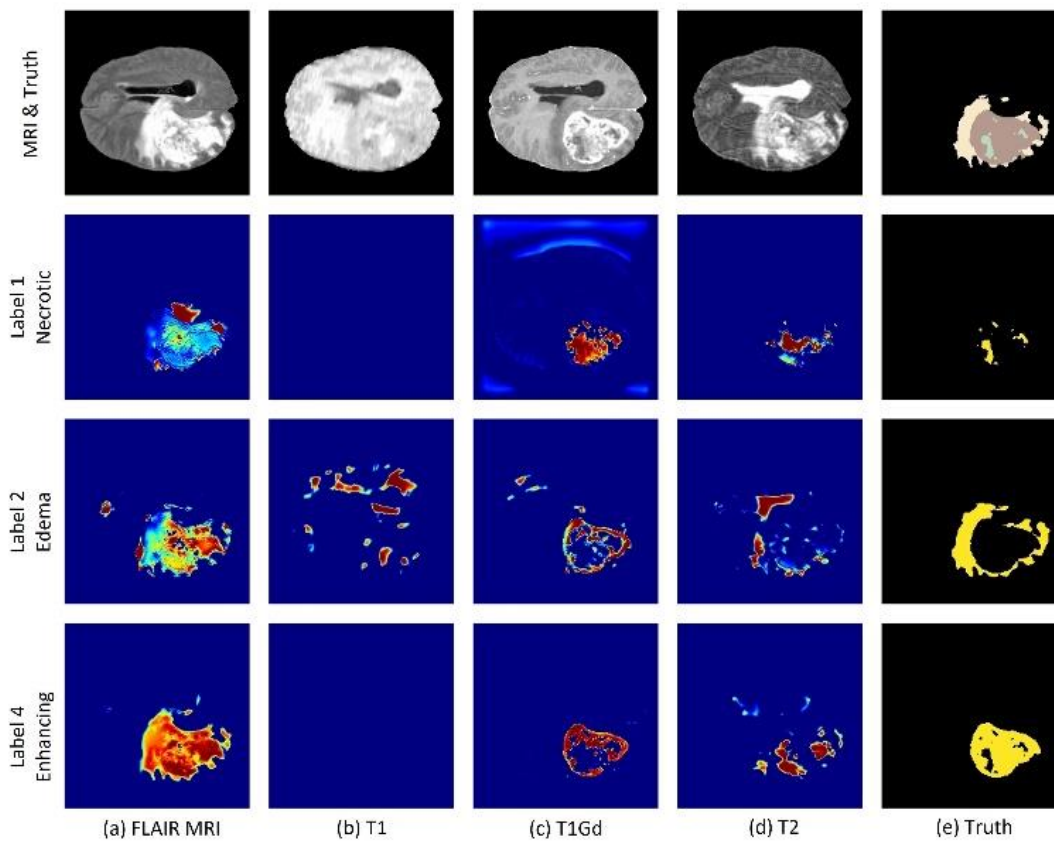


Figure 6.13: Impact of MRI input modality in the detection of different tumor labels. The first row shows the input MRI sequences and the ground truth annotations. The following rows correspond to label 1 (the necrotic tumor core), label 2 (the peritumoral edema), and label 4 (the enhancing tumor). In the saliency maps, warmer regions represent a high score for the specified label detection.

detection of each tumor sub-region is related to one or more of the input MRI volumes coherent with expert radiologists' and raters' observations in reference (Menze et al., 2015). For instance, T1Gd and T2 contribute most to the detection of the gross TC, including both label 1 (NC) and label 4 (ET), while label 2 (ED) and the WT region are predicted using FLAIR. Though, the visual explanations of T1 are the least important maps with very little contribution to the tumor sub-components segmentation and could, therefore, be removed for computational performance advantage without model accuracy degradation.

### 6.5.2.3 GCAM for CNN Layers

In this section, GCAM has been applied to interpret the proposed TransXAI for tumor segmentation. Fig. 6.14 and Fig. 6.15 show saliency maps for the internal convolutional layers of the investigated CNN model. These visual explanations provide details on the information flow inside individual filters of the network and how it learns some meaningful concepts. In this hybrid network, the encoder typically consists of successive layers to capture contextual

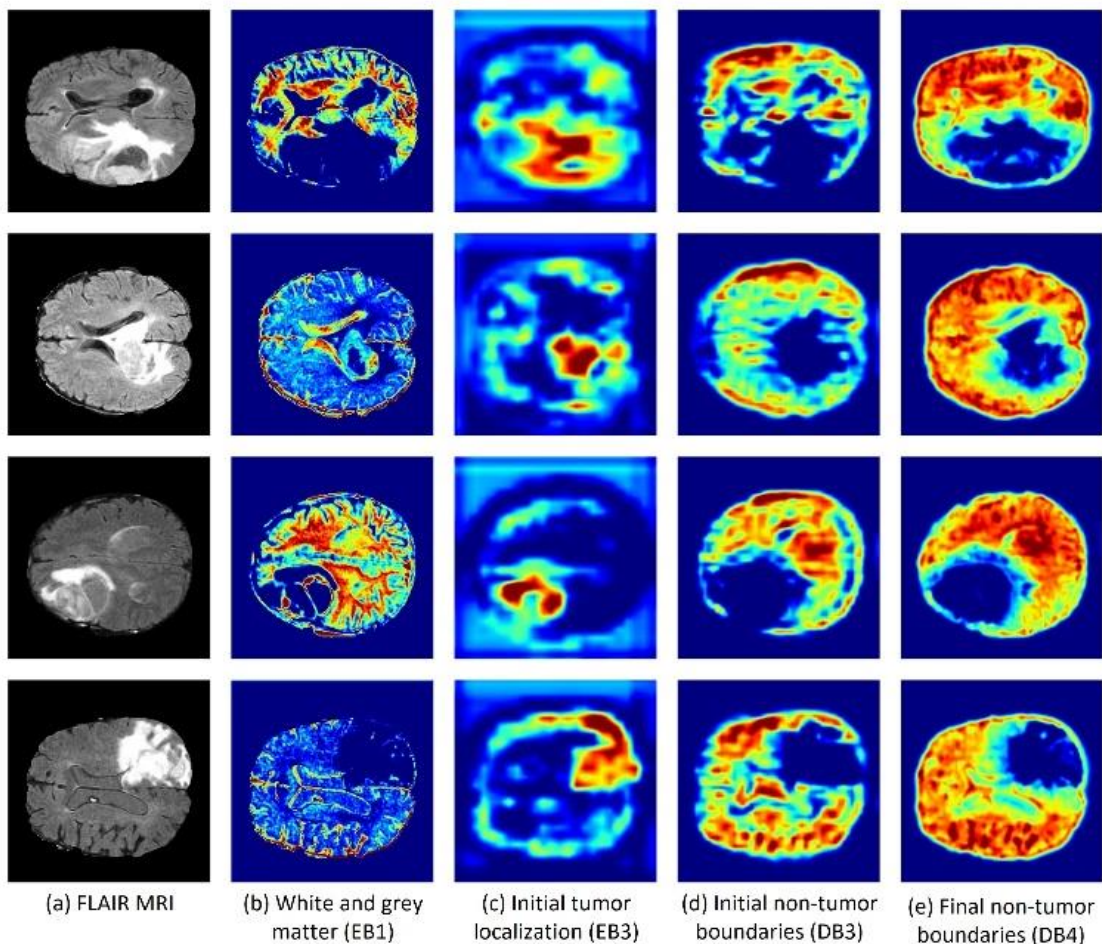


Figure 6.14: Saliency maps for implicit concepts learned by individual filters of the CNN model. It is interesting to note that there are no labels for these concepts in the training dataset. Warmer regions represent a high score for the specified concept in the prediction map. Note that EB and DB denote the encoder and decoder block layers, individually.

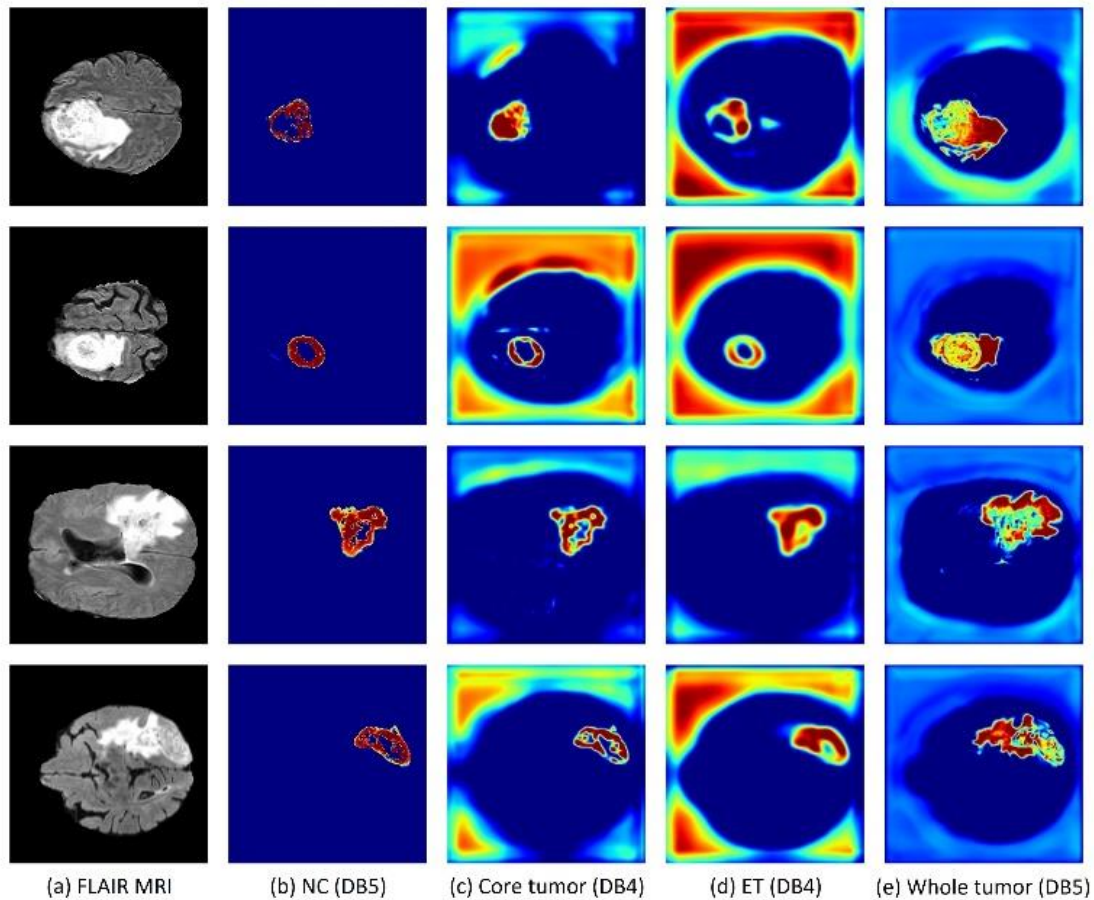


Figure 6.15: Saliency maps for explicit concepts learned by individual filters of the CNN model which are labeled in the training dataset. Warmer regions represent a high score for the specified concept in the prediction map. Note that EB and DB denote the encoder and decoder block layers, individually.

information, Transformer blocks embedded in the bottleneck, and the expanding decoder path contains upsampling operators to enable high-resolution localization of the target tumor voxels.

It is important to observe that, internal layers of the DNN learn some implicit as well as explicit concepts although the training stage included only explicit tumor labels. For example, Fig. 6.14 (b) demonstrates how the model implicitly differentiates white matter and gray matter region in encoder block 1 which the network has not been trained to learn. Similarly, the network understands other implicit concepts such as initial and final non-tumor boundaries in decoder block 3 (Fig. 6.14 (d)) and block 4 (Fig. 6.14 (e)), correspondingly. In addition, the CNN model learns explicit brain tumor sub-regions as depicted in Fig. 6.15.

Furthermore, the results show that the proposed network follows a top-down approach for detecting and segmenting brain glioma. First, the model starts with learning the entire brain tissue, followed by the initial tumor boundaries, and finally, small objects and fine details are localized. In Fig. 6.15, some examples of finer segmentations are presented for the expansive path. Such filters outline the NC (label 1) in Fig. 6.15 (b), the ET (label 4) in Fig. 6.15 (d), the

TC region (label 1 and label 4) in Fig. 6.15 (c), and the WT region (all labels) in Fig. 6.15 (e). Neural networks are mathematical representations that have been motivated by the functioning of the brain (Guo et al., 2016). The findings are consistent with the top-down coupling approach that the human brain follows for the comprehension of relevant visual features in global first then local areas (Dijkstra, Zeidman, Ondobaka, van Gerven, & Friston, 2017).

#### 6.5.2.4 CNN Node Failure Detection

To allow for CNN transparency, an important question should be answered such as why the model has failed to predict some tumor sub-regions or provided a partial segmentation map. This allows us to enhance the model's accuracy by avoiding these problems during training time. In addition, the visual explanation may help medical experts gain human-understandable interpretations for the inference. Therefore, this analysis was performed to identify the neural network's failure nodes. The model first predicts segmentation maps, then the DSC scores were calculated for each tumor component in the BraTS validation set, as presented in Fig. 6.16, and finally, saliency maps were generated for subjects with the lowest performance.

Figure 6.17 gives the ground truth annotations as well as GCAM visual explanations obtained at selected intermediate layers for three validation subjects from BraTS 2021. These concept-oriented activation maps aid in debugging why the CNN model has failed to detect such areas. In Fig. 6.17 (a), the DSC of TC (label 1 and label 4) is the highest at 0.92 followed by ET (label 4) at 0.85. However, the network failed to predict the ED (label 2) accurately resulting in WT (label 1, label 2, and label 4) of dice score of 0.48. The reason behind that can be seen in the corresponding saliency maps where it seems that the model localized the tumor region incorrectly at the beginning as seen in encoder block 4. Further, the model misclassified the tumor as part of the healthy brain in the encoder block 4. Consequently, the resulting whole tumor region is only about half of the original one leading to that DSC value.

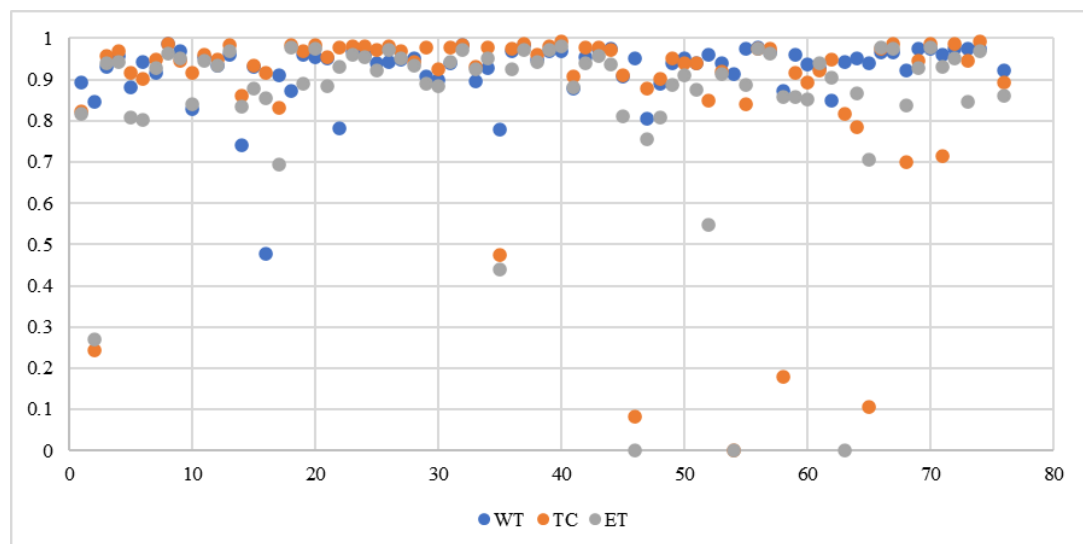
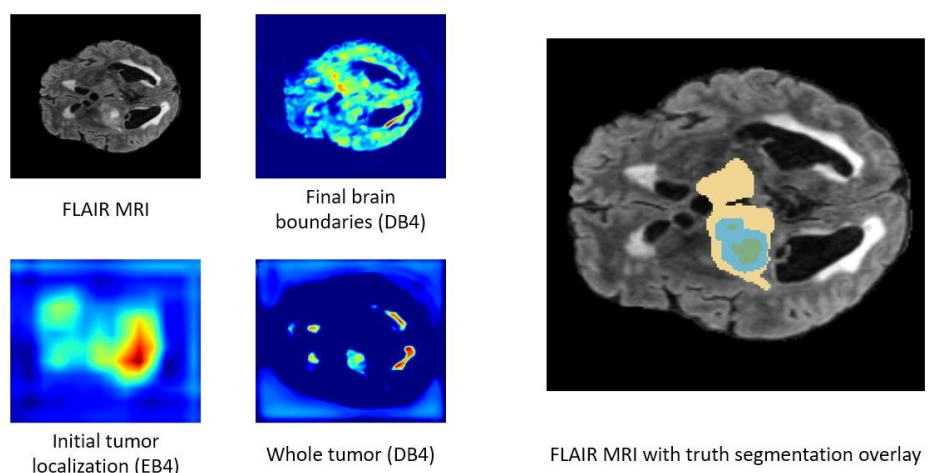
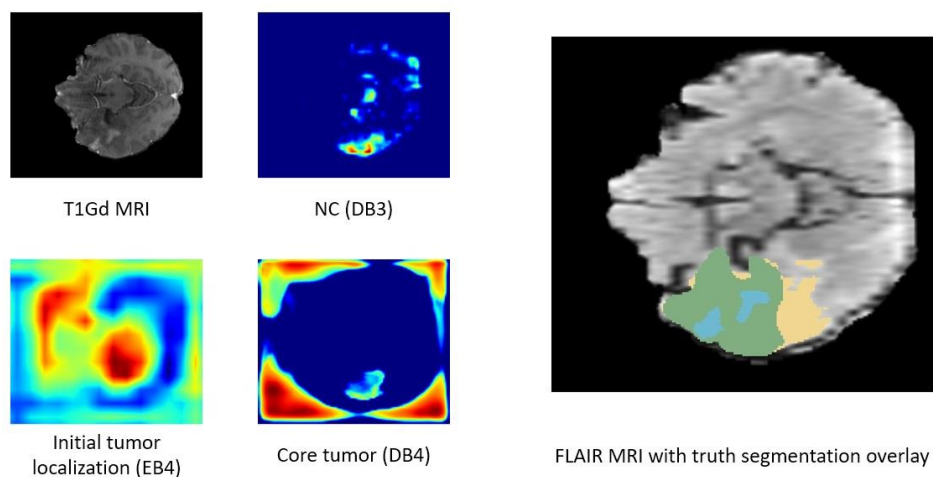


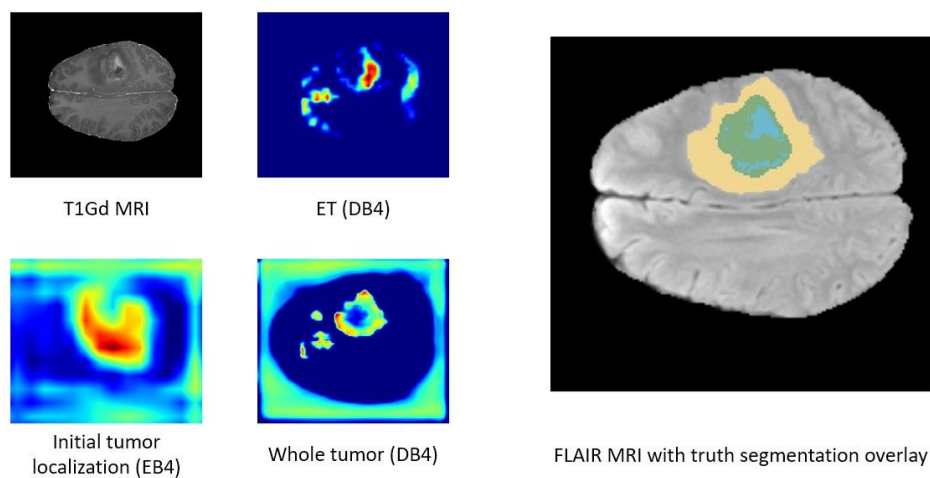
Figure 6.16: Prediction failure detection DSC metrics on 70 selected subjects from the BraTS 2021 validation database. Each point per color represents one of the brain tumor sub-regions. WT, TC, and ET stand for Whole Tumor, Tumor Core, and Enhancing Tumor, respectively.



(a) Subject #00493 with DSC 0.48, 0.92, 0.85 for WT, TC, and ET



(b) Subject #01438 with DSC 0.87, 0.18, 0.86 for WT, TC, and ET



(c) Subject #01291 with DSC 0.96, 0.85, 0.55 for WT, TC, and ET

Figure 6.17: Overview of the relevant gradient-based class activation maps for failure cases.

Another subject #01438 is presented in Fig. 6.17 (b), where the dice score for the tumor core (label 1 and label 4) is the lowest of 0.18 and the ET core (label 4) is 0.86. To reveal what made the network misidentify the NC (label 1), the GCAM heatmaps were generated for corresponding filters. Although the model localized the tumor correctly, it focused on a small part of the core tumor voxels, and this is apparent in the saliency map of the core tumor region afterward. Similarly, the visual interpretations in Fig. 6.17 (c) illustrate that the neural network predicted the whole tumor and core tumor regions accurately but fails to obtain a similar result for the ET (label 4). From this figure, it can be seen that the CNN model somewhat captured correct information about the ET but it focused on other parts of the tumor as well.

## 6.6 Clinical Relevance

In the medical domain, explainable and interpretable AI systems are crucial for many applications such as research, education, and clinical treatments (Holzinger, Biemann, Pattichis, & Kell, 2017). XAI systems can support medical professionals to understand the prediction process followed by deep learning models, and thus, enhance human experts' trust in the system's decisions. In this section, clinical partners at the Department of Neurosurgery, University of Ulm evaluated the above extensive experiments in Section 6.5.

First, it is helpful to provide different methodologies of explanation, however, the second XAI method, namely GCAM was found to be easier to understand and interpret from the surgical point of view. Moreover, it showed a better way to evaluate the capability of the software to identify the tumor boundaries and to differentiate between different structures in the MRI images.

Second, utilizing FLAIR and T2 MRI scans in the clinical routine could lead to a better and more accurate estimation of the perifocal edema surrounding the tumor. While applying the T1Gd could lead to a better assessment of the tumor boundaries in high-grade intra-axial lesions. The decreased accuracy in detecting the tumor boundaries in low-grade lesions can be explained by the failure to identify the tumors in T1, where these lesions do not show a marked enhancement. In these conditions, tumor segmentations would rely more on the T1Gd, T2, and FLAIR sequences incoherent with the reported results.

Finally, the software showed a logical systematic process during the segmentation process, which has a significant similarity to the way used to identify the neoplastic tissues in the clinical practice, the perifocal edema, and the surrounding brain parenchyma. Finally, providing activation maps of internal filters in the neural network could help to provide more confidence about the prediction output and transparency about the "black box" machine learning systems. This may be valuable to generate on-demand human-understandable interpretations to help medical specialists in assessing the legitimacy of the segmentations and thus motivate them to evaluate it in their clinical trials.

## 6.7 Summary

This chapter presented a new explainability framework, named NeuroXAI, for assisting the interpretation of the behavior of deep learning networks using state-of-the-art visualization attention maps. NeuroXAI is post-hoc and can therefore be applied to any deep neural models gaining insight into the behavior of these already trained models. Since visual pixel-based representations are not enough to give meaningful information, extensive experiments were conducted to provide interpretability by evaluating their clinical significance. Additionally, the two showcases have demonstrated the significance of incorporating XAI methods in medical image analysis tasks. NeuroXAI can also support the analysis of CNNs by providing an individual activation map for every internal filter.

The obtained results supported the technical research work to realize that deep neural models behave in a human-understandable manner and are consistent with the surgical experts' domain knowledge. Moreover, XAI methods showed the importance of explainability for medical imaging tasks to understand deep learning models to accelerate their clinical acceptance by medical staff in the field.



# 7 DeepIGN Integration and Neuronavigation Display

---

## Contents

<b>7.1</b>	<b>Introduction</b>	<b>104</b>
<b>7.2</b>	<b>Related Work</b>	<b>104</b>
<b>7.3</b>	<b>System Design</b>	<b>104</b>
<b>7.3.1</b>	<b>Design Requirements</b>	<b>105</b>
<b>7.3.2</b>	<b>System Components</b>	<b>106</b>
<b>7.3.2.1</b>	<b>3D Slicer Software</b>	<b>106</b>
<b>7.3.2.2</b>	<b>Hardware Interfacing Layer</b>	<b>107</b>
<b>7.3.2.3</b>	<b>SlicerIGT Calibration</b>	<b>108</b>
<b>7.3.2.4</b>	<b>DeepIGN Extensions</b>	<b>108</b>
<b>7.4</b>	<b>Results and Discussion</b>	<b>110</b>
<b>7.4.1</b>	<b>Pre-operative Application to BraTS Dataset</b>	<b>110</b>
<b>7.4.1.1</b>	<b>Use Case Example</b>	<b>110</b>
<b>7.4.1.2</b>	<b>Runtime Results</b>	<b>111</b>
<b>7.4.2</b>	<b>Phantom Study</b>	<b>111</b>
<b>7.4.2.1</b>	<b>Experimental Setup</b>	<b>111</b>
<b>7.4.2.2</b>	<b>DeepIGN Workflow</b>	<b>112</b>
<b>7.4.2.3</b>	<b>System Accuracy</b>	<b>117</b>
<b>7.4.3</b>	<b>Comparison with IGN Platforms</b>	<b>118</b>
<b>7.5</b>	<b>Summary</b>	<b>120</b>

The fourth module of the proposed IGN pipeline (see Figure 1.2 “DeepIGN”) called *Development of an intuitive neuronavigational display in the operating room* is described in this chapter. This includes a short introduction to the neuronavigation system in Section 7.1. Section 7.2 summarizes the relevant neuronavigational systems. Then, the overall concept of the proposed AI-assisted toolkit for neuronavigation is described in Section 7.3. In Section 7.4 the results of the extensive experiments are presented and discussed. A brief summary of this chapter is provided in Section 7.5. Parts of this chapter have previously appeared in (Ramy A. Zeineldin, Weimann, Karar, Mathis-Ullrich, & Burgert, 2021).

## 7.1 Introduction

Neuronavigation is crucial in all phases of the IGN and contributes to the success of brain surgeries, for instance, it has an effect on the residual tumor mass and the post-operative follow-up treatment plan. Therefore, reliable and robust neuronavigation software is mandatory for precise and effective neurosurgical guidance. Neuronavigation allows the tracking of surgical tools in real-time and the presentation of surgical tools on an intuitive display in the correct orientation with respect to a virtual patient model.

## 7.2 Related Work

Several commercial software platforms have been widely developed for the basic principles of image guidance such as Medtronic's StealthStation and Brainlab's Curve system. However, these proprietary systems are typically built for specific applications and due to their restrictive licenses, lack flexibility and extensibility, which are two main factors in developing research toolkits. Consequently, various initiatives have been launched for open-source medical research toolkits, such as 3D Slicer (Fedorov et al., 2012), MITK (Goch, Metzger, & Nolden, 2017), and ITK-Snap (Yushkevich et al., 2006). With the addition of OpenIGTLink (Tokuda et al., 2009) and PLUS (Lasso et al., 2014), open-source general medical imaging platforms can also provide integrated multimodal imaging acquisition and navigation such as SlicerIGT (Ungi et al., 2016) and NifTK (Clarkson et al., 2015). IBIS and CustusX are both open-source software platforms developed for assisting brain surgery; however, integrating with commercially available navigation systems is non-trivial and the scope of integrating iUS with additional functionalities such as neuromonitoring is limited.

Lately, 3D Slicer and MITK have integrated the AI framework NVIDIA Clara (W. Zhu et al., 2020) for automatic segmentation using deep learning. NVIDIA Clara is a cloud-powered application framework that enables automatic tumor volume inference using some pre-trained models. Similarly, Mehrtash et al. (Mehrtash et al., 2017) proposed DeepInfer, which is an application package developed as an extension for Slicer 3D to provide over-the-cloud deep learning models. Shortcomings of the above extensions include dependency on the network speed for uploading the MRI image data and downloading the predicted segmented model. Additionally, major privacy concerns become apparent since sensitive imaging data must be transferred over the internet before being processed using the backend network.

## 7.3 System Design

The development of image-guided surgery (IGS) applications is supported by several existing open-source platforms which can be categorized according to their level of system integration. Toolkits with a high level of system integration are ready to use in specialized image-guiding applications, while less integrated platforms that require significant investment in software engineering. On the other hand, less integrated toolkits have the advantage of being more flexi-

ble and supporting a range of applications, whereas high integrated systems sacrifice flexibility in favor of being specialized on single or multiple tasks with more enhanced usability.

Figure 7.1 depicts existing open-source platforms dedicated to IGS and their dependencies. At the lowest level of integration, *Hardware*, and common libraries are required for building IGS applications. The following level is called *General Platform*, which contains general imaging platforms such as 3D Slicer and MITK. By installing plugins or external libraries, general imaging software can be applied for IGS applications such as SlicerIGT and NifTK which is the third layer of integration called the *IGS Plugin*. Fully integrated IGS platforms for specific applications can be found at the final layer of the *Application*. For instance, CustusX and IBIS research software focuses on the integration of iUS for brain shift visualization toward assisting neurosurgery. The proposed AI-based IGN toolkit, namely DeepIGN, lies at the fully integrated end layer.

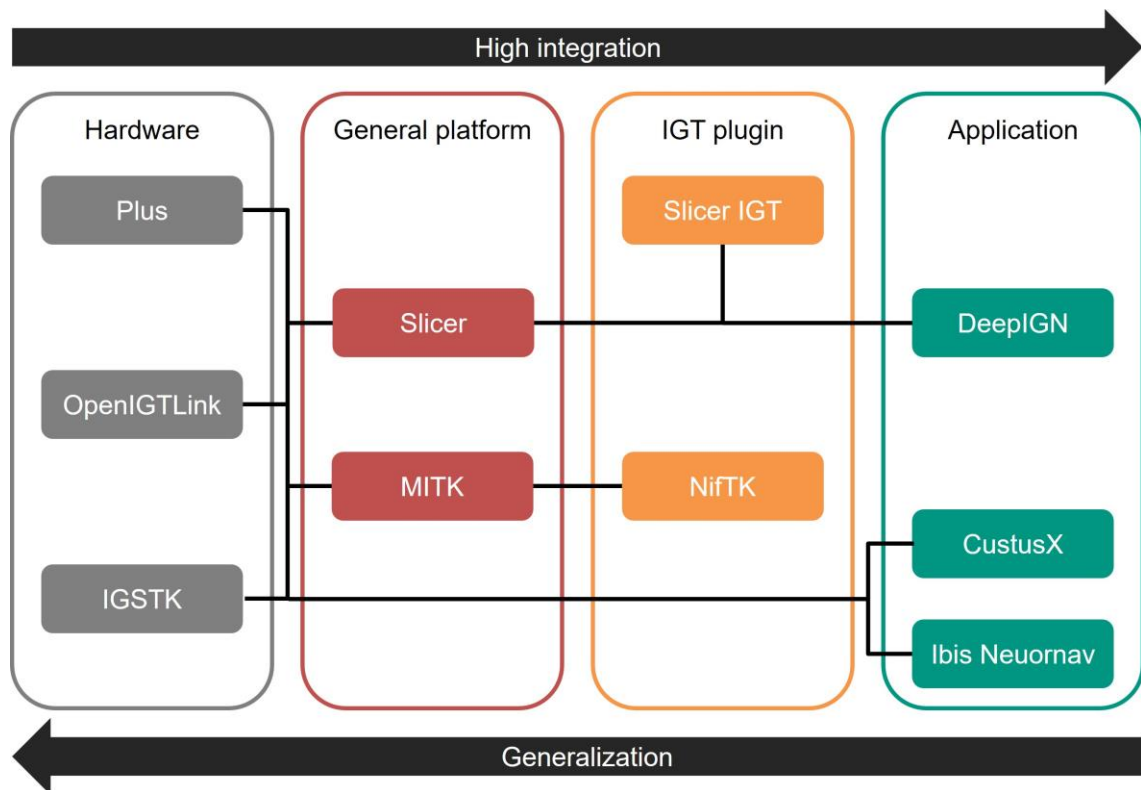


Figure 7.1: Existing open-source platforms dedicated to IGS and their dependencies. From left to right: (1) Common libraries for interfacing with the hardware in the operating room. (2) General imaging platforms that provide visualization capabilities with customizable interfaces. (3) Add-on plugins for the general-purpose platforms in level (2) to provide image-guidance capabilities. (4) Fully integrated IGS platforms for specific target applications.

### 7.3.1 Design Requirements

This section presents the design requirements for an integrated multimodal imaging and navigation system for assisting neurosurgery called the DeepIGN system. Table 1 provides an overview of the design requirements (RX) of the proposed system, which can be categorized into

two groups: clinical environment requirements (CR) and technical requirements (TR). For the purposes of this study, clinical requirements were determined through consultation with the surgical team and understanding the medical device regulations (Pesapane, Volonte, Codari, & Sardanelli, 2018). Specific technical requirements were then determined to satisfy these clinical requirements.

Table 7.1: Design requirements for a multimodal image-guided navigation system for assisting neurosurgery. CRX and TRX stand for clinical and technical requirements, respectively.

Requirement	Description
CR1. System assembly	System assembly should be straightforward and should ideally be completed within 15 min. It should not interfere with the routine surgical workflow
CR2. Surgical safety	Surgical sterility and safety are maintained for each system component
CR3. Tracking accuracy	TRE < 5 mm
CR4. System calibration	Interventional system calibration should be completed in less than 5 min and should achieve satisfactory spatial accuracy (RMS < 5 mm)
TR1. System compatibility	The system should be compatible with US systems that support sharing video data over a network connection as well as standard clinical devices
TR2. Intuitive display	The GUI of the system should be simple and easy to use displaying all required information with intuitive user interaction and switching between modules
TR3. Ultrasound visualization	US imaging should be displayed in real-time in conventional axial, sagittal, and coronal planes
TR4. Frame rate	Fast frame rate to provide real-time information and thus helps in fast surgical decision-making (minimum of 10 FPS)

## 7.3.2 System Components

The described system hardware included a commercial ultrasound system (Esaote, Italy), a standard laptop (Intel Core i7-6700HQ, 3.5GHz, 32GB Ram, 64-bit Windows 10), and NDI Vicra optical tracking camera (NDI Medical, USA). 3D Slicer (version 5.0.3) [9] and the PLUS Toolkit (version 2.8.0) [10] form the core of the navigation software. Figure 7.2 provides an overview of the hardware and software components of the proposed navigation system.

### 7.3.2.1 3D Slicer Software

3D Slicer (Fedorov et al., 2012), an open-source, cross-platform, and extensible software program for medical image computing and analysis, serves as the main neuronavigation software. Over two decades, 3D Slicer was implemented and developed through multi-institutional support from the National Institute of Health (NIH) as well as a worldwide developer community. Slicer 3D provides 2D, 3D, and 4D visualization capabilities for different imaging modalities (for example MRI, CT, and Ultrasound). Besides, 3D Slicer supports importing and exporting imaging data from multiple standard data formats including NIFTI, DICOM, and NRRD.

The 3D Slicer application follows a modular paradigm allowing the development of additional modules for feature-specific functionalities. Numerous modules providing a wide range of medical applications are distributed with the standard 3D Slicer software. Core modules are primarily categorized according to their function. One example is the Filtering module providing tools for basic pre-processing functions on medical images including arithmetic operations, Gaussian, bias field correction, and denoising filters.

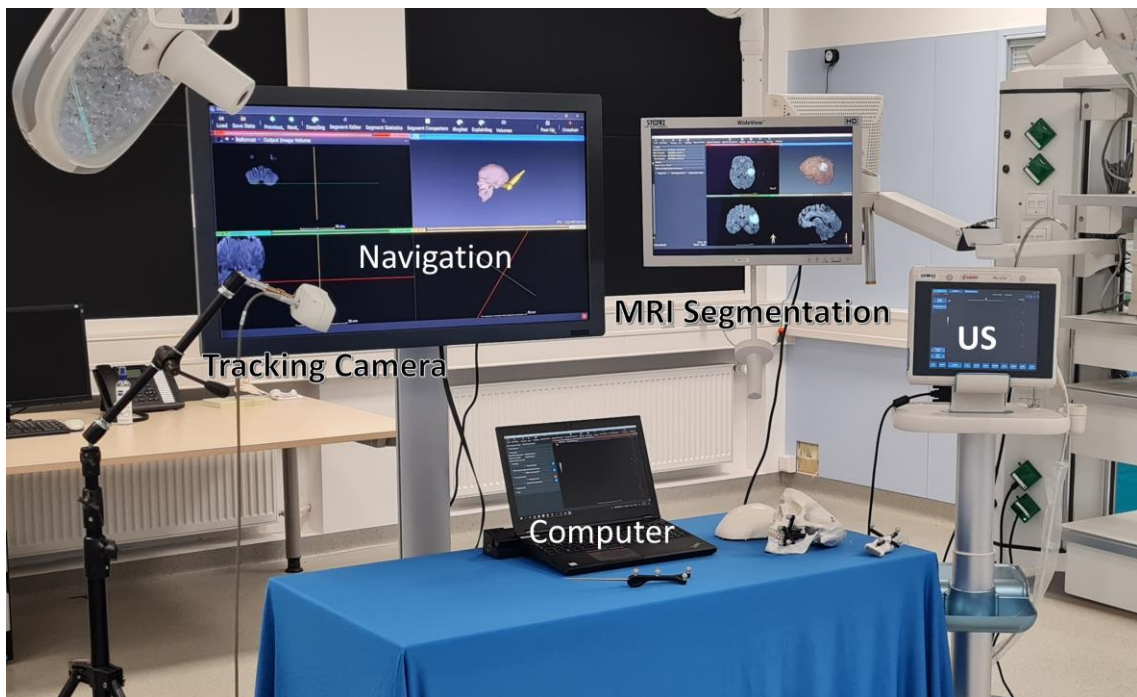


Figure 7.2: Overview of the proposed DeepIGN navigation system for assisting neurosurgery. NDI Vicra system is tracking the three optical markers: attached to the brain phantom, pointer tool (stylus), and the ultrasound transducer probe. Each marker is composed of four optically reflective spheres.

### 7.3.2.2 Hardware Interfacing Layer

PLUS Toolkit (Lasso et al., 2014) is a public open-source library for real-time data acquisition, streaming of imaging, and position tracking. PLUS was originally developed for ultrasound-guided interventions, and hence its name is the Public software Library for UltraSound imaging research Toolkit. PLUS implements an abstraction layer for low-level data acquisition from various commercial vendors and data forwarding in a standardized format through a network connection. For the purpose of the IGN system, PLUS is used mainly to provide data from the navigated ultrasound transducer as well as the NDI tracking device. Ultrasound data and tracking information for each tracked object, specifically a  $4 \times 4$  homogenous transformation matrix, is provided from PLUS Toolkit to 3D Slicer via OpenIGTLink (Tokuda et al., 2009). OpenIGTLink (Tokuda et al., 2009) is a standardized communication protocol among devices in operating rooms, which provides real-time data input-output. This configuration allowed a frame rate of approximately 15 FPS with a clinically acceptable latency.

### 7.3.2.3 SlicerIGT Calibration

SlicerIGT (Ungi et al., 2016) is an open-source extension of 3D which provides additional functionality for the rapid development of image-guided interventions. In particular, SlicerIGT provides a convenient user interface for the registration of tracked stylus and ultrasound probe to their virtual complements in 3D Slicer. The calibration procedure consists of a two-step process whereby the tracked instrument was pivoted and then rotated around a fixed point for 15 s each in a sequential manner<sup>1</sup>. After each movement, the root-mean-square error (RMSE) value is computed and the final value following the whole pivot calibration procedure should be less than 0.5 mm. This calibration process is performed once at the start of each operation using the pivot calibration module provided as part of SlicerIGT software. From this set of corresponding landmarks, SlicerIGT computes the transformation matrix from the reference to 3D Slicer RAS (namely Right, Anterior and Superior) coordinate system, ReferenceToRAS, as depicted in Fig. 7.3. Other transforms of the stylus and the US probe are provided by the PLUS Toolkit, appended in the scene graph below ReferenceToRAS (Fig. 7.3).

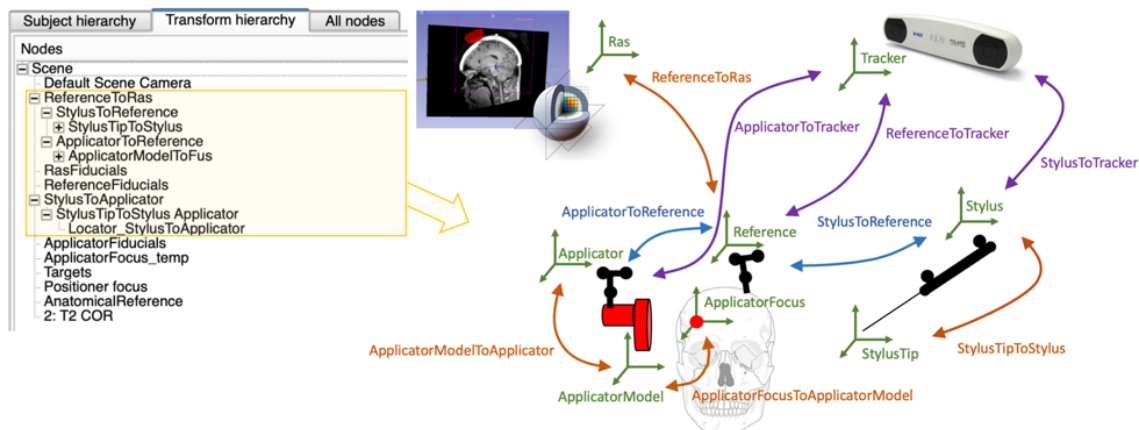


Figure 7.3: The hierarchy of transformations within the 3D slicer software (left), and their representations (right). *purple transforms* represent coordinate transformations from all tracked tools to the tracking camera, whereas PLUS Toolkit is responsible for generating the transforms of tracked tools to the reference (*blue transforms*). *Orange transforms* show fixed transforms that are the result of all registration procedures (Preiswerk, Brinker, McDannold, & Mariano, 2019).

### 7.3.2.4 DeepIGN Extensions

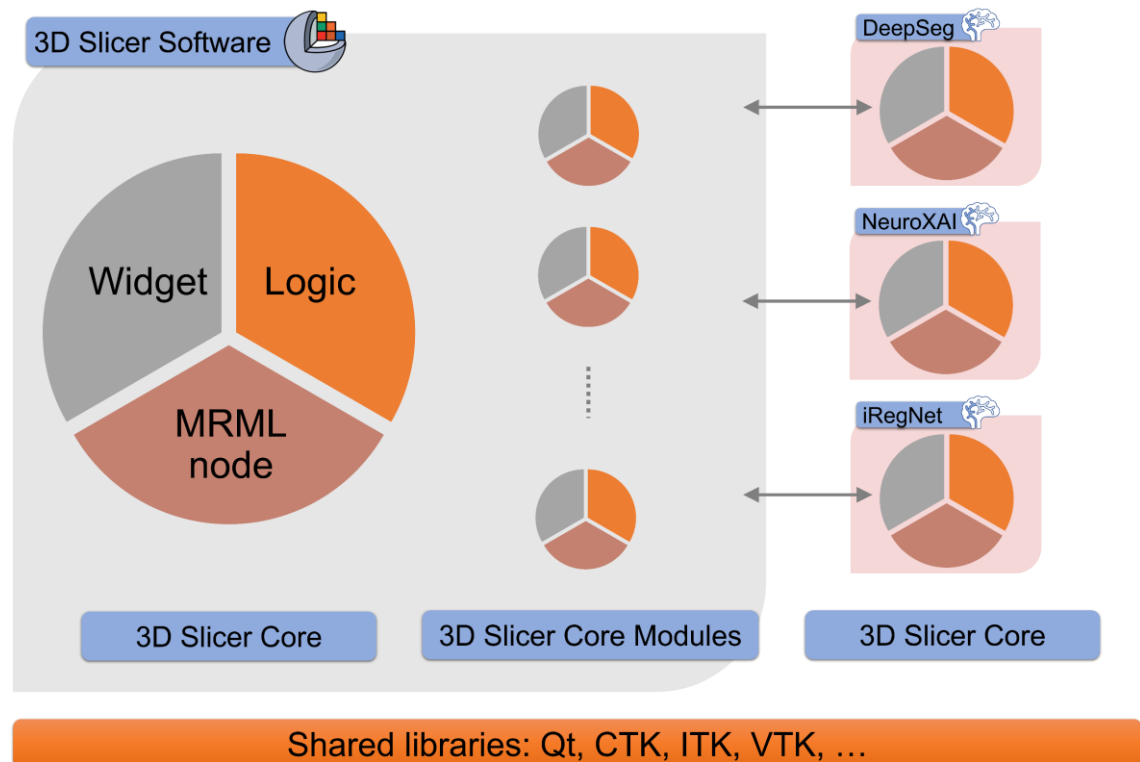
The modular design of 3D Slicer allows the development of additional modules for delivering further feature-specific functionalities. The core of Slicer is written in C++ programming language; however, the APIs of the core modules are also implemented in Python programming language allowing great support for rapid and simple prototyping applications. Similarly, DeepIGN follows a modular layered architecture, as presented in Fig. 7.1. DeepIGN shares common libraries with 3D Slicer such as the CTK and Qt libraries to create a custom user interface for deep learning-assisted neurosurgery. In each extension, additional libraries for image pro-

<sup>1</sup> Pivot Calibration, Tutorial U-11, SlicerIGT; <http://www.slicerigt.org/wp/user-tutorial/>

cessing, tracking, resampling, and registration are included separately according to the required function.

As discussed in Section 3.2, an IGN system consists of segmentation, registration, tracking, and visualization core functions. For each module, DeepIGN creates a custom end-user interface using 3D Slicer, as a slicelet<sup>2</sup> where extraneous GUI components are removed, providing a greatly simplified workflow ready for clinical use, as shown in Fig. 7.4. Figure 7.5 presents the segmentation component, also called DeepSeg Module, as a sample DeepIGN module that is implemented as an external 3D Slicer extension coded in Python as a scripted module. DeepSeg module extends the functionality of the 3D Slicer providing automatic segmentation of brain glioma using the approaches in Chapter 4. Similar to the DeepSeg module, the NeuroXAI module provides human-understandable attention maps to make the deep learning models transparent and thus increase medical users' trust and encourage their usage clinically. Furthermore, non-linear pre-operative MRI to iUS registration is performed by the iRegNet module to compensate for the brain shift during neurosurgical interventions.

Both DeepIGN and 3D Slicer are distributed under the BSD-style license that allows free use for both academic and commercial purposes without any restrictions. Source codes will be released on the project website<sup>3</sup>, where users, developers, data scientists, and clinical researchers are welcome to contribute and add their models and comments.



<sup>2</sup> Slicelet, 3D Slicer; <https://www.slicer.org/wiki/Documentation/Nightly/Developers/Slicelets/>

<sup>3</sup> DeepIGN project website; <https://www.github.com/razeinedin/DeepIGN/>

Figure 7.4: A high-level overview of the Proposed AI-powered IGN platform architecture and its relationship with 3D Slicer software and other libraries. Each extension has its dependencies for image processing, tracking information, resampling, or registration.



Figure 7.5: DeepSeg module during visualization of the resultant tumor boundaries. The GUI panel is shown on the left, where the user can specify inputs, output volumes, model parameters, and other 3D visualization options. Brain tumor segmentation results (*green*) of a sample HGG case from the BraTS 2021 challenge are presented on the right. The application layout shows the axial (*top middle*), sagittal (*bottom middle*), and coronal (*bottom right*), as well as 3D views (*top right*).

## 7.4 Results and Discussion

### 7.4.1 Pre-operative Application to BraTS Dataset

#### 7.4.1.1 Use Case Example

In order to demonstrate the capabilities of using the DeepSeg module, part of the DeepIGN system, with the 3D Slicer for addressing brain imaging segmentation problems, a sample HGG case from the BraTS dataset was employed.

The user must install the Slicer 3D program and then download the DeepSeg extension via the 3D Slicer Extension Manager<sup>4</sup>. After that, DeepSeg can be selected from the machine learning category in the modules list. The default parameter settings include two different pre-trained deep learning models based on the input MRI image modalities. The first model is from the previous work, DeepSeg (Section 4.3.1) which requires only the FLAIR MRI data as an input

<sup>4</sup> Slicer-DeepSeg module, DeepIGN project, <https://github.com/razeinedin/Slicer-DeepSeg>.



and automatically predicts the tumor region. The second model is the winning approach in the segmentation task of the MICCAI BraTS 2022 challenge, 3D DeepSeg (Section 4.3.2), which requires the four MRI modalities like the BraTS challenge: FLAIR, T1, T1Gd, and T2.

After the DeepSeg installation, the user can choose one model, specifies its input data, creates a new segmentation volume, and presses the “apply” button, as shown in Fig. 7.5. Then, an automatic pre-processing stage, including resampling, cropping, and registration, is applied before the resultant tumor region is predicted using the specified pre-trained DNNs. Finally, the segmented tumor is displayed in both Slicer 2- and 3D scenes as presented in Fig. 7.5.

### 7.4.1.2 Runtime Results

Table 7.1 lists the time runtime speed analysis for the two built-in neural networks, namely, DeepSeg and nnU-Net. The experiments were run on a computer with AMD Ryzen 2920X (32M Cache, 3.50 GHz) CPU, 32 GB RAM, and a single NVIDIA RTX 3060 GPU 12GB GDDR6. Slicer 3D version 4.11 was used with Python 3.6 running on 64-bit Ubuntu 18.04. Each measurement has been repeated 10 times and the average values are reported. Outstandingly, the DeepSeg module was able to detect and localize the tumor region in an average of 1.87 and 4.25 seconds using 3D DeepSeg (Section 4.3.2) and nnU-Net (Section 4.2) models, respectively. Compared to other cloud-based Slicer 3D extensions such as NVIDIA Clara and DeepInfer which take about 2 to 3 minutes for inference, the DeepSeg extension demonstrates significant performance advantages using pre-trained local models.

Table 7.2: Runtime measurement comparing the two integrated DNNs: DeepSeg and nnU-Net on both CPU and GPU implementations.

Process	DeepSeg (s)	nnU-Net (s)
Loading data and pre-processing	0.07	0.70
Building the DNN	0.58	0.81
Tumor prediction	1.20	2.35
Results visualization	0.02	0.02
Total time	1.87	4.25

## 7.4.2 Phantom Study

### 7.4.2.1 Experimental Setup

In this study, a tailored-made brain phantom based on tissue-mimicking material was used to simulate intra-operative navigation and evaluate the accuracy of the system’s tracking and registration capabilities. The detailed protocol followed to make the abstract model is provided in Appendix A. Figure 7.6 shows the use of the brain phantom to simulate the use of the DeepIGN navigation system in a clinical operating room. Public MRI Data from the BraTS 2021 dataset was used as the basis for manufacturing the phantom. The resultant brain phantom was placed in a classic skull model manufactured by 3B Scientific Corporation (Fig. 7.6 (a)). MRI

images were first used to register the phantom model data in the DeepIGN. Then, an iUS transducer connected to the Esaote MyLabSat ultrasound system captures real-time US data which is displayed in the DeepIGN display with its 3D reconstructed volumetric data using 3D Slicer-IGT.

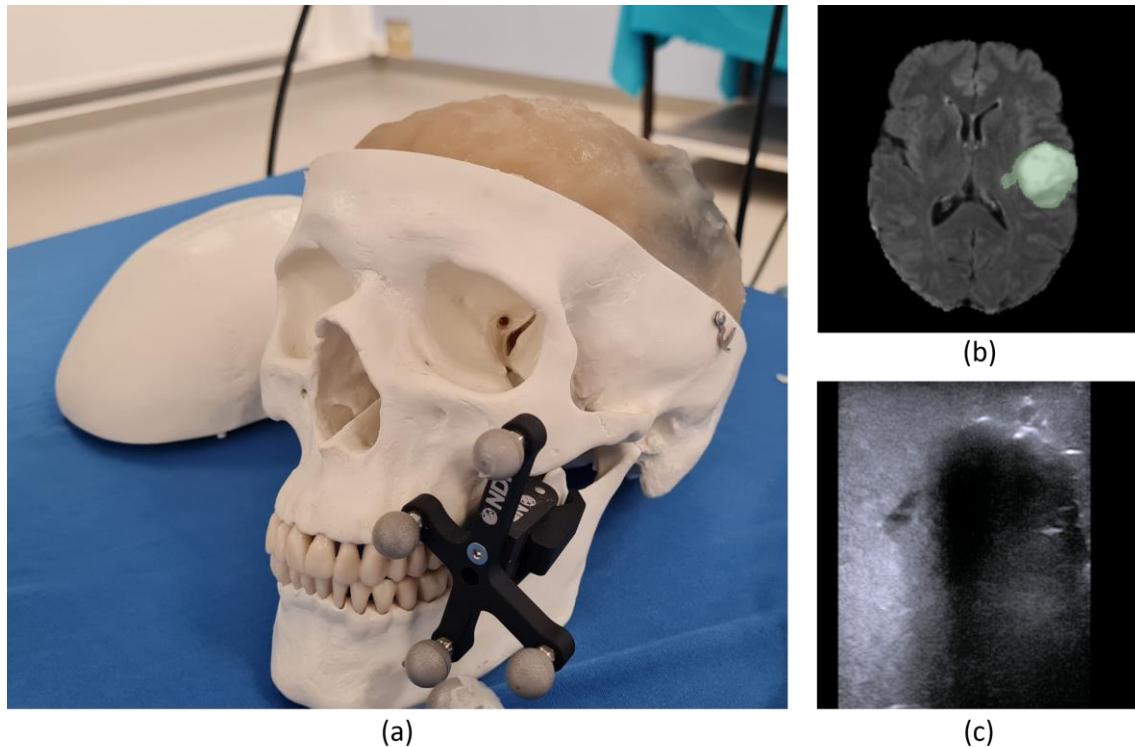


Figure 7.6: Validation of the DeepIGN system for use in guided neurosurgery in the simulated OR using the fabricated brain phantom. (a) custom brain phantom; (b) pre-operative FLAIR MRI with tumor highlighted in *green*; (c) live iUS image captured with the Esaote MyLabSat ultrasound linear transducer.

#### 7.4.2.2 DeepIGN Workflow

##### *Pre-operative preprocessing*

Pre-operative preprocessing refers to the preparation of images through data visualization, image processing, tumor segmentation, and the generation of a virtual 3D model of the patient. This process involves volume data, geometric models, AI methods, and spatial and temporal relations between them. The underlying building blocks for 3D computer graphics and visualization are provided by the Visualization Toolkit (VTK)<sup>5</sup>. Automatic segmentation of brain tumors in MRI volumes is provided by deep learning in the DeepSeg module (Figure 7.5) whereas semi-automatic segmentation can be done using a variety of tools in the segment editor module (Figure 7.7). Different file formats can be imported directly such as DICOM<sup>6</sup>, NIFTI<sup>7</sup>,

<sup>5</sup> The Visualization Toolkit; <http://www.vtk.org/>

<sup>6</sup> Digital Imaging and Communication in Medicine; <https://www.dicomstandard.org/>

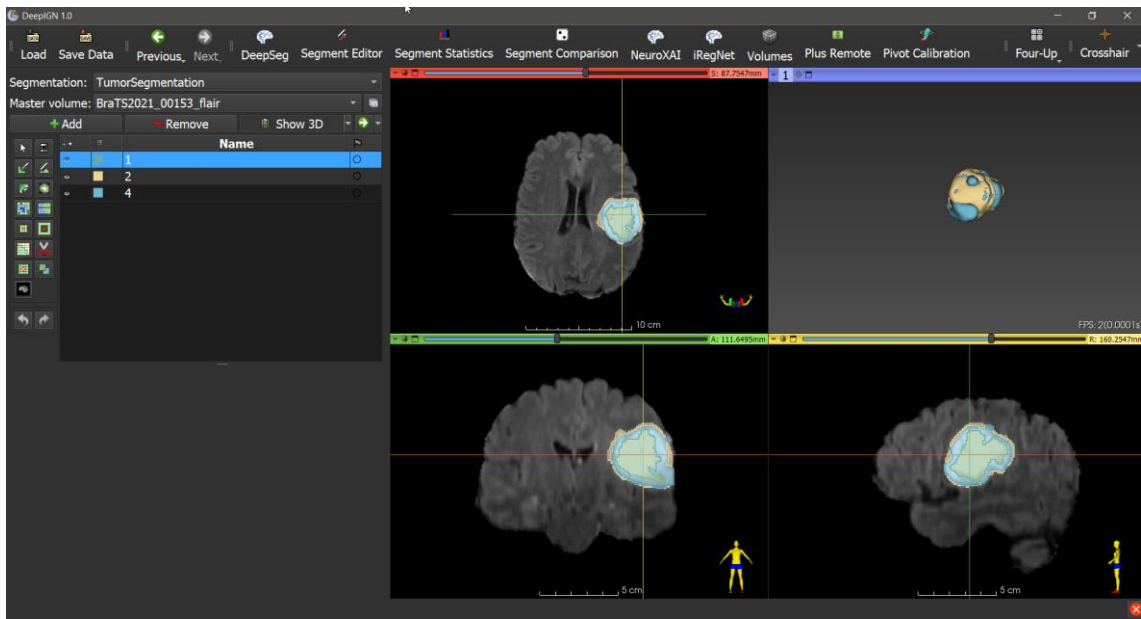


Figure 7.7: Interactive segmentation using the segment editor module as a built-in module of 3D Slicer. Tools such as threshold, paint, draw, erase, and grow from seed are available on the left (*module panel*).

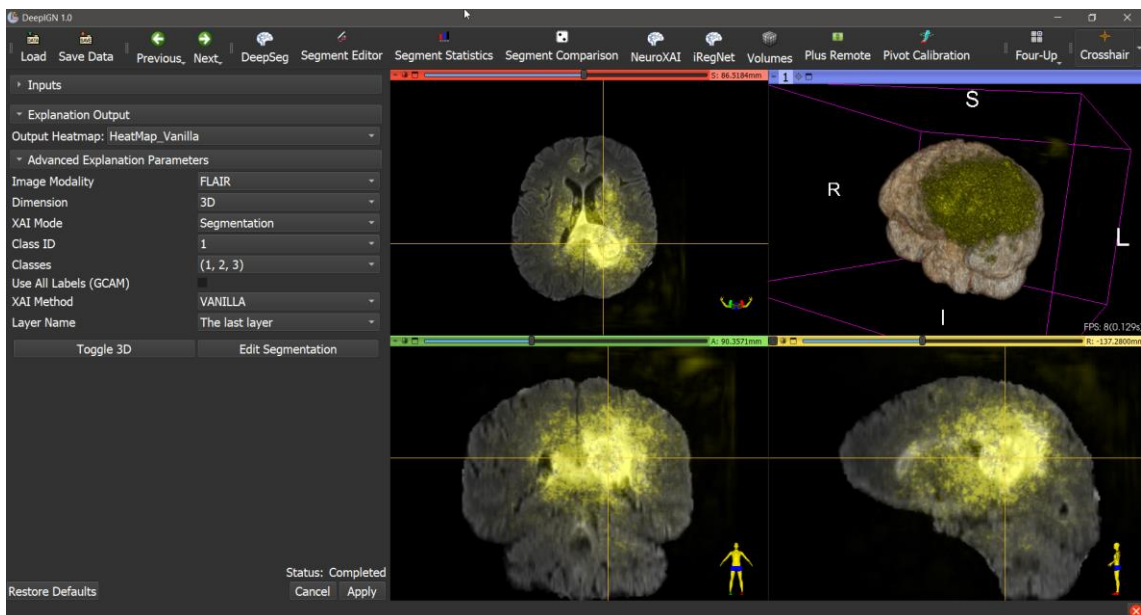


Figure 7.8: The NeuroXAI module provides sensitivity maps to explain the prediction of the tumor segmentation in brain MRI.

ITK<sup>8</sup>, and VTK file formats. Figure 7.7 the user interface of the Segment Editor module, which offers editing of overlapping segments, volume displaying in both 2D and 3D views, editing in

<sup>7</sup> Neuroimaging Informatics Technology Initiative; <https://nifti.nimh.nih.gov/>

<sup>8</sup> The Insight Toolkit; <http://www.itk.org/>

3D views, and creating segmentation by interpolating or extrapolating segmentation on a few slices. Explanation maps can be provided on-demand to justify the decision of the deep neural network via the NeuroXAI module (Figure 7.8).

### **Registration**

The term registration refers to the process of determining the spatial relation between two coordinate systems. In general, there are two general forms of registration implemented in the DeepIGN system. First, image-to-image registration where one set of images (moving) is transformed into the coordinate space of another set of images (fixed). iRegNet module allows fast and accurate MRI-iUS registration during neurosurgical interventions using an AI-based approach (Figure 7.9). Second, image-to-patient registration transforms the virtual patient model into the physical patient, thus enabling image navigation using physical tools. This type is provided using specific methods of VTK, in which the same point set is defined in the two spaces to be registered, and then finding the relation between them using the least squares fit.

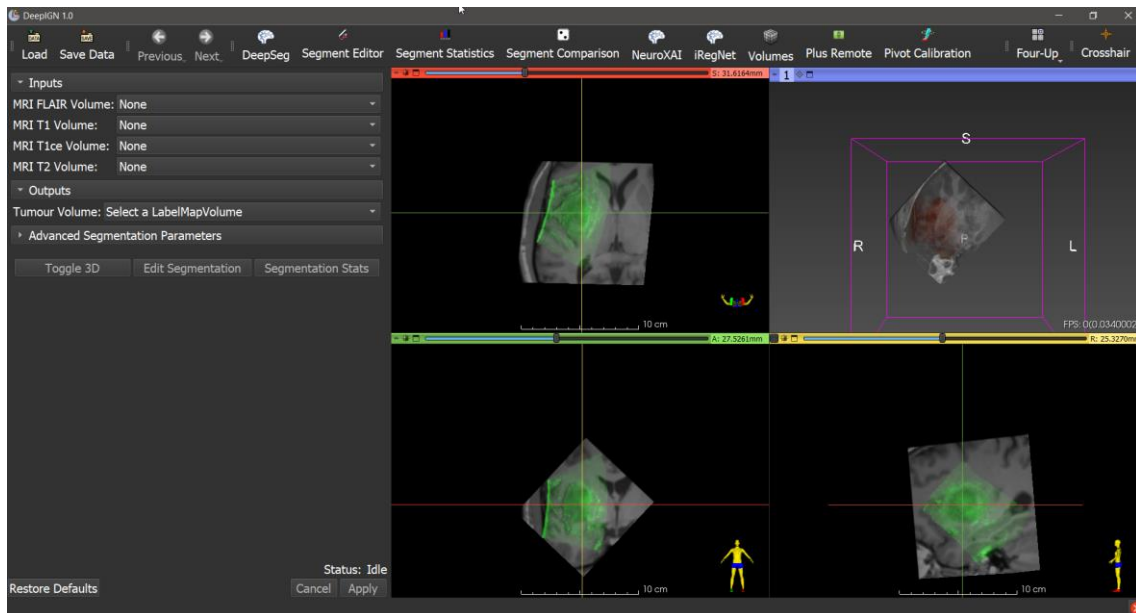


Figure 7.9: Automatic image-to-image registration using the iRegNet module. MRI and iUS scans are utilized as moving and fixed images, respectively, since the main goal is to reflect the brain shift in the pre-operative MRI data.

### **Position Tracking**

Position tracking technology is a key component of navigation, in which a dynamic visualization of instruments and tools is achieved at the correct location in the 3D guidance display. Tracking Tools are physical devices that are used by the operator to interact with the patient. Surgical instruments, stylus pointers, and US probes are examples of tracking tools. All IGSTK devices, namely the optical tracking system (Polaris, NDI, Canada) and the ultrasound imaging (MyLabSat, Esaote, Italy), are connected via the PLUS toolkit. The instruments to be tracked are equipped with special sensors with attached reflective markers, and the tracking system

provides realtime information about their position and orientation. For quality control reasons, all tracked devices are calibrated before each procedure using the SlicerIGT and the configurations are stored in XML files. Figure 7.10 shows an example of the stylus probe calibration using fiducial landmarks registration as a two-step approach: First, the transform between a tracker tool attached to the probe and the stylus tip is estimated. Second, the registration of the device's physical coordinates to the simulated 3D model is achieved.

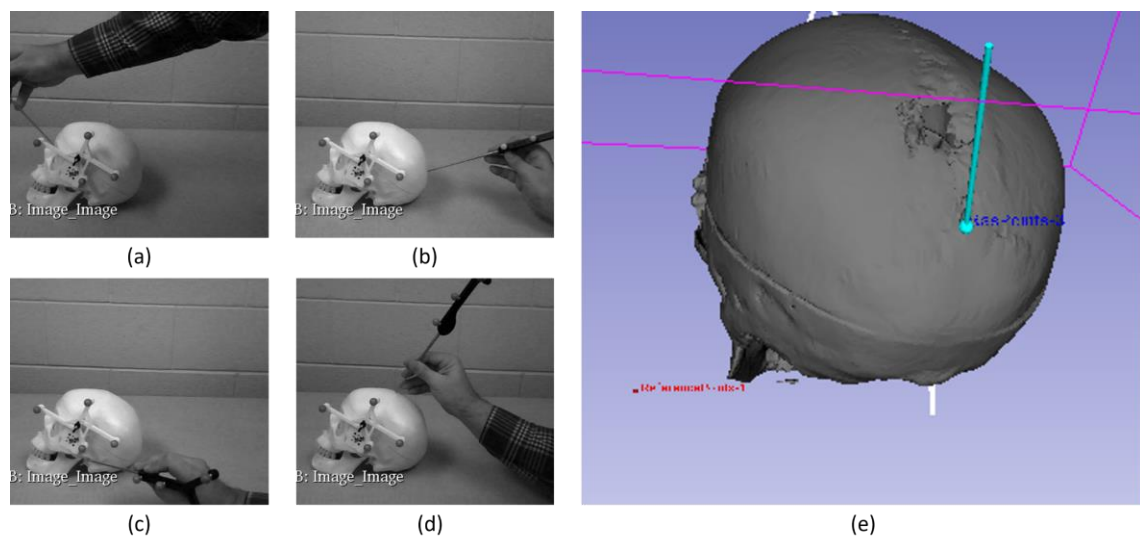


Figure 7.10: Illustration of the stylus calibration using the VTK landmark registration method as part of the SlicerIGT extension. (a) – (d) show different fiducial landmarks, whereas the resultant registered stylus pointer is displayed in *cyan* (e)<sup>9</sup>.

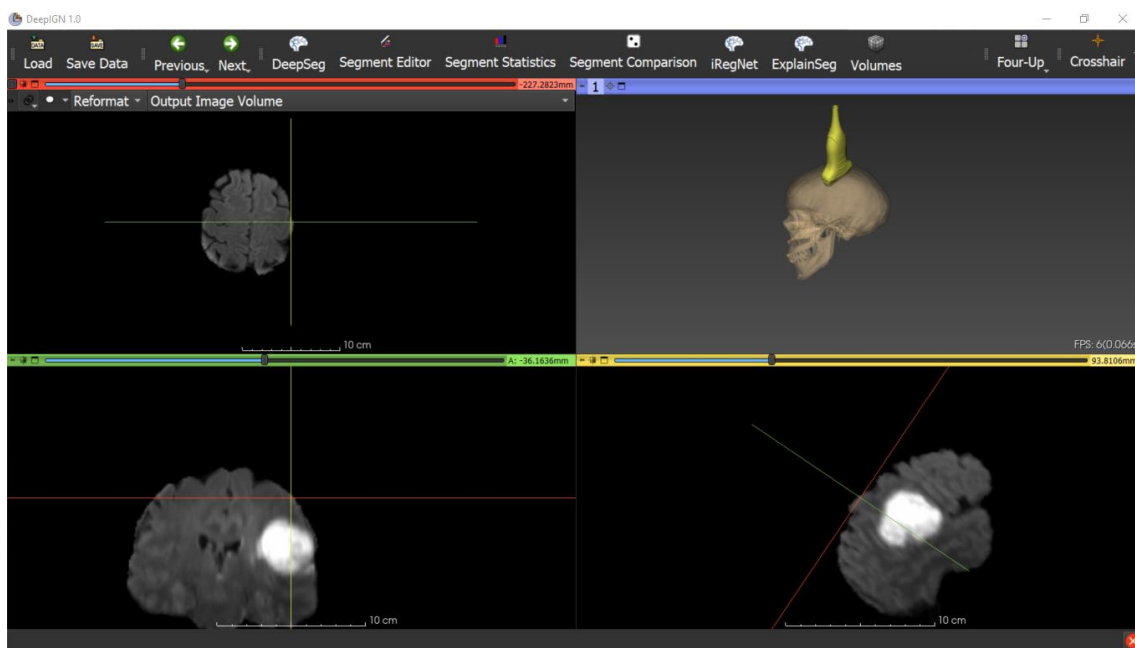
### ***Image Acquisition***

The clinician navigates through the patient's anatomy using tools, and the views show the position inside the virtual patient. For each phase of the brain surgery, the user has the option to choose between a number of pre-defined layouts, which may contain 2D slices, 3D patient models, and table views for statistics. During the intervention, the acquisition of updated image data may be necessary due to patient movement, anatomical changes, or lack of pre-operative data. DeepIGN can use image modalities such as the US to provide this updated information. With a properly calibrated and tracked US probe, new 3D volumes can be created. Figure 7.11 display a live iUS visualization using the DeepIGN system.

<sup>9</sup> Landmark registration, Tutorial U-12, SlicerIGT; <http://www.slicerigt.org/wp/user-tutorial/>



(a)



(b)

Figure 7.11: Illustration of the live US acquisition as part of the DeepIGN system. (a) provides US view with recording and reconstruction of 3D US images. (b) provide a view of the tracked US probe and its location with respect to the pre-operative MRI.

### 7.4.2.3 System Accuracy

To enable research and clinical trials in the operating room, overall clinical accuracy, robustness, and responsiveness of the system and the anatomical representation of relevant structures should meet standard levels. The proposed IGN system shares similar components with other computer-aided interventional systems including image segmentation algorithms, multimodal registration, data visualization, and tracking system. Segmentation errors were measured using DSC and HD95 between the predicted tumor and ground truth volumes in 3D. The proposed DeepSeg module has been successfully tested and evaluated on-line based on MRI datasets of the BraTS 2022 challenge, including more than 1200 cases. It is worth noting that the ensemble method is the winner of this MICCAI challenge among more than one thousand teams worldwide achieving an average DSC of 0.9270 and HD95 of 3.60 mm for the whole tumor region.

To evaluate the registration module, experiments were conducted using 36 patients from two publicly available multi-location databases: BITE and RESECT. The TRE was calculated between the center of the fitted spheres in 3D. Quantitatively, iRegNet reduced the mean landmark errors from pre-registration values of  $(4.18 \pm 1.84$  and  $5.35 \pm 4.19$  mm) to the lowest value of  $(1.47 \pm 0.61$  and  $0.84 \pm 0.16$  mm) for the BITE and RESECT datasets, respectively.

The integrated navigation system was also tested in a pre-clinical operating room (Figure 7.12). A volunteer computer-assisted medicine student correctly assembled the system hardware and completed the initial set-up in about 11 min (CR1). The computing device can be put on a separate sterilized table and connected to other operating devices via a network connection (CR2). Calibration and position tracking of rigid surgical tools (stylus for simulation, and iUS transducer) has achieved an accuracy of less than 0.5 mm (CR3), and the process was completed in less than one minute (CR4).

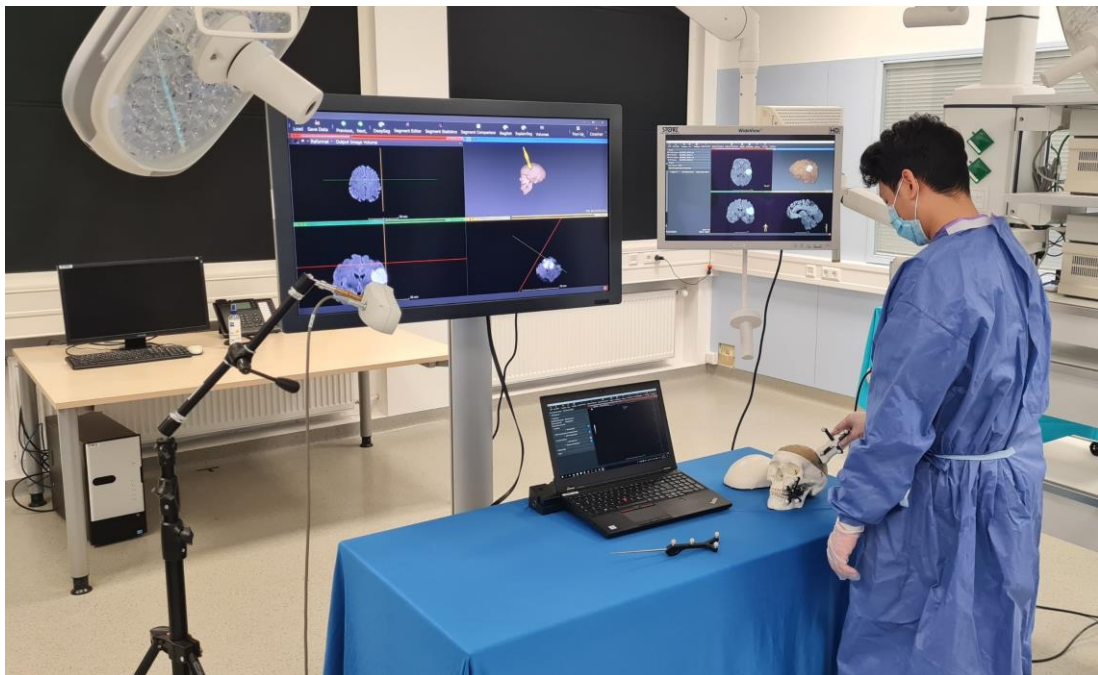


Figure 7.12: Intra-operative simulation of 3D Ultrasound reconstruction using patient data and a patient-specific phantom model.

The system is assumed to be compatible with ultrasound systems that support sharing video data over a network connection or using an image grabber device (TR1). Further, a pre-clinical evaluation of the system's usability was undertaken independently by an experienced neurosurgeon who considered the system to be highly useful, with an intuitive display and clinically acceptable accuracy (TR2). Frame rate imaging of an average of 19 FPS was achieved, as per the target requirement (TR4). The system provides live US imaging (Figure 7.11 (a)) fulfilling the requirement of (TR3). Nevertheless, the available US imaging system (MyLabSat, Esaote, Italy) does not provide 3D US imaging and future work will include testing the proposed system with neurosurgical navigation US systems (TR3). Overall, the proposed prototype research system met all of the minimum requirements specified in Table 7.1, but further refinement in collaboration with commercial partners will be employed.

### 7.4.3 Comparison with IGN Platforms

As discussed in Section 3.2.2, two categories can be distinguished including commercial systems and open-source research platforms. Table 7.2 summarizes the differences between existing systems for image-guided interventions as well as the proposed DeepIGN platform. Commercial IGN systems, for example, Brainlab, Medtronic, and imFusion, are built for routine clinical use and are widely used for surgical planning and guidance worldwide. The other group comprises open research systems that are built specifically to be used as intra-operative navigation systems and are designed with a focus on neurosurgery. Another major difference is that a research system should be at the forefront of the development and has a rapid introduction of new features into the clinic compared to commercial systems which emphasize stability and ease of use.

The proposed DeepIGN system belongs to the second group of open-source guidance systems, which allows components to be improved upon or replaced to accommodate research projects while replicating the common functionality of a commercial IGNS. Two distinguishing features of DeepIGN are its fully automatic deep learning solution for brain tumor segmentation in MRI images as well as its capabilities of explaining neural network prediction using XAI methods. Besides the effective visualization of brain shift intra-operatively using iUS navigation, DeepIGN provides an automatic methodology for the correction of deformed brain images with the help of recent advances in DL. Since the backend software of the proposed IGN system is 3D Slicer, interfaces can be overwritten with a simple Python script, which can greatly ease the development process.

Another important aspect to consider when choosing an open-source research system is the level of support by the community of developers and users. This aspect is fully fulfilled with 3D Slicer as it has one of the largest and most active community that applied it in a variety of clinical and pre-clinical research applications (Fedorov et al., 2012). For instance, a recent study indicated the possibility of using 3D Slicer in pre-operative planning of brain lesions surgery (L. Zhou et al., 2022) although cranial planning has been a unique feature of Brainlab so far. Similarly, some research studies have investigated the use of augmented reality to assist neurosurgical procedures such as (Dominguez-Velasco, Perez-Lomeli, Padilla-Castaneda, Tello-Mata, & Alcocer-Barradas, 2022; Haouchine et al., 2022; Yavas, Caliskan, & Cagli, 2021).



Table 7.3: Comparative analysis of the proposed DeepIGN platform against the available image guidance systems.

Features	NifTK	CustusX	IBIS	SlicerIGT	ImFusion	Brainlab	Medtronic	DeepIGN
2D/ 3D Visualization	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Customizable User Interface	No	No	No	Yes	No	Yes	Yes	Yes
Cranial Planning	No	No	No	No	No	Yes	Yes	No
Interactive Manual Segmentation	Yes	No	No	No	Yes	Yes	Yes	Yes
<b>Automatic Tumor Segmentation</b>	No	No	No	No	No	No	No	Yes
Patient Registration	No	Yes	Yes	Yes	No	Yes	Yes	Yes
MRI-MRI Registration	Yes	Yes	No	No	Yes	No	No	Yes
MRI-US Registration	No	Yes	Yes	No	Yes	No	No	Yes
iUS Navigation	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Brain shift Visualization	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Brain shift Correction	No	No	Yes	No	No	No	No	Yes
Instrument Tracking	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Augmented Reality	No	No	Yes	No	No	No	No	No
Deep Learning	Yes	No	No	No	Yes	No	No	Yes
<b>Explainable AI*</b>	No	No	No	No	No	No	No	Yes
FDA Approved	No	No	No	No	No	Yes	Yes	No
Open source	Yes	Yes	Yes	Yes	No	No	No	Yes

\* Bold values highlight distinctive features of the proposed DeepIGN system

## 7.5 Summary

An open-source integrated intra-operative navigation toolkit, called DeepIGN, was developed and tailored to neurosurgery. With its simple GUI, easy installation guide, and range of available modules, DeepIGN gives users and clinical researchers the opportunity to utilize advanced deep learning approaches within brain cancer research without the need for advanced programming skills or detailed knowledge of several software packages. DeepIGN has a modular architecture with the ability to incorporate the following functions: (1) multimodal 3D pre-operative MRI and iUS volume visualizations of the brain tumor and surrounding healthy structures, (2) pre-operative brain tumor segmentation in MRI images, (3) neuronavigation based on the pre-operative MRI volumes, (4) intra-operative guidance with the help of iUS volumes, (5) brain shift corrections using the registration of pre-operative MRI to iUS scanning, (6) real-time tracking of the surgical instruments and the iUS transducer. Except for the commercial iUS system, all other system's software components including the DeepIGN application, 3D Slicer platform, PLUS toolkit, and SlicerIGT plugin software are open source under the BSD license. Nonetheless, the proposed navigation system can be integrated with any type of clinical hardware for tracking surgical tools and iUS streaming.

# 8 Conclusions and Outlook

---

## Contents

<b>8.1</b>	<b>Conclusions</b>	<b>122</b>
<b>8.2</b>	<b>Outlook</b>	<b>124</b>
<b>8.2.1</b>	<b>Multimodal Registration</b>	<b>124</b>
<b>8.2.2</b>	<b>Need for large datasets</b>	<b>125</b>
<b>8.2.3</b>	<b>Mixed Reality</b>	<b>125</b>
<b>8.2.4</b>	<b>Explainable Deep Models</b>	<b>126</b>

This chapter summarizes the main results of this thesis and provides plans and an outlook for future research.

## 8.1 Conclusions

In this thesis, an image-guided system is proposed for assisting neurosurgical interventions using deep learning methods. Hence, an efficient research pipeline (see Figure 1.2) is defined to address the four necessary modules including an automatic and accurate definition of the brain tumor boundaries, registration of pre-operative MRI to iUS images for brain shift compensation, interpreting the developed DNNs using XAI, and the development of an intuitive neuronavigational display in the operating room. The main part of this thesis, therefore, comprises four chapters dedicated to each of the following research questions.

*How to correctly delineate brain tumor boundaries from adjacent healthy structures using automatic deep-learning models?*

Accurate segmentation of brain tumor boundaries and their neighboring healthy structures in multimodal MRI is the basis for successful neurosurgery. In particular, the precise localization of pathological targets (lesions) within the brain anatomy is a major issue in neurosurgery. This challenge is related to the difficulty in visually delineating these pathological structures from healthy tissue. Typically, manual segmentation is the gold standard and is usually performed by an expert: often a radiologist or a specialized clinician. While it has the advantage of incorporating expert knowledge, the drawbacks of this method are the time-consuming nature of the procedure and the proneness of inter-observer variability (refer to Section 3.3.1).

Different novel approaches are proposed in Chapter 4 as the first step of the research pipeline. Initially, the 2D DeepSeg framework is proposed for automatic segmentation of the brain glioma in MR FLAIR images based on seven state-of-the-art CNN models. By incorporating multimodal MRI inputs, utilizing region-based training, applying a larger batch size of 5, post-processing stage, and on-the-fly data augmentation, the 3D DeepSeg model achieved a better DSC score of 0.91 for all tumor sub-regions, respectively. Utilizing the STAPLE ensemble strategy contributed further to the improvement of the global performance of the deep networks with DSC scores of 0.84, 0.88, and 0.93 for the ET, TC, and WT glioma sub-regions, respectively. It is worth noting that the proposed Ensemble 2022 model is the winner of the BraTS 2022 segmentation challenge and on two different datasets: the BraTS test dataset and the Sub-Saharan Africa population (SSA) dataset as proof of its generality to unseen data.

*Which algorithmic methods should be applied to use iUS data to correct the shifted pre-operative MRI images?*

Intra-operative brain deformation in response to surgical manipulation, swelling, gravity, and anesthesia, called brain shift, reduces the utility of pre-operative imaging data for neurosurgical guidance. Pre-operative MRI presents perfect imaging modalities for neurosurgical guidance, due to its excellent visualization of the brain tissues, its sub-structure, and surrounding tissues. However, MRI suffers from long scan times and requires special precautions in the operating room in order not to affect the MRI scanning quality or cause any image artifacts. On the other hand, the US modality offers portable, low-cost, non-ionizing radiation, fast scan times ranging from seconds to minutes, easy to use, good visualization of the internal soft tissue and structures, and could achieve a spatial resolution within 0.20 mm. The disadvantages of intervention-

al US are the difficulty in imaging through bone and image inconsistency because of inter-operator variability.

Thereby, the iRegNet framework is presented in Chapter 5 as an automatic, fast, and accurate framework for the alignment of pre-interventional MRI to iUS to correct tissue shift and enable guided surgery. Pre-operative MRI (as a moving image) and iUS (as a fixed image) are first appended to the convolutional neural network, after which a non-rigid transformation field is estimated. The MRI image is then transformed using the output displacement field to the iUS coordinate system. Extensive quantitative experiments of six iRegNet configurations demonstrated the robustness and generality of the proposed method (refer to Section 5.5). Statistically, non-rigid registration results showed that TRE values decreased from initial values of  $(4.18 \pm 1.84 \text{ mm}$  and  $5.35 \pm 4.19 \text{ mm})$  to reduced values of  $(1.47 \pm 0.61$  and  $0.84 \pm 0.16 \text{ mm})$  outperforming other state-of-the-art methods on the BITE and RESECT datasets, respectively, as illustrated in Tables 5.2 and 5.3. Moreover, qualitative evaluation by two experienced neurosurgeons shows that registered MRI-iUS pairs have significantly improved over the original alignment.

*Can Explainable AI help to make deep learning networks transparent and understand the reason behind their predictions toward gaining human trust?*

AI, in particular DL, has achieved remarkable results for medical image analysis in several applications. Yet the lack of human-like explanations of such systems is considered the principal restriction before utilizing these methods in clinical practice. XAI provides a human-explainable and interpretable description of the “black-box” nature of DL, as illustrated in Chapter 6. An effective XAI diagnosis generator, namely NeuroXAI (refer to Section 6.3.1), has been developed to extract 3D explanations from CNN models of brain gliomas. By providing visual justification maps, NeuroXAI can help make deep learning models transparent and thus increase the trust of medical experts

NeuroXAI has been applied to two applications of the most widely investigated problems in brain imaging analysis, i.e. image classification and segmentation using MRI. Visual attention maps of multiple XAI methods have been generated and compared for both applications, which could help to provide transparency about the performance of deep learning systems. NeuroXAI helps to understand the prediction process of 3D CNN networks for brain glioma using human-understandable explanations. Since visual pixel-based representations are not enough to give meaningful information, extensive experiments were conducted to provide interpretability by evaluating their clinical significance. Results revealed that the investigated deep learning models behave in a logical human-like manner and can improve the analytical process of the MRI images systematically. Due to its open architecture, ease of implementation, and scalability to new XAI methods, NeuroXAI could be utilized to assist medical professionals in the detection and diagnosis of brain tumors.

*Which toolkits are required to develop a meaningful display for neuronavigation in a simple and intuitive fashion in the operating room?*

Neuronavigation has become an indispensable tool in the management of brain tumors, allowing for smaller, more precisely positioned incisions and the accurate localization of tumors and

surrounding structural and functional regions which may be at risk during surgery. However, dynamic changes in the surgical environment regularly occur during the intervention, making the surgeon face continuously challenging intra-operative circumstances, as illustrated in Section 3.3.2. Consequently, image guidance using computerized navigation based on pre-operative imaging data does not provide enough information to complete the surgical guidance, and the demand for intra-operative acquired images raises to provide online information as the surgery progresses.

Chapter 7 gives a detailed outline of the proposed DeepIGN software including several main components organized in a particular way for neuronavigation procedures. A commercial Esaote system is utilized to provide real-time live iUS data, while accurate tracking information is transmitted by the PLUS toolkit to allow for the accurate definition of the tumor and surgical tools positions relative to patient anatomy (Section 7.3). Visualization of multimodal imaging data in a meaningful display simply and intuitively in the operating room is another feature of DeepIGN provided by 3D Slicer software. Position tracking of the Stylus and iUS transducer was calibrated at the start of each intervention using the SlicerIGT software as an integrated part of the DeepIGN.

In conclusion, the proposed DeepIGN system is an effective open toolkit for assisting neurosurgery which greatly facilitates the surgeon's understanding of the relationship between normal and pathological anatomy. Among its benefits in pre-operative planning and visualization in neurosurgery is the automatic segmentation of glioma in pre-operative Brain MRI, displaying the position and exact location of the lesion in 2D and 3D views, and extracting human-understandable attention maps to interpret the deep learning networks and explain their decision-making procedure. Furthermore, DeepIGN has been proven to be highly valuable at volumetrically representing iUS data and using them to compensate for brain deformation during neurosurgical interventions.

## 8.2 Outlook

The main focus of this thesis was set on the development of multimodal image-guided methods for brain surgery guidance using the recent advances in deep learning. The main part of this thesis therefore comprised four chapters dedicated to each of the IGN tasks of glioma segmentation in Brain MRI, Multimodal MRI to iUS registration, explainability of DNNs, and DeepIGN system integration. Thereby, the evaluation of the results depicts further capabilities of the proposed approaches (refer to Table 7.2), which are discussed within the following subsections.

### 8.2.1 Multimodal Registration

Further research work should be conducted to investigate the optimal cropping radius for MRI images to minimize the missing data as possible. Automating this procedure will contribute towards rendering iRegNet an end-to-end pipeline. Furthermore, employing the recent learning-based approaches and registration advances would lead to improving the overall registration performance, specifically atrous spatial pyramid pooling (ASPP) (Chen, Papandreou, Kokkinos,

Murphy, & Yuille, 2018), resection cavities segmentation (Canalini, Klein, Miller, & Kikinis, 2020), multi-resolution (Nan, Tennant, Rubin, & Ray, 2020), spatial attention (Y. Zhu et al., 2021), and locally adaptive regularization (Dalca, Balakrishnan, Guttag, & Sabuncu, 2019), and unsupervised learning (Balakrishnan et al., 2019).

Another limitation of the multimodal registration module is concerning the training datasets. In general, the quantity and diversity of training data are limited to particular areas and recording conditions, thereby impairing the generality and transferability of learned models. While the proposed registration pipeline exhibits good accuracy results on unseen data, limitations of its transferability become obviously recognizable, especially in the case of a real intra-operative application to brain phantom.

## 8.2.2 Need for large datasets

Despite the great progress, deep learning-based algorithms have made in solving various medical segmentation problems, their performance is highly dependent on the availability of large, high-quality expert-annotated datasets (Fu et al., 2021; Kim et al., 2019). Another challenge facing researchers, particularly in the field of medical imaging, is scarce annotations where only a limited number of real clinical images are available for the training and validation of new image analysis algorithms (Tajbakhsh et al., 2020). In the presence of these shortcomings, even the most advanced learning models may fail to generalize to real-world applications.

Accordingly, researchers from the medical imaging community have actively sought solutions to these problems of preparing both medical imaging data and their annotations (Tajbakhsh et al., 2020). One idea was to invest effort in developing new methods for the expansion and enhancement of existing datasets would reduce the annotation time and operational costs (Lundervold & Lundervold, 2019). Synthetic data generation was also proposed using generative adversarial networks (GANs) to actively enlarge the training dataset through effective algorithms (Donnez et al., 2021). The combination of synthetic and real datasets could considerably increase the performance of neural networks under the condition that a sufficient quantity of original data is accessible. This assumption may be addressed in future studies.

## 8.2.3 Mixed Reality

Mixed reality allows for the superimposition of virtual objects with the real world. In medicine, especially in complex surgical interventions, the availability of patient- and pathology-related information can be the decisive factor for the success of the intervention. There are first approaches to use mixed reality in surgical interventions (Guha et al., 2017), e.g. the overlay of tumor boundaries in a surgical microscope (Drouin et al., 2017; Mascitelli et al., 2018) or the use of Microsoft HoloLens for open procedures (Galati et al., 2020). The question is still unknown, which information shall be displayed using which specific visual representation at what point in time to guarantee that the surgeons are not distracted by the mixed reality overlay or the interaction with the system?

### **8.2.4 Explainable Deep Models**

For future work, the generalization architecture of the proposed TransXAI can be extended by adding new deep-learning models. Further studies should explore feeding on-demand 3D concept activation maps back to the neural network as on-demand deep supervision. That will provide additional guidance to the network and thus enhance the overall accuracy of assisting the surgeons during interventional procedures. Moreover, future work will be focused on the quantitative evaluation of XAI methods to assess the quality of the generated sensitivity maps and study their relationship with the deep learning accuracy metrics with additional experiments on multi-modal MRI-guided neurosurgery.



# Appendix A. Brain Phantom Preparation

The use of patient-specific anatomical phantoms with appropriate tissue-mimicking properties is often a critical part of the testing environment and can increase the confidence of clinicians who are learning to use a new image-guiding system (Pacioni et al., 2014). In the following, the steps involved in the fabrication of a durable, patient-specific brain phantom are described. The resultant phantom was utilized for the test and evaluation of the proposed DeepIGN navigation system for assisting brain tumor surgery. Appendix A reproduces parts of the following publication (Earle, Portu, & DeVos, 2016; Mackle et al., 2020).

## A.1 Design Requirements

The brain tumor phantom must possess several key characteristics in order to be suitable for the validation of the iUS neuronavigation application, as follows:

- The phantom should be made of non-toxic materials for safe clinical use.
- Has realistic imaging properties, i.e. suitable for US attenuation.
- Possess mechanical properties similar to human tissue to simulate brain shifts.
- To be based on real patient data to allow for pre-clinical simulation.
- Made of durable materials to be used repeatedly for the testing and validation processes.

## A.2 Data

The publicly available data of the BraTS 2019 challenge was utilized that describes a multi-institutional collection of pre-operative multimodal MRI brain scans of 336 subjects diagnosed with high-grade (HGG) and lower-grade gliomas (LGG) (Spyridon Bakas et al., 2018; Menze et al., 2015). Details of the dataset and pre-processing are described in Section (4.4.1). The phantom described in this study was created using pre-operative T1Gd/ FLAIR MRI data from an HGG patient (#00153) and iUS live data from the Esaote MyLabSat ultrasound system.

## A.3 Material

Polyvinyl alcohol cryogel (PVA-c) is a popular choice of tissue-mimicking material thanks to its acoustic and mechanical properties, which can be tuned by varying its freeze-thaw cycles. However, their mechanical strength and durability are limited, and they do not meet the required design requirements for this application. Similarly, agar and gelatin are aqueous materials used widely as tissue-mimicking materials for organ phantoms (Fromageau et al., 2007). Agar is also durable and has a high degree of elasticity allowing for repetitive use without the need for refrigeration.

## A.4 Parts Segmentations

1. Install the 3D Slicer software<sup>1</sup> to segment the patient data.
2. Tumor segmentation
  1. Open DeepIGN software and load the four MRI modalities provided by the BraTS dataset.
  2. The **‘DeepSeg’** module will be loaded, and a new segmentation named ‘Tumor’ is automatically created.
  3. Change the tumor type to ‘Whole Tumor’ from the advanced segmentation parameters.
  4. Press the **‘Apply’** button to perform automatic segmentation of the brain tumor.
  5. Open the **‘Segment Editor’** module, and correct the segmentation if necessary, using the **‘Paint’**, **‘Draw’**, and **‘Erase’** functions.
  6. Apply the **‘Smoothing’** function (suggested median 2.00 mm 5 x 5 x 3 pixels).
  7. Save ‘Tumor’ segmentations.
  8. Open the **‘Segmentations’** module and export ‘Tumor’ as STL files.
3. Brain tissue segmentation
  1. Open the **‘Segment Editor’** module and create a new segmentation named ‘Brain’.
  2. Use the **‘Threshold’** function to highlight the brain. Make sure to set the threshold value from 1.0 to exclude the empty background pixels.
  3. Use the **‘Smoothing’** function (suggested median 2.00 mm, 5 × 5 × 3 pixels).
  4. Use the **‘Scissors’** function to remove any unwanted or erroneous segmentations.
  5. Save ‘Brain’ segmentations.
  6. Open the **‘Segmentations’** module and export ‘Brain’ as STL files.

## A.5 3D Printing of Molds

1. Create the brain and tumor molds
  1. Import the STL file ‘Brain’ into computer-aided design (CAD) software<sup>2</sup>.
  2. Click the **‘MESH’** tab and use the **‘Reduce’** tool to reduce the size of the model so that it can be handled by the program – the aim is to reduce the size as much as possible, whilst still retaining all the detail necessary.
  3. Use the **‘Covert Mesh’** tool to convert the imported mesh to a body that can be manipulated. If this action cannot be completed, the mesh was not reduced enough in the last step.
  4. Click **‘Create’** in the **‘SOLID’** tab then **‘Box’** and draw a box around the tumor. Select to create this as a **‘New Body’** and rotate the view to ensure the box completely encloses the tumor on all sides.

---

<sup>1</sup> 3D Slicer, USA; <https://www.slicer.org/>

<sup>2</sup> Fusion 360, Autodesk, USA; <https://www.autodesk.com/products/fusion-360/>

5. In the modify tab, use the '**Combine**' tool to cut the tumor (the '**Brain Body**') from the box (the '**Target Body**'). This will then leave a box with a hollow shape of the tumor inside it.
6. Check that the hollowed-out box is present. Cut this box into an appropriate number of pieces so that once the mold is filled, it can be prized apart without damaging the phantom inside. For the tumor here, it is enough to split the box in two, but for the other parts of the phantom, more pieces are needed.
7. Create planes through the box in the places where the mold needs to be cut. Click '**Construct**' and then '**Midplane**' to create a plane through the center of the box. Right-click on the created plane and choose '**Offset Plane**' to position the plane more precisely.
8. Use the '**Split Body**' function in the '**Modify**' tab to split the mold along the planes created.
9. Move the individual pieces of the mold, by right-clicking and selecting '**Move/Copy**', so that all the pieces are facing outwards.
10. Add rivets to the faces of each piece of the mold (so it can fit together securely), by clicking '**Create sketch**' then '**Centre diameter circle**' and on each face, drawing small circles. Right click then '**Extrude**' these circles outwards a few millimeters on one face and extrude them inwards on the corresponding face.  
NOTE: The circles that are extruded inwards need to be slightly bigger – approximately 1.5 mm - than those that are extruded outwards so that they will fit together snugly.
11. Create a hole at the side of the mold to pour the material into the tumor. Click '**Hole**' in the '**SOLID**' tab and draw a hole on the tumor side. Select the hole type to '**Countersink**'.
12. Similarly, create another hole at the bottom of the molds to pour the material in.
13. Save each piece of the mold as a separate STL file.

## 2. Print the 3D molds

1. Install or open 3D printing software.
2. Open the STL file for each piece of the mold in the printing software and rotate it so that it lies flat against the build plate. It is possible to add multiple mold pieces to the build plate and to print these simultaneously.
3. Choose a large layer height (around 0.2 mm) and low infill value (around 20%) for faster printing. Print the molds using a rigid material such as Polylactic acid (PLA). If the molds are positioned appropriately, the support material is not necessary.

## A.6 Preparation of Agar

1. Measure 75 g of agar powder and set it to the side.
2. Mix 1500 mL of cold water with the agar powder.
3. Stir until the agar is suspended in water without clumps.
4. Briefly bring the mixture to a boiler while stirring periodically.

5. The agar can be left at room temperature for a few hours until the temperature reaches 50 °C.

## **A.7 Phantom Assembly**

1. Measure out enough agar to fill the tumor mold into a beaker.
2. To the agar gel for the tumor, add 1 w/w% glass microspheres for ultrasound contrast and stir by hand.
3. Sonicate the beaker to ensure homogenous mixing of the additives.
4. Leave to cool and allow any bubbles formed to escape, around 10 min, then scrape any bubbles from the surface.
5. Secure the tumor mold together (tape can be used to cover the joins in the mold) and pour about 10% of the prepared mix through the hole in the top of the mold.
6. Let set at room temperature for 20 min.
7. Pour the remaining agar gel into the mold.
8. The agar phantom can be left uncovered at room temperature for one week or up to four weeks in the fridge.

## **A.8 Phantom US Imaging**

1. Apply ultrasound gel to the imaging probe.  
NOTE: Gel is not used intraoperatively but may be used in simulation and does not significantly change the clinical workflow or the quality of the acquired images.
2. Image the brain and tumor through the craniotomy, with a clinical scanner and burr hole probe.

# Appendix B. List of Publications

- Zeineldin, R., Karar, M., Coburger, J., Wirtz, C., Mathis-Ullrich, F. & Burgert, O. (2020). Towards automated correction of brain shift using deep deformable magnetic resonance imaging-intraoperative ultrasound (MRI-iUS) registration. *Current Directions in Biomedical Engineering*, 6(1), 20200039. <https://doi.org/10.1515/cdbme-2020-0039>
- Zeineldin, R. A., Karar, M. E., Coburger, J., Wirtz, C. R., & Burgert, O. (2020). DeepSeg: deep neural network framework for automatic brain tumor segmentation using magnetic resonance FLAIR images. *International journal of computer assisted radiology and surgery*, 15(6), 909-920. <https://doi.org/10.1007/s11548-020-02186-z>
- Zeineldin, R., Weimann, P., Karar, M., Mathis-Ullrich, F. & Burgert, O. (2021). Slicer-DeepSeg: Open-Source Deep Learning Toolkit for Brain Tumour Segmentation. *Current Directions in Biomedical Engineering*, 7(1), 30-34. <https://doi.org/10.1515/cdbme-2021-1007>
- Zeineldin, R.A., Karar, M.E., Mathis-Ullrich, F., Burgert, O. (2021). A Hybrid Deep Registration of MR Scans to Interventional Ultrasound for Neurosurgical Guidance. In *International Workshop on Machine Learning in Medical Imaging* (pp. 586-595). Springer, Cham. [https://doi.org/10.1007/978-3-030-87589-3\\_60](https://doi.org/10.1007/978-3-030-87589-3_60)
- Zeineldin, R. A., Karar, M. E., Elshaer, Z., Schmidhammer, M., Coburger, J., Wirtz, C. R., ... & Mathis-Ullrich, F. (2021). iRegNet: non-rigid registration of MRI to interventional US for brain-shift compensation using convolutional neural networks. *IEEE Access*, 9, 147579-147590. <https://doi.org/10.1109/ACCESS.2021.3120306>
- Zeineldin, R.A., Karar, M.E., Mathis-Ullrich, F., Burgert, O. (2022). Ensemble CNN Networks for GBM Tumors Segmentation Using Multi-parametric MRI. In: Crimi, A., Bakas, S. (eds) *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2021. Lecture Notes in Computer Science*, vol 12962. Springer, Cham. [https://doi.org/10.1007/978-3-031-08999-2\\_41](https://doi.org/10.1007/978-3-031-08999-2_41)
- Zeineldin, R. A., Karar, M. E., Elshaer, Z., Coburger, J., Wirtz, C. R., Burgert, O., & Mathis-Ullrich, F. (2022). Explainability of deep neural networks for MRI analysis of brain tumors. *Int J Comput Assist Radiol Surg*, 17(9), 1673-1683. <https://doi.org/10.1007/s11548-022-02619-x>
- Zeineldin, R.A., Pollok, A., Mangliers, T., Karar, M., Mathis-Ullrich, F. & Burgert, O. (2022). Deep automatic segmentation of brain tumours in interventional ultrasound data. *Current Directions in Biomedical Engineering*, 8(1), 133-137. <https://doi.org/10.1515/cdbme-2022-0034>
- Zeineldin, R.A., Karar, M.E., Mathis-Ullrich, F., Burgert, O. (2023). Multimodal CNN Networks for Brain Tumor Segmentation in MRI: A BraTS 2022 Challenge Solution. Accepted in BraTS 2022 Challenge, BrainLes 2022. *Lecture Notes in Computer Science*. Springer, Cham.
- Zeineldin, R.A., Karar, M.E., Mathis-Ullrich, F., Burgert, O. (2023). Self-supervised iRegNet for the Registration of Longitudinal Brain MRI of Diffuse Glioma Patients. Accepted in BraTS-Reg 2022 Challenge, BrainLes 2022. *Lecture Notes in Computer Science*. Springer, Cham.
- Zeineldin, R. A., Karar, M. E., Elshaer, Z., Coburger, J., Wirtz, C. R., ... & Mathis-Ullrich, F. (2023). Explainable Hybrid Vision Transformers and Convolutional Network for Multimodal Glioma Segmentation in Brain MRI. *Under Review at IEEE Journal of Biomedical and Health Informatic*



# References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., . . . Devin, M. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*.
- Adebayo, J., Gilmer, J., Muelly, M., Goodfellow, I., Hardt, M., & Kim, B. (2018). *Sanity checks for saliency maps*. Paper presented at the Proceedings of the 32nd International Conference on Neural Information Processing Systems.
- Amari, S.-i. (1993). Backpropagation and stochastic gradient descent method. *Neurocomputing*, 5(4-5), 185-196. doi:10.1016/0925-2312(93)90006-o
- Angelov, P. P., Soares, E. A., Jiang, R., Arnold, N. I., & Atkinson, P. M. (2021). Explainable artificial intelligence: an analytical review. *WIREs Data Mining and Knowledge Discovery*, 11(5). doi:10.1002/widm.1424
- Apostolopoulos, I. D., & Mpesiana, T. A. (2020). Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine*, 43(2), 635-640. doi:10.1007/s13246-020-00865-4
- Armanious, K., Jiang, C., Fischer, M., Kustner, T., Hepp, T., Nikolaou, K., . . . Yang, B. (2020). MedGAN: Medical image translation using GANs. *Comput Med Imaging Graph*, 79, 101684. doi:10.1016/j.compmedimag.2019.101684
- Askeland, C., Solberg, O. V., Bakeng, J. B., Reinertsen, I., Tangen, G. A., Hofstad, E. F., . . . Lindseth, F. (2016). CustusX: an open-source research platform for image-guided therapy. *Int J Comput Assist Radiol Surg*, 11(4), 505-519. doi:10.1007/s11548-015-1292-0
- Baheti, B., Waldmannstetter, D., Chakrabarty, S., Akbari, H., Bilello, M., Wiestler, B., . . . Abidi, S. (2021). The brain tumor sequence registration challenge: Establishing correspondence between pre-operative and follow-up mri scans of diffuse glioma patients. *arXiv preprint arXiv:2112.06979*.
- Baid, U., Ghodasara, S., Bilello, M., Mohan, S., Calabrese, E., Colak, E., . . . Bakas, S. (2021). The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification. arXiv:2107.02314. Retrieved from <https://ui.adsabs.harvard.edu/abs/2021arXiv210702314B>
- Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J. S., . . . Davatzikos, C. (2017). Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci Data*, 4(1), 170117. doi:10.1038/sdata.2017.117
- Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., . . . Rozycki, M. (2018). Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. *arXiv preprint arXiv:1811.02629*.
- Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J., & Dalca, A. V. (2019). VoxelMorph: A Learning Framework for Deformable Medical Image Registration. *IEEE Trans Med Imaging*. doi:10.1109/TMI.2019.2897538
- Bastos, D. C. A., Juvekar, P., Tie, Y., Jowkar, N., Pieper, S., Wells, W. M., . . . Kapur, T. (2021). Challenges and Opportunities of Intraoperative 3D Ultrasound With Neuronavigation in Relation to Intraoperative MRI. *Front Oncol*, 11, 656519. doi:10.3389/fonc.2021.656519

- Baxter, J. S. H., Gibson, E., Eagleson, R., & Peters, T. M. (2018). The semiotics of medical image Segmentation. *Medical Image Analysis, 44*, 54-71. doi:10.1016/j.media.2017.11.007
- Bucholz, R. D., Smith, K. R., Laycock, K. A., & McDurmont, L. L. (2001). Three-dimensional localization: from image-guided surgery to information-guided therapy. *Methods, 25*(2), 186-200. doi:10.1006/meth.2001.1234
- Buckner, J. C. (2003). Factors Influencing Survival in High-Grade Gliomas. *Seminars in Oncology*: W.B. Saunders.
- Canalini, L., Klein, J., Miller, D., & Kikinis, R. (2020). Enhanced registration of ultrasound volumes by segmentation of resection cavity in neurosurgical procedures. *Int J Comput Assist Radiol Surg, 15*(12), 1963-1974. doi:10.1007/s11548-020-02273-1
- Censi, F., Mattei, E., Triventi, M., Bartolini, P., & Calcagnini, G. (2012). Radiofrequency identification and medical devices: the regulatory framework on electromagnetic compatibility. Part I: medical devices. *Expert Rev Med Devices, 9*(3), 283-288. doi:10.1586/erd.12.4
- Chan, H. P., Hadjiiski, L. M., & Samala, R. K. (2020). Computer-aided diagnosis in the era of deep learning. *Med Phys, 47*(5), e218-e227. doi:10.1002/mp.13764
- Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 40*, 834-848. doi:10.1109/TPAMI.2017.2699184
- Cheng, X., Zhang, L., & Zheng, Y. F. (2018). Deep similarity learning for multimodal medical images. *Computer Methods in Biomechanics and Biomedical Engineering-Imaging and Visualization, 6*(3), 248-252. doi:10.1080/21681163.2015.1135299
- Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*: Institute of Electrical and Electronics Engineers Inc.
- Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016* (pp. 424-432).
- Cireşan, D. C., Giusti, A., Gambardella, L. M., & Schmidhuber, J. (2012). Deep neural networks segment neuronal membranes in electron microscopy images. *Advances in Neural Information Processing Systems*.
- Clarkson, M. J., Zombori, G., Thompson, S., Totz, J., Song, Y., Espak, M., . . . Ourselin, S. (2015). The NifTK software platform for image-guided interventions: platform overview and NiftyLink messaging. *Int J Comput Assist Radiol Surg, 10*(3), 301-316. doi:10.1007/s11548-014-1124-7
- Cleary, K., & Peters, T. M. (2010). Image-guided interventions: technology review and clinical applications. *Annu Rev Biomed Eng, 12*(1), 119-142. doi:10.1146/annurev-bioeng-070909-105249
- Coburger, J., König, R. W., Scheuerle, A., Engelke, J., Hlavac, M., Thal, D. R., & Wirtz, C. R. (2014). Navigated high frequency ultrasound: description of technique and clinical comparison with conventional intracranial ultrasound. *World Neurosurg, 82*(3-4), 366-375. doi:10.1016/j.wneu.2014.05.025
- Coburger, J., Merkel, A., Scherer, M., Schwartz, F., Gessler, F., Roder, C., . . . Wirtz, C. R. (2016). Low-grade Glioma Surgery in Intraoperative Magnetic Resonance Imaging: Results of a Multicenter Retrospective Assessment of the German Study Group for Intraoperative Magnetic Resonance Imaging. *Neurosurgery, 78*(6), 775-786. doi:10.1227/NEU.0000000000001081



- Coburger, J., & Wirtz, C. R. (2019). Fluorescence guided surgery by 5-ALA and intraoperative MRI in high grade glioma: a systematic review. *J Neurooncol*, *141*(3), 533-546. doi:10.1007/s11060-018-03052-4
- Dahl, G. E., Sainath, T. N., & Hinton, G. E. (2013). *Improving deep neural networks for LVCSR using rectified linear units and dropout*. Paper presented at the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing.
- Dalca, A. V., Balakrishnan, G., Guttag, J., & Sabuncu, M. R. (2019). Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. *Med Image Anal*, *57*, 226-236. doi:10.1016/j.media.2019.07.006
- Datta, S. K., Shaikh, M. A., Srihari, S. N., & Gao, M. (2021). Soft Attention Improves Skin Cancer Classification Performance. In *Interpretability of Machine Intelligence in Medical Image Computing, and Topological Data Analysis and Its Applications for Medical Data* (pp. 13-23).
- De Momi, E., Ferrigno, G., Bosoni, G., Bassanini, P., Blasi, P., Casaceli, G., . . . Cardinale, F. (2016). A method for the assessment of time-varying brain shift during navigated epilepsy surgery. *Int J Comput Assist Radiol Surg*, *11*(3), 473-481. doi:10.1007/s11548-015-1259-1
- De Nigris, D., Collins, D. L., & Arbel, T. (2013). Fast rigid registration of pre-operative magnetic resonance images to intra-operative ultrasound for neurosurgery based on high confidence gradient orientations. *Int J Comput Assist Radiol Surg*, *8*(4), 649-661. doi:10.1007/s11548-013-0826-6
- de Vos, B. D., Berendsen, F. F., Viergever, M. A., Sokooti, H., Staring, M., & Isgum, I. (2019). A deep learning framework for unsupervised affine and deformable image registration. *Med Image Anal*, *52*, 128-143. doi:10.1016/j.media.2018.11.010
- Delorenzo, C., Papademetris, X., Staib, L. H., Vives, K. P., Spencer, D. D., & Duncan, J. S. (2010). Image-guided intraoperative cortical deformation recovery using game theory: application to neocortical epilepsy surgery. *IEEE Trans Med Imaging*, *29*(2), 322-338. doi:10.1109/TMI.2009.2027993
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1-22. doi:10.1111/j.2517-6161.1977.tb01600.x
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Diba, A., Sharma, V., Pazandeh, A., Pirsiavash, H., & Gool, L. V. (2017, 21-26 July 2017). *Weakly Supervised Cascaded Convolutional Networks*. Paper presented at the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
- Dietterich, T. G. (2000). Ensemble Methods in Machine Learning. In *Multiple Classifier Systems* (pp. 1-15).
- Dijkstra, N., Zeidman, P., Ondobaka, S., van Gerven, M. A. J., & Friston, K. (2017). Distinct Top-down and Bottom-up Brain Connectivity During Visual Perception and Imagery. *Sci Rep*, *7*(1), 5677. doi:10.1038/s41598-017-05888-8
- Dolz, J., Desrosiers, C., Wang, L., Yuan, J., Shen, D., & Ben Ayed, I. (2020). Deep CNN ensembles and suggestive annotations for infant brain MRI segmentation. *Comput Med Imaging Graph*, *79*, 101660. doi:10.1016/j.compmedimag.2019.101660
- Dominguez-Velasco, C. F., Perez-Lomeli, J. S., Padilla-Castaneda, M. A., Tello-Mata, I. E., & Alcocer-Barradas, V. (2022). *A Ventriculostomy Simulation through Augmented Reality Navigation System for Learning and Improving Skills in Neurosurgery*. Paper presented at the 2022 IEEE Mexican International Conference on Computer Science (ENC).
- Donnez, M., Carton, F.-X., Le Lann, F., De Schlichting, E., Chabanas, M., Linte, C. A., & Siewerdsen, J. H. (2021). *Realistic synthesis of brain tumor resection ultrasound*

- images with a generative adversarial network*. Paper presented at the Medical Imaging 2021: Image-Guided Procedures, Robotic Interventions, and Modeling.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., . . . Gelly, S. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- Drobny, D., Vercauteren, T., Ourselin, S., & Modat, M. (2018). Registration of MRI and iUS Data to Compensate Brain Shift Using a Symmetric Block-Matching Based Approach. In *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation* (pp. 172-178).
- Drouin, S., Kochanowska, A., Kersten-Oertel, M., Gerard, I. J., Zelmann, R., De Nigris, D., . . . Collins, D. L. (2017). IBIS: an OR ready open-source platform for image-guided neurosurgery. *Int J Comput Assist Radiol Surg*, 12(3), 363-378. doi:10.1007/s11548-016-1478-0
- Du, M., Liu, N., & Hu, X. (2019). Techniques for interpretable machine learning. *Communications of the ACM*, 63(1), 68-77. doi:10.1145/3359786
- Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(7).
- Dussik, K. (1942). On the possibility of using ultrasound waves as a diagnostic aid. *Neurol Psychiat*, 174, 153-168.
- Earle, M., Portu, G., & DeVos, E. (2016). Agar ultrasound phantoms for low-cost training without refrigeration. *Afr J Emerg Med*, 6(1), 18-23. doi:10.1016/j.afjem.2015.09.003
- Eitel, F., & Ritter, K. (2019). Testing the Robustness of Attribution Methods for Convolutional Neural Networks in MRI-Based Alzheimer's Disease Classification. In *Interpretability of Machine Intelligence in Medical Image Computing and Multimodal Learning for Clinical Decision Support* (pp. 3-11).
- Ellingson, B. M., Wen, P. Y., & Cloughesy, T. F. (2017). Modified Criteria for Radiographic Response Assessment in Glioblastoma Clinical Trials. *Neurotherapeutics*, 14(2), 307-320. doi:10.1007/s13311-016-0507-6
- Enquobahrie, A., Cheng, P., Gary, K., Ibanez, L., Gobbi, D., Lindseth, F., . . . Cleary, K. (2007). The image-guided surgery toolkit IGSTK: an open source C++ software toolkit. *J Digit Imaging*, 20 Suppl 1(Suppl 1), 21-33. doi:10.1007/s10278-007-9054-3
- Esmaeili, M., Vettukattil, R., Banitalebi, H., Krogh, N. R., & Geitung, J. T. (2021). Explainable Artificial Intelligence for Human-Machine Interaction in Brain Tumor Localization. *J Pers Med*, 11(11). doi:10.3390/jpm11111213
- Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J. C., Pujol, S., . . . Kikinis, R. (2012). 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging*, 30(9), 1323-1341. doi:10.1016/j.mri.2012.05.001
- Fromageau, J., Gennisson, J. L., Schmitt, C., Maurice, R. L., Mongrain, R., & Cloutier, G. (2007). Estimation of polyvinyl alcohol cryogel mechanical properties with four ultrasound elastography methods and comparison with gold standard testings. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 54(3), 498-509. doi:10.1109/tuffc.2007.273
- Fu, Y., Lei, Y., Wang, T., Curran, W. J., Liu, T., & Yang, X. (2021). A review of deep learning based methods for medical image multi-organ segmentation. *Phys Med*, 85, 107-122. doi:10.1016/j.ejmp.2021.05.003
- Galati, R., Simone, M., Barile, G., De Luca, R., Cartanese, C., & Grassi, G. (2020). Experimental Setup Employed in the Operating Room Based on Virtual and Mixed Reality: Analysis of Pros and Cons in Open Abdomen Surgery. *Journal of Healthcare Engineering*, 2020, 1-11. doi:10.1155/2020/8851964

- Galloway, J. R. L., Wolf, I., Vetter, M., Wegner, I., Nolden, M., Bottger, T., . . . Meinzer, H.-P. (2004). *The medical imaging interaction toolkit (MITK): a toolkit facilitating the creation of interactive software by extending VTK and ITK*. Paper presented at the Medical Imaging 2004: Visualization, Image-Guided Procedures, and Display.
- Ge, C., Gu, I. Y.-H., Jakola, A. S., & Yang, J. (2018). *Deep Learning and Multi-Sensor Fusion for Glioma Classification Using Multistream 2D Convolutional Networks*. Paper presented at the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).
- Gerard, I. J., Kersten-Oertel, M., Hall, J. A., Sirhan, D., & Collins, D. L. (2020). Brain Shift in Neuronavigation of Brain Tumors: An Updated Review of Intra-Operative Ultrasound Applications. *Front Oncol*, *10*, 618837. doi:10.3389/fonc.2020.618837
- Glorot, X., & Bengio, Y. (2010). *Understanding the difficulty of training deep feedforward neural networks*. Paper presented at the Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research. <https://proceedings.mlr.press/v9/glorot10a.html>
- Goch, C. J., Metzger, J., & Nolden, M. (2017). Abstract: Medical Research Data Management Using MITK and XNAT. In *Bildverarbeitung für die Medizin 2017* (pp. 305-305).
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*: MIT press.
- Guha, D., Alotaibi, N. M., Nguyen, N., Gupta, S., McFaul, C., & Yang, V. X. D. (2017). Augmented Reality in Neurosurgery: A Review of Current Concepts and Emerging Applications. *Canadian Journal of Neurological Sciences / Journal Canadien des Sciences Neurologiques*, *44*(3), 235-245. doi:10.1017/cjn.2016.443
- Gulum, M. A., Trombley, C. M., & Kantardzic, M. (2021). A Review of Explainable Deep Learning Cancer Detection Models in Medical Imaging. *Applied Sciences*, *11*(10). doi:10.3390/app11104573
- Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, *187*, 27-48. doi:10.1016/j.neucom.2015.09.116
- Habibi Aghdam, H., & Jahani Heravi, E. (2017). *Guide to Convolutional Neural Networks*.
- Haouchine, N., Juvekar, P., Necessian, M., Wells, W., Golby, A., & Frisken, S. (2022). Pose Estimation and Non-Rigid Registration for Augmented Reality During Neurosurgery. *IEEE Trans Biomed Eng*, *69*(4), 1310-1317. doi:10.1109/TBME.2021.3113841
- Haskins, G., Kruecker, J., Kruger, U., Xu, S., Pinto, P. A., Wood, B. J., & Yan, P. (2019). Learning deep similarity metric for 3D MR-TRUS image registration. *Int J Comput Assist Radiol Surg*, *14*(3), 417-425. doi:10.1007/s11548-018-1875-7
- Haskins, G., Kruger, U., & Yan, P. (2020). Deep learning in medical image registration: a survey. *Machine Vision and Applications*, *31*(1-2). doi:10.1007/s00138-020-01060-x
- Hastreiter, P., Rezk-Salama, C., Soza, G., Bauer, M., Greiner, G., Fahlbusch, R., . . . Nimsky, C. (2004). Strategies for brain shift evaluation. *Med Image Anal*, *8*(4), 447-464. doi:10.1016/j.media.2004.02.001
- Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., . . . Larochelle, H. (2017). Brain tumor segmentation with Deep Neural Networks. *Medical Image Analysis*, *35*, 18-31. doi:10.1016/j.media.2016.05.004
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.
- Heinrich, M. P. (2018). Intra-operative Ultrasound to MRI Fusion with a Public Multimodal Discrete Registration Tool. In *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation* (pp. 159-164).

- Heinrich, M. P., Jenkinson, M., Papież, B. W., Brady, S. M., & Schnabel, J. A. (2013). *Towards Realtime Multimodal Fusion for Image-Guided Interventions Using Self-similarities*, Berlin, Heidelberg.
- Hervey-Jumper, S. L., & Berger, M. S. (2016). Maximizing safe resection of low- and high-grade glioma. *J Neurooncol*, *130*(2), 269-282. doi:10.1007/s11060-016-2110-4
- Holland, E. C. (2001). Progenitor cells and glioma formation. *Current opinion in neurology*, *14*, 683-688.
- Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B. (2017). What do we need to build explainable AI systems for the medical domain? *arXiv preprint arXiv:1712.09923*.
- Hong, J., & Park, H. (2018). Non-linear Approach for MRI to intra-operative US Registration Using Structural Skeleton. In *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation* (pp. 138-145).
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., . . . Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.
- Hu, Y., Modat, M., Gibson, E., Li, W., Ghavami, N., Bonmati, E., . . . Vercauteren, T. (2018). Weakly-supervised convolutional neural networks for multimodal image registration. *Med Image Anal*, *49*, 1-13. doi:10.1016/j.media.2018.07.002
- Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*: Institute of Electrical and Electronics Engineers Inc.
- Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *32nd International Conference on Machine Learning, ICML 2015*: International Machine Learning Society (IMLS).
- Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*, *18*(2), 203-211. doi:10.1038/s41592-020-01008-z
- Isensee, F., Jäger, P. F., Full, P. M., Vollmuth, P., & Maier-Hein, K. H. (2021). nnU-Net for Brain Tumor Segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (pp. 118-132).
- Jain, K. K. (2018). A Critical Overview of Targeted Therapies for Glioblastoma. *Front Oncol*, *8*, 419. doi:10.3389/fonc.2018.00419
- Jiang, D., Shi, Y., Yao, D., Wang, M., & Song, Z. (2016). miLBP: a robust and fast modality-independent 3D LBP for multimodal deformable registration. *Int J Comput Assist Radiol Surg*, *11*(6), 997-1005. doi:10.1007/s11548-016-1407-2
- Johnson, H., Harris, G., & Williams, K. (2007). BRAINSFit: Mutual Information Registrations of Whole-Brain 3D Images, Using the Insight Toolkit. *The Insight Journal*. doi:10.54294/hmb052
- Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S., Hudspeth, A. J., & Mack, S. (2000). *Principles of neural science* (Vol. 4): McGraw-hill New York.
- Kang, J., Ullah, Z., & Gwak, J. (2021). MRI-Based Brain Tumor Classification Using Ensemble of Deep Features and Machine Learning Classifiers. *Sensors (Basel)*, *21*(6). doi:10.3390/s21062222
- Kapishnikov, A., Venugopalan, S., Avci, B., Wedin, B., Terry, M., & Bolukbasi, T. (2021). *Guided Integrated Gradients: An Adaptive Path Method for Removing Noise*. Paper presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.
- Karar, M. E., Merk, D. R., Falk, V., & Burgert, O. (2016). A simple and accurate method for computer-aided transapical aortic valve replacement. *Computerized Medical Imaging and Graphics*, *50*, 31-41. doi:10.1016/j.compmedimag.2014.09.005

- Kim, M., Yun, J., Cho, Y., Shin, K., Jang, R., Bae, H. J., & Kim, N. (2019). Deep Learning in Medical Imaging. *Neurospine*, *16*(4), 657-668. doi:10.14245/ns.1938396.198
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kneöaurek, K., Ivanovic, M., Machac, J., & Weber, D. A. (2000). Medical image registration. *Europhysics News*, *31*(4), 5-8. doi:10.1051/ePN:2000401
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). *ImageNet classification with deep convolutional neural networks* (Vol. 60): Association for Computing Machinery.
- Kubben, P. L., Postma, A. A., Kessels, A. G., van Overbeeke, J. J., & van Santbrink, H. (2010). Intraobserver and interobserver agreement in volumetric assessment of glioblastoma multiforme resection. *Neurosurgery*, *67*(5), 1329-1334. doi:10.1227/NEU.0b013e3181efbb08
- Lapointe, S., Perry, A., & Butowski, N. A. (2018). Primary brain tumours in adults. *Lancet*, *392*(10145), 432-446. doi:10.1016/S0140-6736(18)30990-5
- Lasso, A., Heffter, T., Rankin, A., Pinter, C., Ungi, T., & Fichtinger, G. (2014). PLUS: open-source toolkit for ultrasound-guided intervention systems. *IEEE Trans Biomed Eng*, *61*(10), 2527-2537. doi:10.1109/TBME.2014.2322864
- Lauterbur, P. C. (1973). Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance. *Nature*, *242*(5394), 190-191. doi:10.1038/242190a0
- Le, N. Q. K., Hung, T. N. K., Do, D. T., Lam, L. H. T., Dang, L. H., & Huynh, T. T. (2021). Radiomics-based machine learning model for efficiently classifying transcriptome subtypes in glioblastoma patients from MRI. *Comput Biol Med*, *132*, 104320. doi:10.1016/j.compbiomed.2021.104320
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444. doi:10.1038/nature14539
- Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324. doi:10.1109/5.726791
- Lee, J., Liu, P., Cheng, J., & Fu, H. (2019). *A Deep Step Pattern Representation for Multimodal Retinal Image Registration*. Paper presented at the 2019 IEEE/CVF International Conference on Computer Vision (ICCV).
- Li, H., & Fan, Y. (2018). Non-Rigid Image Registration Using Self-Supervised Fully Convolutional Networks without Training Data. *Proc IEEE Int Symp Biomed Imaging*, *2018*, 1075-1078. doi:10.1109/ISBI.2018.8363757
- Li, J., Huo, H., Li, C., Wang, R., Sui, C., & Liu, Z. (2021). Multigrained Attention Network for Infrared and Visible Image Fusion. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1-12. doi:10.1109/tim.2020.3029360
- Li, Y. M., Suki, D., Hess, K., & Sawaya, R. (2016). The influence of maximum safe resection of glioblastoma on survival in 1229 patients: Can we do better than gross-total resection? *J Neurosurg*, *124*(4), 977-988. doi:10.3171/2015.5.JNS142087
- Liang, Z.-P., & Lauterbur, P. C. (2000). *Principles of magnetic resonance imaging*: SPIE Optical Engineering Press Bellingham.
- Liew, A., Lee, C. C., Lan, B. L., & Tan, M. (2021). CASPIANET++: A multidimensional Channel-Spatial Asymmetric attention network with Noisy Student Curriculum Learning paradigm for brain tumor segmentation. *Comput Biol Med*, *136*, 104690. doi:10.1016/j.compbiomed.2021.104690
- Liu, J., Singh, G., Al'Aref, S., Lee, B., Oleru, O., Min, J. K., . . . Mosadegh, B. (2019). Image Registration in Medical Robotics and Intelligent Systems: Fundamentals and Applications. *Advanced Intelligent Systems*, *1*(6). doi:10.1002/aisy.201900048
- Louis, D. N., Wesseling, P., Aldape, K., Brat, D. J., Capper, D., Cree, I. A., . . . Ellison, D. W. (2020). cIMPACT-NOW update 6: new entity and diagnostic principle

- recommendations of the cIMPACT-Utrecht meeting on future CNS tumor classification and grading. *Brain Pathol*, 30(4), 844-856. doi:10.1111/bpa.12832
- Lundberg, S. M., & Lee, S.-I. (2017). *A unified approach to interpreting model predictions*. Paper presented at the Proceedings of the 31st international conference on neural information processing systems.
- Lundervold, A. S., & Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on MRI. *Z Med Phys*, 29(2), 102-127. doi:10.1016/j.zemedi.2018.11.002
- Luo, X., Wang, G., Song, T., Zhang, J., Aertsen, M., Deprest, J., . . . Zhang, S. (2021). MIDeepSeg: Minimally interactive segmentation of unseen objects from medical images using deep learning. *Medical Image Analysis*, 72. doi:10.1016/j.media.2021.102102
- Lynch, C. J., & Liston, C. (2018). New machine-learning technologies for computer-aided diagnosis. *Nat Med*, 24(9), 1304-1305. doi:10.1038/s41591-018-0178-4
- Ma, K., Wang, J., Singh, V., Tamersoy, B., Chang, Y.-J., Wimmer, A., & Chen, T. (2017). *Multimodal Image Registration with Deep Context Reinforcement Learning*, Cham.
- Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). *Rectifier nonlinearities improve neural network acoustic models*. Paper presented at the Proc. icml.
- Machado, I., Toews, M., George, E., Unadkat, P., Essayed, W., Luo, J., . . . Ou, Y. (2019). Deformable MRI-Ultrasound registration using correlation-based attribute matching for brain shift correction: Accuracy and generality in multi-site data. *Neuroimage*, 202, 116094. doi:10.1016/j.neuroimage.2019.116094
- Machado, I., Toews, M., Luo, J., Unadkat, P., Essayed, W., George, E., . . . Ou, Y. (2018). Deformable MRI-Ultrasound Registration via Attribute Matching and Mutual-Saliency Weighting for Image-Guided Neurosurgery. In *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation* (pp. 165-171).
- Mackle, E. C., Shapey, J., Maneas, E., Saeed, S. R., Bradford, R., Ourselin, S., . . . Desjardins, A. E. (2020). Patient-Specific Polyvinyl Alcohol Phantom Fabrication with Ultrasound and X-Ray Contrast for Brain Tumor Surgery Planning. *Journal of Visualized Experiments*(161). doi:10.3791/61344
- Mahapatra, D., Bozorgtabar, B., & Ge, Z. (2021). *Medical Image Classification Using Generalized Zero Shot Learning*. Paper presented at the Proceedings of the IEEE/CVF International Conference on Computer Vision.
- Maloca, P. M., Muller, P. L., Lee, A. Y., Tufail, A., Balaskas, K., Niklaus, S., . . . Denk, N. (2021). Unraveling the deep learning gearbox in optical coherence tomography image segmentation towards explainable artificial intelligence. *Commun Biol*, 4(1), 170. doi:10.1038/s42003-021-01697-y
- Mascitelli, J. R., Schlachter, L., Chartrain, A. G., Oemke, H., Gilligan, J., Costa, A. B., . . . Bederson, J. B. (2018). Navigation-Linked Heads-Up Display in Intracranial Surgery: Early Experience. *Operative Neurosurgery*, 15(2), 184-193. doi:10.1093/ons/oxp205
- Masoumi, N., Xiao, Y., & Rivaz, H. (2019). ARENA: Inter-modality affine registration using evolutionary strategy. *Int J Comput Assist Radiol Surg*, 14(3), 441-450. doi:10.1007/s11548-018-1897-1
- McKinley, R., Meier, R., & Wiest, R. (2019). Ensembles of Densely-Connected CNNs with Label-Uncertainty for Brain Tumor Segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (pp. 456-465).
- Mehrtash, A., Pesteie, M., Hetherington, J., Behringer, P. A., Kapur, T., Wells, W. M., 3rd, . . . Abolmaesumi, P. (2017). DeepInfer: Open-Source Deep Learning Deployment Toolkit for Image-Guided Therapy. *Proc SPIE Int Soc Opt Eng*, 10135. doi:10.1117/12.2256011
- Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., . . . Van Leemput, K. (2015). The Multimodal Brain Tumor Image Segmentation Benchmark

- (BRATS). *IEEE Trans Med Imaging*, 34(10), 1993-2024.  
doi:10.1109/TMI.2014.2377694
- Mercier, L., Del Maestro, R. F., Petrecca, K., Araujo, D., Haegelen, C., & Collins, D. L. (2012). Online database of clinical MR and ultrasound images of brain tumors. *Med Phys*, 39(6), 3253-3261. doi:10.1118/1.4709600
- Mercier, L., Fonov, V., Haegelen, C., Del Maestro, R. F., Petrecca, K., & Collins, D. L. (2012). Comparing two approaches to rigid registration of three-dimensional ultrasound and magnetic resonance images for neurosurgery. *Int J Comput Assist Radiol Surg*, 7(1), 125-136. doi:10.1007/s11548-011-0620-2
- Miller, D., Benes, L., & Sure, U. (2011). Stand-alone 3D-ultrasound navigation after failure of conventional image guidance for deep-seated lesions. *Neurosurg Rev*, 34(3), 381-387; discussion 387-388. doi:10.1007/s10143-011-0314-9
- Miner, R. C. (2017). Image-Guided Neurosurgery. *J Med Imaging Radiat Sci*, 48(4), 328-335. doi:10.1016/j.jmir.2017.06.005
- Muhammad, K., Khan, S., Ser, J. D., & Albuquerque, V. H. C. (2021). Deep Learning for Multigrade Brain Tumor Classification in Smart Healthcare Systems: A Prospective Survey. *IEEE Trans Neural Netw Learn Syst*, 32(2), 507-522. doi:10.1109/TNNLS.2020.2995800
- Mustaf, M., Sali, A., Illzam, E., Sharifa, A., & Nang, M. (2018). Brain cancer: Current concepts, diagnosis and prognosis. *IOSR Journal of Dental and Medical Sciences*, 17(3), 41-46.
- naceur, M. B., Saouli, R., Akil, M., & Kachouri, R. (2018). Fully Automatic Brain Tumor Segmentation using End-To-End Incremental Deep Neural Networks in MRI images. *Computer Methods and Programs in Biomedicine*, 166, 39-49. doi:10.1016/j.cmpb.2018.09.007
- Nan, A., Tennant, M., Rubin, U., & Ray, N. (2020). *DRMIME: Differentiable Mutual Information and Matrix Exponential for Multi-Resolution Image Registration*. Paper presented at the Proceedings of the Third Conference on Medical Imaging with Deep Learning, Proceedings of Machine Learning Research. <https://proceedings.mlr.press/v121/nan20a.html>
- Natekar, P., Kori, A., & Krishnamurthi, G. (2020). Demystifying Brain Tumor Segmentation Networks: Interpretability and Uncertainty Analysis. *Front Comput Neurosci*, 14, 6. doi:10.3389/fncom.2020.00006
- Novelline, R. A., & Squire, L. F. (2004). *Squire's fundamentals of radiology*: La Editorial, UPR.
- Ouyang, W., Zeng, X., Wang, X., Qiu, S., Luo, P., Tian, Y., . . . Tang, X. (2017). DeepID-Net: Object Detection with Deformable Part Based Convolutional Neural Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39, 1320-1334. doi:10.1109/TPAMI.2016.2587642
- Pacioni, A., Carbone, M., Freschi, C., Vigliani, R., Ferrari, V., & Ferrari, M. (2014). Patient-specific ultrasound liver phantom: materials and fabrication method. *International Journal of Computer Assisted Radiology and Surgery*, 10(7), 1065-1075. doi:10.1007/s11548-014-1120-y
- Pagani, E., Bizzi, A., Di Salle, F., De Stefano, N., & Filippi, M. (2008). Basic concepts of advanced MRI techniques. *Neurol Sci*, 29 Suppl 3(S3), 290-295. doi:10.1007/s10072-008-1001-7
- Pala, A., Durner, G., Braun, M., Schmitz, B., Wirtz, C. R., & Coburger, J. (2021). The Impact of an Ultra-Early Postoperative MRI on Treatment of Lower Grade Glioma. *Cancers (Basel)*, 13(12). doi:10.3390/cancers13122914
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., . . . Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in*

- Neural Information Processing Systems 32 (Nips 2019)*, 32, 8026-8037. Retrieved from <Go to ISI>://WOS:000534424308009
- Pereira, S., Meier, R., Alves, V., Reyes, M., & Silva, C. A. (2018). Automatic Brain Tumor Grading from MRI Data Using Convolutional Neural Networks and Quality Assessment. In *Understanding and Interpreting Machine Learning in Medical Image Computing Applications* (pp. 106-114).
- Pereira, S., Meier, R., McKinley, R., Wiest, R., Alves, V., Silva, C. A., & Reyes, M. (2018). Enhancing interpretability of automatically extracted machine learning features: application to a RBM-Random Forest system on brain lesion segmentation. *Med Image Anal*, 44, 228-244. doi:10.1016/j.media.2017.12.009
- Pesapane, F., Volonte, C., Codari, M., & Sardanelli, F. (2018). Artificial intelligence as a medical device in radiology: ethical and regulatory issues in Europe and the United States. *Insights Imaging*, 9(5), 745-753. doi:10.1007/s13244-018-0645-y
- Pooley, R. A. (2005). AAPM/RSNA physics tutorial for residents: fundamental physics of MR imaging. *RadioGraphics*, 25(4), 1087-1099. doi:10.1148/rg.254055027
- Pope, W. B., & Brandal, G. (2018). Conventional and advanced magnetic resonance imaging in patients with high-grade glioma. *Q J Nucl Med Mol Imaging*, 62(3), 239-253. doi:10.23736/S1824-4785.18.03086-8
- Preiswerk, F., Brinker, S. T., McDannold, N. J., & Mariano, T. Y. (2019). Open-source neuronavigation for multimodal non-invasive brain stimulation using 3D Slicer. arXiv:1909.12458. Retrieved from <https://ui.adsabs.harvard.edu/abs/2019arXiv190912458P>
- Qian, N. (1999). On the momentum term in gradient descent learning algorithms. *Neural Netw*, 12(1), 145-151. doi:10.1016/s0893-6080(98)00116-6
- Rajan, S. S. (1997). MRI: a conceptual overview.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should i trust you?" Explaining the predictions of any classifier. Paper presented at the Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining.
- Rivaz, H., Chen, S. J., & Collins, D. L. (2015). Automatic deformable MR-ultrasound registration for image-guided neurosurgery. *IEEE Trans Med Imaging*, 34(2), 366-380. doi:10.1109/TMI.2014.2354352
- Rivaz, H., Karimaghloo, Z., Fonov, V. S., & Collins, D. L. (2014). Nonrigid registration of ultrasound and MRI using contextual conditioned mutual information. *IEEE Trans Med Imaging*, 33(3), 708-725. doi:10.1109/TMI.2013.2294630
- Rohé, M.-M., Datar, M., Heimann, T., Sermesant, M., & Pennec, X. (2017). *SVF-Net: Learning Deformable Image Registration Using Shape Matching*, Cham.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (pp. 234-241).
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., . . . Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3), 211-252. doi:10.1007/s11263-015-0816-y
- Saleem, H., Shahid, A. R., & Raza, B. (2021). Visual interpretability in 3D brain tumor segmentation network. *Comput Biol Med*, 133, 104410. doi:10.1016/j.compbimed.2021.104410
- Saleh, K., Zeineldin, R. A., Hossny, M., Nahavandi, S., & El-Fishawy, N. (2018). *End-to-End Indoor Navigation Assistance for the Visually Impaired Using Monocular Camera*. Paper presented at the 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC).



- Sanai, N., Polley, M. Y., McDermott, M. W., Parsa, A. T., & Berger, M. S. (2011). An extent of resection threshold for newly diagnosed glioblastomas. *J Neurosurg*, *115*(1), 3-8. doi:10.3171/2011.2.jns10998
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 4510-4520. doi:10.1109/CVPR.2018.00474
- Sastry, R., Bi, W. L., Pieper, S., Frisken, S., Kapur, T., Wells, W., 3rd, & Golby, A. J. (2017). Applications of Ultrasound in the Resection of Brain Tumors. *J Neuroimaging*, *27*(1), 5-15. doi:10.1111/jon.12382
- Sayres, R., Taly, A., Rahimy, E., Blumer, K., Coz, D., Hammel, N., . . . Webster, D. R. (2019). Using a Deep Learning Algorithm and Integrated Gradients Explanation to Assist Grading for Diabetic Retinopathy. *Ophthalmology*, *126*(4), 552-564. doi:10.1016/j.ophtha.2018.11.016
- Schipmann-Miletić, S., & Stummer, W. (2020). Image-Guided Brain Surgery. In *Molecular Imaging in Oncology* (pp. 813-841).
- Sedghi, A., O'Donnell, L. J., Kapur, T., Learned-Miller, E., Mousavi, P., & Wells, W. M. (2021). Image registration: Maximum likelihood, minimum entropy and deep learning. *Medical Image Analysis*, *69*. doi:10.1016/j.media.2020.101939
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). *Grad-cam: Visual explanations from deep networks via gradient-based localization*. Paper presented at the Proceedings of the IEEE international conference on computer vision.
- Shah, A. H., & Heiss, J. D. (2022). Neurosurgical Clinical Trials for Glioblastoma: Current and Future Directions. *Brain Sciences*, *12*(6). doi:10.3390/brainsci12060787
- Shams, R., Boucher, M.-A., & Kadoury, S. (2018). Intra-operative Brain Shift Correction with Weighted Locally Linear Correlations of 3DUS and MRI. In *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation* (pp. 179-184).
- Shapey, J., Dowrick, T., Delaunay, R., Mackle, E. C., Thompson, S., Janatka, M., . . . Vercauteren, T. (2021). Integrated multi-modality image-guided navigation for neurosurgery: open-source software platform using state-of-the-art clinical hardware. *Int J Comput Assist Radiol Surg*, *16*(8), 1347-1356. doi:10.1007/s11548-021-02374-5
- Shelhamer, E., Long, J., & Darrell, T. (2017). Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*, 640-651. doi:10.1109/TPAMI.2016.2572683
- Shetty, P., & Moiyadi, A. V. (2016). Clinical Ultrasound: Historical Aspects. In *Intraoperative Ultrasound (IOUS) in Neurosurgery* (pp. 3-8).
- Shorfuzzaman, M. (2021). An explainable stacked ensemble of deep learning models for improved melanoma skin cancer detection. *Multimedia Systems*, *28*(4), 1309-1323. doi:10.1007/s00530-021-00787-5
- Shrikumar, A., Greenside, P., Shcherbina, A., & Kundaje, A. (2016). Not just a black box: Learning important features through propagating activation differences. *arXiv preprint arXiv:1605.01713*.
- Siegel, R. L., Miller, K. D., & Jemal, A. (2019). Cancer statistics, 2019 (US statistics). *CA: A Cancer Journal for Clinicians*, *69*, 7-34. doi:10.3322/caac.21551
- Siekmann, M., Lothes, T., Konig, R., Wirtz, C. R., & Coburger, J. (2018). Experimental study of sector and linear array ultrasound accuracy and the influence of navigated 3D-reconstruction as compared to MRI in a brain tumor model. *Int J Comput Assist Radiol Surg*, *13*(3), 471-478. doi:10.1007/s11548-018-1705-y

- Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). *Deep inside convolutional networks: Visualising image classification models and saliency maps*. Paper presented at the In Workshop at International Conference on Learning Representations.
- Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition.
- Singh, R. K., Pandey, R., & Babu, R. N. (2021). COVIDScreen: explainable deep learning framework for differential diagnosis of COVID-19 using chest X-rays. *Neural Comput Appl*, 33(14), 8871-8892. doi:10.1007/s00521-020-05636-6
- Smilkov, D., Thorat, N., Kim, B., Viégas, F., & Wattenberg, M. (2017). Smoothgrad: removing noise by adding noise. *arXiv preprint arXiv:1706.03825*.
- Sotiras, A., Davatzikos, C., & Paragios, N. (2013). Deformable medical image registration: a survey. *IEEE Trans Med Imaging*, 32(7), 1153-1190. doi:10.1109/TMI.2013.2265603
- Springenberg, J., Dosovitskiy, A., Brox, T., & Riedmiller, M. (2015). *Striving for Simplicity: The All Convolutional Net*. Paper presented at the ICLR (workshop track).
- Srivastava, N., Hinton, G., Krizhevsky, A., & Salakhutdinov, R. (2014). *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*. Paper presented at the Journal of Machine Learning Research.
- Steno, A., Buvala, J., Babkova, V., Kiss, A., Toma, D., & Lysak, A. (2021). Current Limitations of Intraoperative Ultrasound in Brain Tumor Surgery. *Front Oncol*, 11, 659048. doi:10.3389/fonc.2021.659048
- Stoll, J. (2014). Technology of Ultrasound-Guided Therapy. In *Intraoperative Imaging and Image-Guided Therapy* (pp. 155-162).
- Sun, L., & Zhang, S. (2018). Deformable MRI-Ultrasound Registration Using 3D Convolutional Neural Network. In *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation* (pp. 152-158).
- Sundararajan, M., Taly, A., & Yan, Q. (2017). *Axiomatic attribution for deep networks*. Paper presented at the International Conference on Machine Learning.
- Tacher, V., Lin, M., Chao, M., Gjestebj, L., Bhagat, N., Mahammedi, A., . . . Geschwind, J.-F. (2013). Semiautomatic Volumetric Tumor Segmentation for Hepatocellular Carcinoma. *Academic Radiology*, 20(4), 446-452. doi:10.1016/j.acra.2012.11.009
- Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J. N., Wu, Z., & Ding, X. (2020). Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Med Image Anal*, 63, 101693. doi:10.1016/j.media.2020.101693
- Temme, M. (2017). Algorithms and Transparency in View of the New General Data Protection Regulation. *European Data Protection Law Review*, 3(4), 473-485. doi:10.21552/edpl/2017/4/9
- Tokuda, J., Fischer, G. S., Papademetris, X., Yaniv, Z., Ibanez, L., Cheng, P., . . . Hata, N. (2009). OpenIGTLink: an open network protocol for image-guided therapy environment. *Int J Med Robot*, 5(4), 423-434. doi:10.1002/rcs.274
- Ungi, T., Lasso, A., & Fichtinger, G. (2016). Open-source platforms for navigated image-guided interventions. *Med Image Anal*, 33, 181-186. doi:10.1016/j.media.2016.06.011
- Unsgaard, G., Gronningsaeter, A., Ommedal, S., & Nagelhus Hernes, T. A. (2002). Brain Operations Guided by Real-time Two-dimensional Ultrasound: New Possibilities as a Result of Improved Image Quality. *Neurosurgery*, 51(2), 402-412. Retrieved from [https://journals.lww.com/neurosurgery/Fulltext/2002/08000/Brain\\_Operations\\_Guided\\_by\\_Real\\_time.19.aspx](https://journals.lww.com/neurosurgery/Fulltext/2002/08000/Brain_Operations_Guided_by_Real_time.19.aspx)
- Upadhyay, N., & Waldman, A. D. (2011). Conventional MRI evaluation of gliomas. *Br J Radiol*, 84 Spec No 2(special\_issue\_2), S107-111. doi:10.1259/bjr/65711810
- van Geuns, R. J., Wielopolski, P. A., de Bruin, H. G., Rensing, B. J., van Ooijen, P. M., Hulshoff, M., . . . de Feyter, P. J. (1999). Basic principles of magnetic resonance imaging. *Prog Cardiovasc Dis*, 42(2), 149-156. doi:10.1016/s0033-0620(99)70014-9

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Vincent, R. D., Neelin, P., Khalili-Mahani, N., Janke, A. L., Fonov, V. S., Robbins, S. M., . . . Evans, A. C. (2016). MINC 2.0: A Flexible Format for Multi-Modal Images. *Front Neuroinform*, 10, 35. doi:10.3389/fninf.2016.00035
- Visser, M., Muller, D. M. J., van Duijn, R. J. M., Smits, M., Verburg, N., Hendriks, E. J., . . . de Munck, J. C. (2019). Inter-rater agreement in glioma segmentations on longitudinal MRI. *Neuroimage Clin*, 22, 101727. doi:10.1016/j.nicl.2019.101727
- Warfield, S. K., Zou, K. H., & Wells, W. M. (2004). Simultaneous Truth and Performance Level Estimation (STAPLE): An Algorithm for the Validation of Image Segmentation. *IEEE Transactions on Medical Imaging*, 23(7), 903-921. doi:10.1109/tmi.2004.828354
- Wein, W. (2018). Brain-Shift Correction with Image-Based Registration and Landmark Accuracy Evaluation. In *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation* (pp. 146-151).
- Wein, W., Ladikos, A., Fuerst, B., Shah, A., Sharma, K., & Navab, N. (2013). *Global Registration of Ultrasound to MRI Using the LC2 Metric for Enabling Neurosurgical Guidance*, Berlin, Heidelberg.
- Weller, M., van den Bent, M., Preusser, M., Le Rhun, E., Tonn, J. C., Minniti, G., . . . Wick, W. (2021). EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood. *Nat Rev Clin Oncol*, 18(3), 170-186. doi:10.1038/s41571-020-00447-z
- Wickstrom, K., Kampffmeyer, M., & Jenssen, R. (2020). Uncertainty and interpretability in convolutional neural networks for semantic segmentation of colorectal polyps. *Med Image Anal*, 60, 101619. doi:10.1016/j.media.2019.101619
- Windisch, P., Weber, P., Furweger, C., Ehret, F., Kufeld, M., Zwahlen, D., & Muacevic, A. (2020). Implementation of model explainability for a basic brain tumor detection using convolutional neural networks on MRI slices. *Neuroradiology*, 62(11), 1515-1518. doi:10.1007/s00234-020-02465-1
- Wu, W., Chen, A. Y. C., Zhao, L., & Corso, J. J. (2014). Brain tumor detection and segmentation in a CRF (conditional random fields) framework with pixel-pairwise affinity and superpixel-level features. *International Journal of Computer Assisted Radiology and Surgery*, 9, 241-253. doi:10.1007/s11548-013-0922-7
- Wu, Y. H., Gao, S. H., Mei, J., Xu, J., Fan, D. P., Zhang, R. G., & Cheng, M. M. (2021). JCS: An Explainable COVID-19 Diagnosis System by Joint Classification and Segmentation. *IEEE Trans Image Process*, 30, 3113-3126. doi:10.1109/TIP.2021.3058783
- Xiao, Y., Fortin, M., Unsgard, G., Rivaz, H., & Reinertsen, I. (2017). RETrospective Evaluation of Cerebral Tumors (RESECT): A clinical database of pre-operative MRI and intra-operative ultrasound in low-grade glioma surgeries. *Med Phys*, 44(7), 3875-3882. doi:10.1002/mp.12268
- Xiao, Y., Rivaz, H., Chabanas, M., Fortin, M., Machado, I., Ou, Y., . . . Reinertsen, I. (2020). Evaluation of MRI to Ultrasound Registration Methods for Brain Shift Correction: The CuRIOUS2018 Challenge. *IEEE Trans Med Imaging*, 39(3), 777-786. doi:10.1109/TMI.2019.2935060
- Xie, X., Niu, J., Liu, X., Chen, Z., Tang, S., & Yu, S. (2021). A survey on incorporating domain knowledge into deep learning for medical image analysis. *Med Image Anal*, 69, 101985. doi:10.1016/j.media.2021.101985
- Yang, G., Ye, Q., & Xia, J. (2022). Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond. *Inf Fusion*, 77, 29-52. doi:10.1016/j.inffus.2021.07.016
- Yang, P., Hwa Yang, Y., B Zhou, B., & Y Zomaya, A. (2010). A review of ensemble methods in bioinformatics. *Current Bioinformatics*, 5(4), 296-308.

- Yang, X., Kwitt, R., Styner, M., & Niethammer, M. (2017). Quicksilver: Fast predictive image registration - A deep learning approach. *Neuroimage*, *158*, 378-396. doi:10.1016/j.neuroimage.2017.07.008
- Yavas, G., Caliskan, K. E., & Cagli, M. S. (2021). Three-dimensional-printed marker-based augmented reality neuronavigation: a new neuronavigation technique. *Neurosurg Focus*, *51*(2), E20. doi:10.3171/2021.5.FOCUS21206
- Young, K., Booth, G., Simpson, B., Dutton, R., & Shrapnel, S. (2019). Deep Neural Network or Dermatologist? In *Interpretability of Machine Intelligence in Medical Image Computing and Multimodal Learning for Clinical Decision Support* (pp. 48-55).
- Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., & Gerig, G. (2006). User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*, *31*(3), 1116-1128. doi:10.1016/j.neuroimage.2006.01.015
- Zeiler, M. D., & Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. In *Computer Vision – ECCV 2014* (pp. 818-833).
- Zeineldin, R. A., Karar, M. E., Burgert, O., & Mathis-Ullrich, F. (2023). Multimodal CNN Networks for Brain Tumor Segmentation in MRI: A BraTS 2022 Challenge Solution. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*.
- Zeineldin, R. A., Karar, M. E., Coburger, J., Wirtz, C. R., & Burgert, O. (2020). DeepSeg: deep neural network framework for automatic brain tumor segmentation using magnetic resonance FLAIR images. *Int J Comput Assist Radiol Surg*, *15*(6), 909-920. doi:10.1007/s11548-020-02186-z
- Zeineldin, R. A., Karar, M. E., Coburger, J., Wirtz, C. R., Mathis-Ullrich, F., & Burgert, O. (2020). Towards automated correction of brain shift using deep deformable magnetic resonance imaging-intraoperative ultrasound (MRI-iUS) registration. *Current Directions in Biomedical Engineering*, *6*(1). doi:10.1515/cdbme-2020-0039
- Zeineldin, R. A., Karar, M. E., Elshaer, Z., Coburger, J., Wirtz, C. R., Burgert, O., & Mathis-Ullrich, F. (2022). Explainability of deep neural networks for MRI analysis of brain tumors. *Int J Comput Assist Radiol Surg*, *17*(9), 1673-1683. doi:10.1007/s11548-022-02619-x
- Zeineldin, R. A., Karar, M. E., Elshaer, Z., Schmidhammer, M., Coburger, J., Wirtz, C. R., . . . Mathis-Ullrich, F. (2021). iRegNet: Non-Rigid Registration of MRI to Interventional US for Brain-Shift Compensation Using Convolutional Neural Networks. *IEEE Access*, *9*, 147579-147590. doi:10.1109/access.2021.3120306
- Zeineldin, R. A., Karar, M. E., Mathis-Ullrich, F., & Burgert, O. (2021). A Hybrid Deep Registration of MR Scans to Interventional Ultrasound for Neurosurgical Guidance. In *Machine Learning in Medical Imaging* (pp. 586-595).
- Zeineldin, R. A., Karar, M. E., Mathis-Ullrich, F., & Burgert, O. (2022a). Ensemble CNN Networks for GBM Tumors Segmentation Using Multi-parametric MRI. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (pp. 473-483).
- Zeineldin, R. A., Karar, M. E., Mathis-Ullrich, F., & Burgert, O. (2022b). Self-supervised iRegNet for the Registration of Longitudinal Brain MRI of Diffuse Glioma Patients. arXiv:2211.11025. Retrieved from <https://ui.adsabs.harvard.edu/abs/2022arXiv221111025Z>
- Zeineldin, R. A., Pollok, A., Mangliers, T., Karar, M. E., Mathis-Ullrich, F., & Burgert, O. (2022). Deep automatic segmentation of brain tumours in interventional ultrasound data. *Current Directions in Biomedical Engineering*, *8*(1), 133-137. doi:10.1515/cdbme-2022-0034
- Zeineldin, R. A., Weimann, P., Karar, M. E., Mathis-Ullrich, F., & Burgert, O. (2021). Slicer-DeepSeg: Open-Source Deep Learning Toolkit for Brain Tumour Segmentation.

- Current Directions in Biomedical Engineering*, 7(1), 30-34. doi:10.1515/cdbme-2021-1007
- Zhang, Q., Yu, H., Barbiero, M., Wang, B., & Gu, M. (2019). Artificial neural networks enabled by nanophotonics. *Light Sci Appl*, 8(1), 42. doi:10.1038/s41377-019-0151-0
- Zhong, X., Bayer, S., Ravikumar, N., Strobel, N., Birkhold, A., Kowarschik, M., . . . Maier, A. (2018). Resolve Intraoperative Brain Shift as Imitation Game. In *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation* (pp. 129-137).
- Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). *Learning deep features for discriminative localization*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Zhou, L., Wang, W., Wei, H., Song, P., Li, Z., Cheng, L., . . . Cai, Q. (2022). Clinical application of 3D Slicer combined with Sina/MosoCam multimodal system in preoperative planning of brain lesions surgery. *Scientific Reports*, 12(1). doi:10.1038/s41598-022-22549-7
- Zhu, L., He, Q., Huang, Y., Zhang, Z., Zeng, J., Lu, L., . . . Zhou, F. (2022). DualMMP-GAN: Dual-scale multi-modality perceptual generative adversarial network for medical image segmentation. *Comput Biol Med*, 144, 105387. doi:10.1016/j.compbiomed.2022.105387
- Zhu, W., Zhao, C., Li, W., Roth, H., Xu, Z., & Xu, D. (2020). *LAMP: Large Deep Nets with Automated Model Parallelism for Image Segmentation*, Cham.
- Zhu, Y., Zhou, Z., Liao, G., Yuan, K., Landman, B. A., & Išgum, I. (2021). *A novel unsupervised learning model for diffeomorphic image registration*. Paper presented at the Medical Imaging 2021: Image Processing.
- Zimmer, V. A., Ballester, M. A. G., & Piella, G. (2019). Multimodal image registration using Laplacian commutators. *Information Fusion*, 49, 130-145. doi:10.1016/j.inffus.2018.09.009
- Zoph, B., Vasudevan, V., Shlens, J., & Le, Q. V. (2018). Learning Transferable Architectures for Scalable Image Recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 8697-8710. doi:10.1109/CVPR.2018.00907