IBIMA Publishing
Communications of International Proceedings
https://ibimapublishing.com/p-articles/45AI/2025/4521125/
Vol. 2025 (20), Article ID 4521125

Optimizing U-Net for Brain Tumor Segmentation*

Joanna PIWOŃSKA and Przemysław CZUBA

Faculty of Cybernetics, Military University of Technology, Warsaw, Poland,

Correspondence should be addressed to: Joanna PIWOŃSKA, asia6051@outlook.com

* Presented at the 45th IBIMA International Conference, 25-26 June 2025, Cordoba, Spain

Abstract

Accurate segmentation of brain tumors in magnetic resonance imaging (MRI) is essential for diagnosis, treatment planning, and monitoring. Manual annotation is labor-intensive and prone to variability, which motivates the development of automated, robust segmentation techniques. Despite the popularity of U-Net in biomedical imaging, there is limited systematic evaluation of how training strategies and optimization techniques affect its performance in brain tumor segmentation tasks. This study addresses that gap by optimizing the U-Net architecture through hyperparameter tuning and data augmentation, and comparing two training paradigms: training from scratch versus fine-tuning a pre-trained model with a ResNet34 encoder. Experiments were conducted on a publicly available brain tumor MRI dataset. The methodology involved systematic variation of batch sizes, learning rates, and training epochs, combined with augmentation techniques such as rotation, noise injection, and contrast adjustments. The results demonstrate that careful hyperparameter selection significantly improves segmentation accuracy, with the model trained from scratch slightly outperforming the fine-tuned counterpart. However, the fine-tuned model converged faster, suggesting practical advantages in certain scenarios. These findings highlight the importance of tailored optimization in medical image segmentation and support the continued use of U-Net as a strong baseline, especially in data-constrained clinical contexts.

Keywords: image segmentation, deep learning, neural networks, medical diagnostics

Introduction

Brain tumor segmentation refers to the process of delineating tumor regions from brain scans, and it plays a vital role in medical imaging diagnostics. In MRI-based brain tumor diagnosis, accurate segmentation enables precise localization of lesions, determination of their boundaries, and monitoring of disease progression. This is essential for neurosurgeons and oncologists to assess tumor size, plan treatments, and track changes over time. Traditional manual segmentation by radiologists is time-consuming and subject to inter-observer variability. Automated methods, especially those based on deep learning, offer the promise of improving the accuracy, consistency, and efficiency of tumor detection and delineation. In particular, convolutional neural networks (CNNs) have shown great success in medical image analysis, enabling the processing of complex imaging data to identify anatomical structures and pathologies. Segmenting brain tumors in MRI is challenging due to variability in tumor shape, size, and appearance, but it is one of the most important applications of deep learning in oncological imaging. By producing reliable tumor masks, an automated segmentation system can support clinical decision-making and reduce the workload on medical experts.

Cite this Article as: Joanna PIWOŃSKA and Przemysław CZUBA, Vol. 2025 (20) "Optimizing U-Net for Brain Tumor Segmentation" Communications of International Proceedings, Vol. 2025 (20), Article ID 4521125, https://doi.org/10.5171/2025.4521125

U-Net is a CNN architecture that has become a standard in biomedical image segmentation since its introduction by Ronneberger et al. (2015). The U-Net's popularity stems from its unique encoder-decoder (U-shaped) architecture with skip connections, which enables it to capture both global context and fine details for precise segmentation. In the encoder (contracting path), the model learns abstract semantic features and reduces the image resolution, while in the decoder (expanding path), it recovers spatial detail and image resolution, guided by features passed directly from the encoder through skip connections. This design helps U-Net preserve important anatomical details (such as sharp tumor boundaries) that might be lost in deeper layers. U-Net was originally developed for tasks with very limited training data, demonstrating that with heavy use of data augmentation it can achieve expert-level segmentation performance using only a few dozen images [1]. Since then, U-Net and its variants (including 3D U-Net for volumetric data) have been widely applied in medical imaging. For brain tumor segmentation, U-Net is particularly suitable because MRI scans are high-resolution and often scarce; U-Net's design allows effective learning even from relatively small datasets by leveraging the context bridging and data augmentation. In summary, U-Net provides a powerful backbone for this study due to its proven efficacy in delineating complex structures like tumors while handling the challenges of medical imaging data.

The primary purpose of this work is to investigate how to optimize a U-Net architecture for brain tumor segmentation in magnetic resonance imaging (MRI) scans while emphasizing three major aims. First, the study endeavors to enhance U-Net performance through detailed hyperparameter tuning, focusing on training configurations such as the number of epochs, batch sizes, and learning rates. Second, it seeks to integrate various data augmentation methods (including rotations, brightness and contrast adjustments, and noise injection) to enrich the limited dataset and improve the generalization capabilities of the model. Third, it compares the outcomes of training the U-Net entirely from scratch with those of fine-tuning a model that uses a pre-trained ResNet34 encoder, in order to evaluate differences in convergence speed and final segmentation accuracy. By fulfilling these objectives, the research clarifies the effectiveness of carefully optimized training strategies and elucidates the role of transfer learning in medical image segmentation.

Related Work

Deep learning-based segmentation of medical images has been extensively studied in recent years. The U-Net architecture proposed by Ronneberger et al. (2015) was a breakthrough in biomedical image segmentation [1]. In their original paper, U-Net achieved excellent segmentation results on microscopy images with very small training sets by using aggressive data augmentation. The success of U-Net sparked the development of many variants and has made it a go-to architecture for tasks like organ and tumor segmentation. For example, 3D U-Net models were introduced by Çiçek, Ö. Et al. (2016) to handle volumetric data by extending U-Net's concept to three dimensions, allowing entire MRI or CT volumes to be segmented coherently. In the domain of brain tumors, challenges such as the Multimodal Brain Tumor Segmentation Challenge (BraTS) have driven researchers to develop U-Net based models that incorporate multi-modal MRI inputs (T1, T2, FLAIR, etc.) for more accurate tumor detection. Beyond U-Net, other segmentation architectures (DeepLab, Fully Convolutional Networks, etc.) have been explored, but U-Net remains particularly popular due to its balance of accuracy and efficiency in medical contexts. Recent studies continue to use U-Net for brain tumor segmentation; for instance, Kihira et al. (2022) applied a U-Net for glioma segmentation and characterization, demonstrating its effectiveness on clinical MRI data. U-Net's adaptability has also led to variations like Attention U-Net and cascaded U-Nets, which aim to refine segmentation results, but the classical U-Net often provides a strong baseline.

Another line of relevant work involves transfer learning and pre-trained networks for medical imaging. Given that acquiring large annotated medical datasets is difficult, researchers have investigated using encoders pre-trained on natural image datasets (such as ImageNet) and fine-tuning them for segmentation tasks. Pre-trained classification backbones (e.g. VGG, ResNet) can be plugged into U-Net-like architectures as the encoder. This approach has been used to bootstrap medical image segmentation models with some success, although its benefit varies (Iglovikov et al, 2018).

We investigate whether fine-tuning a pre-trained encoder accelerates convergence and enhances performance under limited data conditions, or if the domain mismatch between natural and medical images results in pre-training providing minimal benefits, with a model trained from scratch potentially achieving comparable or superior performance. This hypothesis is evaluated through our experiments.

Dataset

Medical image analysis poses significant challenges due to legal, ethical, and accessibility constraints. In this study, we utilize a publicly available dataset to ensure compliance with data protection regulations and facilitate reproducibility.

Medical data is classified as highly sensitive and subject to strict regulatory frameworks. In the European Union, the General Data Protection Regulation (GDPR) governs the processing of health-related data, requiring explicit patient consent and adherence to specific conditions for data usage (European Parliament and Council of the European Union, 2016). Ethical considerations in medical research further mandate respect for patient autonomy, beneficence, non-maleficence, and justice (Beauchamp & Childress, 2019). Ensuring patient confidentiality and obtaining informed consent are critical aspects of ethical data usage.

To circumvent legal and ethical challenges associated with proprietary medical data, we employ the Brain Tumor Dataset (Brain Tumor Dataset, n.d.), a publicly available collection of brain MRI images for tumor classification. This dataset comprises 3,064 contrast-enhanced MRI scans, categorized into three tumor types: glioma (1,426 images), meningioma (708 images), and pituitary tumor (930 images). The dataset includes images from 233 patients, although detailed provenance information is not provided. The absence of metadata on data origin raises potential concerns regarding compliance with privacy regulations and research ethics. However, its availability as a standardized dataset facilitates preprocessing and model training without additional data harmonization.

The dataset is provided in MATLAB format, organized as a structured variable (cjdata) containing fields such as tumor classification label (1 for meningioma, 2 for glioma, and 3 for pituitary tumor), patient ID, MRI image data, tumor border coordinates, and a binary segmentation mask highlighting the tumor region. This dataset serves as the foundation for training and evaluating the U-Net model for brain tumor segmentation. Its standardized format and labeled tumor regions enable effective supervised learning and performance assessment of deep learning models.

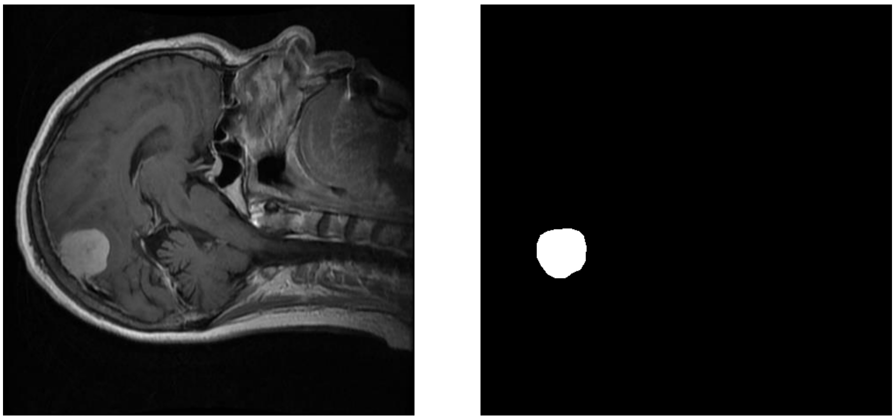

Fig. 1. Example MRI image and corresponding mask.

U-Net Architecture

The U-Net model is a fully convolutional neural network designed for medical image segmentation (Ronneberger et al., 2015). It features a symmetrical, two-part architecture that enables precise segmentation by combining high-level contextual features with fine-grained spatial details. The contracting path consists of a series of convolutional and pooling layers that progressively reduce the spatial dimensions of the input while extracting high-level features. This part of the network functions similarly to a standard convolutional neural network (CNN), capturing essential global features relevant to the overall structure of the image.

The expanding path utilizes transposed convolutional layers to restore the spatial resolution of the feature maps. This upsampling process helps reconstruct fine details and enables precise boundary delineation of segmented objects. To retain fine-grained spatial information lost during downsampling, U-Net employs skip connections between corresponding layers of the contracting and expanding paths. These connections facilitate the fusion of local and global features, enhancing segmentation accuracy and enabling the model to capture both structural context and fine details effectively (Ronneberger et al., 2015).

Methodology

Data Augmentation

Data augmentation is a technique used to artificially increase the diversity of a training dataset by introducing controlled, random modifications to existing samples. In deep learning applications, particularly when data availability is limited, augmentation allows the model to be exposed to a wider range of variations, making it more robust to changes in orientation, lighting conditions, and image quality (Shorten & Khoshgoftaar, 2019). This is especially crucial in medical imaging, where differences in scan acquisition settings and patient anatomy can significantly impact model performance.

In this study, two sets of transformations are applied. The first, used during training, includes standard operations such as resizing and normalization but also incorporates random augmentations, including rotations, horizontal and vertical flipping, brightness and contrast adjustments, and the addition of Gaussian noise. These transformations enhance the model's generalization ability by ensuring it does not become overly reliant on specific image characteristics. The second transformation set, applied to validation and test data, consists solely of resizing and normalization, ensuring consistency in evaluation without introducing artificial variability. By carefully selecting these augmentations, we aim to improve model robustness while preserving the essential features required for accurate brain tumor segmentation.

Training from Scratch

Training the U-Net model involves optimizing a loss function tailored to binary segmentation. We utilize Binary Cross-Entropy with Logits Loss (BCEWithLogitsLoss), which effectively combines binary cross-entropy with an integrated sigmoid activation function (PyTorch, n.d.). This eliminates the need for an additional sigmoid operation on the model's output and ensures accurate probability estimations for distinguishing tumors from the surrounding brain tissue.

The training process follows a mini-batch strategy, where data is processed iteratively in small batches to stabilize weight updates and optimize memory usage. Each iteration involves computing the loss for the current batch, backpropagating gradients, and updating model weights accordingly. The Adam optimizer is employed for this task, as it dynamically adjusts learning rates for individual parameters based on gradient statistics, leading to efficient convergence even in complex datasets such as medical images.

To fine-tune model performance, we experiment with hyperparameters, including the number of epochs, batch size, and learning rate. The number of epochs determines how long the model trains on the dataset, balancing between underfitting and overfitting. The batch size influences computational efficiency and gradient stability, with smaller values improving convergence at the cost of increased processing time. The learning rate, which controls the step size of weight updates, must be carefully selected to ensure stable and effective training. Too high a value can cause instability, while too low a value may slow down convergence.

To monitor generalization, we evaluate the loss on the validation set at each epoch, ensuring that the model improves its predictions on unseen data. This approach helps detect potential overfitting, where the model memorizes training data but struggles to generalize. By maintaining a balance between training efficiency and validation performance, we optimize the U-Net model for accurate brain tumor segmentation.

Fine-Tuning a Pre-trained Model

To compare a model trained from scratch with a pretrained counterpart, we employ the U-Net implementation from the segmentation-models-pytorch library (Segmentation Models PyTorch, n.d.*)*. This model supports various encoders, including ResNet and EfficientNet, which are typically pretrained on large-scale datasets such as ImageNet. These pretrained weights provide a strong initialization, capturing general visual representations that can be adapted for specific tasks.

However, direct application of a pretrained model to brain tumor segmentation yields suboptimal results, as medical images significantly differ from natural images. Without adaptation, the model fails to capture tumor-specific features and generates poor segmentations. To address this, we perform fine-tuning by freezing the encoder's weights initially, allowing the decoder to adjust to the MRI data. After stabilizing performance, we unfreeze the encoder and continue training, enabling the model to refine feature extraction for the given task. This progressive adjustment significantly enhances segmentation accuracy by leveraging prior knowledge while adapting to the unique characteristics of medical imaging data.

Results

Hyperparameter Tuning

Hyperparameters were adjusted based on the evaluation metrics to ensure optimal model performance. Dice Score was the primary measure, assessing the similarity between predicted and actual segmentation masks. This metric is particularly relevant in medical imaging, where precise boundary delineation is crucial. IoU (Intersection over Union) was also considered, as it provides a stricter measure of overlap between predicted and true regions, penalizing minor segmentation discrepancies more strongly than Dice Score. Additionally, the F1-Score was monitored to balance precision and recall, ensuring that the model effectively distinguished tumor regions while minimizing false positives. By systematically tuning hyperparameters such as the number of epochs, batch size, and learning rate, we aimed to maximize segmentation accuracy while mitigating the risk of overfitting or underfitting.

Table 1. Results of experiments for different hyperparameters

Experiment No.	Number of Epochs	Batch size	Learning Rate	Dice score	IoU	F1-score
1	40	16	5e-4	0,6841	0,6047	0,7489
2	200	8	1e-5	0,8046	0,7167	0,8379
3	50	8	1e-3	0,7484	0,6592	0,7980
4	100	32	1e-3	0,7645	0,6746	0,8109
5	150	8	5e-5	0,7959	0,7129	0,8360

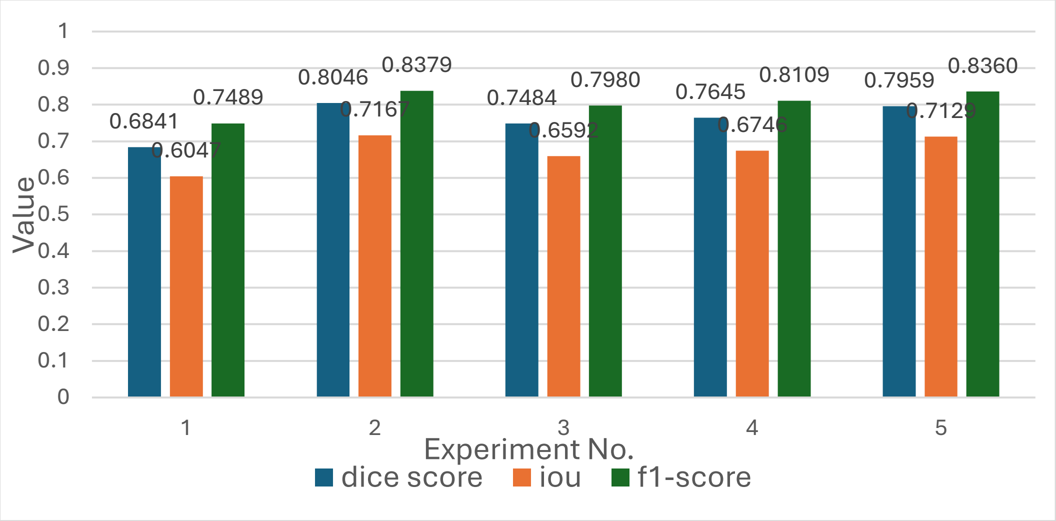

Fig. 2. Chart: Comparison of the results of the experiment metrics.

The experimental results demonstrate the significant impact of hyperparameters, including the number of epochs, batch size, and learning rate, on the quality of brain tumor segmentation using the U-Net model. The highest performance was observed in Experiment 2, where training was conducted for 200 epochs with a small batch size (8) and a low learning rate (1×10^{-5}). This configuration achieved the highest Dice Score (0.8046) and IoU (0.7167), indicating precise tumor boundary delineation and strong alignment with ground truth masks.

In contrast, Experiment 1, which used only 40 epochs, a larger batch size (16), and a higher learning rate (5×10^{-4}), resulted in the lowest segmentation quality, with a Dice Score of 0.6841 and an IoU of 0.6047. These results suggest insufficient convergence due to the limited number of training iterations, while the high learning rate likely contributed to unstable weight updates.

Experiments 3 and 5 produced comparable results, with Dice Scores of 0.7484 and 0.7959, respectively, and IoU values of 0.6592 and 0.7129. The use of a small batch size (8) in both cases contributed to stable gradient updates and improved segmentation accuracy. In contrast, Experiment 4, which employed a larger batch size (32), achieved moderate performance (Dice Score of 0.7645, IoU of 0.6746), suggesting that larger batch sizes may reduce the model's ability to capture fine-grained details.

Overall, the results indicate that longer training duration positively influences segmentation performance by allowing the model to better fit the data. A small batch size (8) contributes to improved gradient stability and feature sensitivity, while a low learning rate (1×10^{-5}) facilitates gradual, precise optimization, particularly in extended training sessions. These findings highlight the necessity of careful hyperparameter selection to achieve high segmentation accuracy, which is critical for medical applications requiring precise tumor boundary detection.

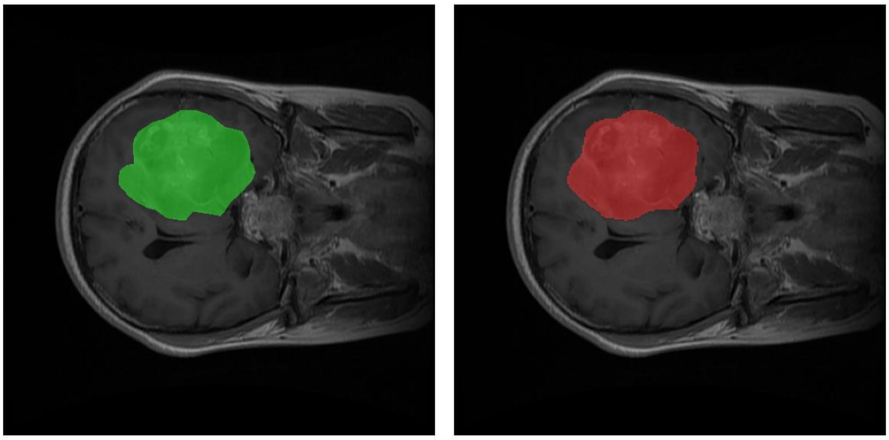

Fig. 3. Example of tumor prediction with correctly predicted tumor boundary.

The green color represents the actual tumor boundaries, while the red color indicates the predicted tumor boundaries.

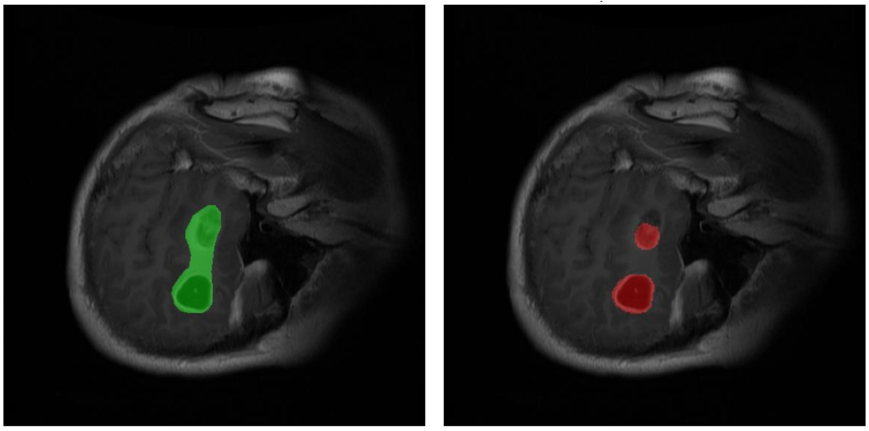

Fig. 4. Example of tumor prediction with incorrectly predicted tumor boundary.

The green color represents the actual tumor boundaries, while the red color indicates the predicted tumor boundaries.

Fine-Tuning vs Training from Scratch

A comparison between two U-Net models - one trained from scratch on brain tumor segmentation data and another fine-tuned using the segmentation-models-pytorch library - reveals notable differences in predictive performance and implementation. The model trained from scratch achieved slightly better results across key metrics, including Dice Score, IoU, F1-score, Precision, and Recall, indicating its superior adaptation to the specific segmentation task. For instance, the Dice Score for the model trained from scratch was 0.8046, whereas the fine-tuned model reached 0.7802. Similarly, the IoU values were 0.7167 and 0.6857, respectively. These results suggest that while both approaches yielded satisfactory segmentation performance, training from scratch provided a slight advantage, likely due to better feature extraction tailored to the dataset. However, fine-tuning remains a viable alternative, offering reduced computational costs and training time, particularly in scenarios with limited labeled data.

Fig. 5. Chart: Comparison of model performance metrics between the fine-tuned library model and the model trained from scratch.

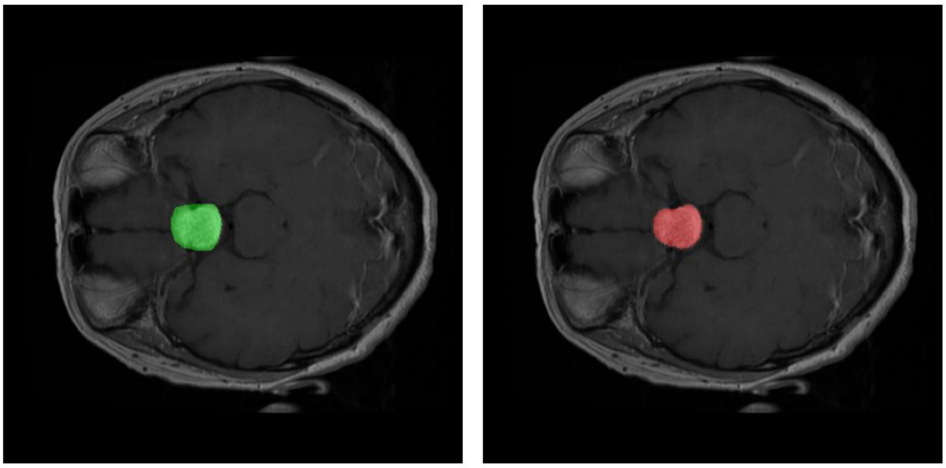

Fig. 6. Example of tumor prediction generated by fine-tuned model.

The green color represents the actual tumor boundaries, while the red color indicates the predicted tumor boundaries

Discussion

The better performance of the model trained from scratch can be attributed to its complete adaptation to the specific characteristics of medical data, allowing it to more effectively distinguish healthy tissue from tumor regions. In contrast, the fine-tuned model relied on pre-trained encoder weights derived from general tasks, such as object segmentation in natural images. While using such weights accelerates the training process, it does not always suffice for full adaptation in tasks that require medical image analysis.

Another factor influencing the results is the dataset size. The model trained from scratch was optimized for a smaller number of examples, which allowed it to better fit the problem. On the other hand, the fine-tuned model from the segmentation-models-pytorch library, with its complex and advanced ResNet34 encoder, performs better with larger datasets. In cases with limited training examples, the fine-tuned model may struggle to fully adapt, leading to slightly poorer results.

Both approaches have their strengths. The model trained from scratch provides greater control over the learning process and better utilizes a limited dataset, while the fine-tuned model significantly reduces implementation time and is more user-friendly. The comparison results show that fine-tuning can be sufficient in many cases, especially when implementation time is a critical factor. However, for data with specific characteristics, such as medical images, additional effort on a model trained from scratch may offer distinct advantages. Both models confirm that the appropriate approach to network architecture and training can achieve high-quality predictions in segmentation tasks.

Conclusions

This study developed and evaluated a U-Net-based deep learning model for automated brain tumor segmentation in MRI images. By optimizing hyperparameters, applying data augmentation, and comparing training strategies, we achieved strong segmentation results. The best-performing model, trained from scratch with 200 epochs, batch size 8, and a learning rate of 1×10^−5, reached a Dice score of 0.8046. A fine-tuned model using a pre-trained ResNet34 encoder achieved a slightly lower Dice score of 0.7802 but required fewer training epochs. These findings show that, with sufficient training and augmentation, a model trained from scratch can outperform a fine-tuned alternative. Overall, the optimized U-Net reliably segmented brain tumors with approximately 80% Dice overlap with expert annotations.

This work offers three main contributions: (1) a detailed hyperparameter study for U-Net on a brain tumor MRI dataset, highlighting the benefits of longer training with a low learning rate; (2) validation of data augmentation techniques, which significantly improved model robustness and segmentation accuracy; and (3) a comparative analysis of training from scratch versus fine-tuning, showing that custom training can match or exceed the performance of transfer learning in this context.

Future work could expand the dataset with more diverse MRI scans to improve model generalization and better utilize transfer learning. Exploring hybrid or advanced architectures such as attention U-Net, 3D U-Net, or transformers may further enhance segmentation performance. Incorporating more sophisticated augmentation—such as simulating MRI acquisition variability or using GANs to generate synthetic images—could strengthen model robustness. Post-processing methods like CRFs or ensembling strategies may also refine outputs. Lastly, extending the method to 3D segmentation would align better with clinical needs and tumor volume analysis.

References

- 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). Springer International Publishing.

- 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)
- Kihira, S., Kobayashi, A., Ueno, K., & Fujioka, H. (2022). "Reduction of Twin Boundary in NbN Films Grown on Annealed AlN." Crystal Growth & Design, 22(8), 4868–4873. DOI: 10.1021/acs.cgd.2c00415
- Iglovikov, V., & Shvets, A. (2018). "TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation." arXiv preprint arXiv:1801.05746.
- European Parliament and Council of the European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Official Journal of the European Union, L119/1.
- Beauchamp, T. L., & Childress, J. F. (2019). Principles of Biomedical Ethics (8th ed.). Oxford University Press.
- Brain Tumor Dataset. (n.d.). Figshare. Retrieved December 8, 2024, from https://figshare.com/articles/dataset/brain_tumor_dataset/1512427?file=7953679A
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data, 6*(1), 60. DOI: 10.1186/s40537-019-0197-0
- PyTorch. (n.d.). torch.nn.BCEWithLogitsLoss. Retrieved December 8, 2024, from https://pytorch.org/docs/stable/generated/torch.nn.BCEWithLogitsLoss.html
- Segmentation Models PyTorch. (n.d.). *UNet*. Retrieved December 8, 2024, from https://segmentation-modelspytorch.readthedocs.io/en/latest/docs/api.html#unet