Brain Tumor Segmentation

Tri Nguyen
Luddy School of Informatics, Computing, and Engineering
Indiana University
Bloomington, USA
trihnguy@iu.edu

Priyanka Prem Kumar
Luddy School of Informatics, Computing, and Engineering
Indiana University
Bloomington, USA
prpremk@iu.edu

Varalakshmi Perumal
Luddy School of Informatics, Computing, and Engineering
Indiana University
Bloomington, USA
vaperu@iu.edu

***Abstract*—Brain tumor segmentation from MRI is a critical task in neuro-oncology, enabling quantitative assessment, treatment planning, and disease monitoring. In this project, we investigate the performance of multiple deep learning architectures for automatic tumor segmentation using the BRATS dataset, which provides multi-modal MRI scans with expert-annotated labels. We evaluate three models: MultiResUNet, SegResNet, and a custom Swin-UNet variant. MultiResUNet offering robust multi-scale representation, SegResNet demonstrating rapid Dice convergence, and the custom Swin-UNet showing promising transformer based spatial feature learning. Our work highlights how architectural differences impact segmentation accuracy and provides insights into selecting appropriate models for brain MRI tumor analysis.**

***Index Terms*—Brain MRI, Tumor Segmentation, BRATS Dataset, Deep Learning, MultiResUNet, SegResNet, Swin-UNet**

I. Introduction

Brain tumors represent one of the most aggressive and clinically challenging neurological diseases, often requiring rapid, accurate, and reproducible evaluation to support therapeutic decision-making. Magnetic Resonance Imaging (MRI) remains the gold standard for non-invasive evaluation of brain tumors due to its superior soft-tissue contrast and its ability to distinguish key tumor components such as edema, necrotic core, and enhancement regions [1] . Despite its diagnostic value, manual delineation of tumor subregions by expert radiologists is time-consuming, labor-intensive, and inherently subjective, limiting its scalability in both clinical and research environments. This has motivated a substantial shift toward automated segmentation methods capable of producing consistent and clinically reliable results.

Deep learning approaches, especially Convolutional Neural Networks (CNNs), have led to major advances in medical image segmentation. U-Net [2] established the encoder–decoder architecture with skip connections, enabling effective learning from limited annotated data and preserving essential spatial features. While U-Net variants have become foundational models for brain tumor segmentation, accurately identifying heterogeneous tumor subregions, including the Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET), remains a difficult task due to variations in intensity distribution, morphology, and tumor progression across patients [3]. Furthermore, class imbalance, particularly the relatively small ET regions, often reduces the effectiveness of conventional pixel-wise losses.

The Brain Tumor Segmentation (Brats) benchmark provides a large-scale, multimodal MRI dataset with high-quality expert annotations, enabling systematic evaluation of segmentation algorithms [4]. Although many state-of-the-art approaches employ computationally intensive 3D architectures, these models require substantial GPU memory and long training times. In contrast, 2D slice-based approaches offer computational efficiency and remain competitive by leveraging the strong anatomical and contextual information available in individual MRI slices.

In this study, we investigate a 2D brain tumor segmentation framework based on the MultiResUNet architecture [5]. MultiResUNet enhances the conventional U-Net by introducing MultiRes blocks and residual pathways that capture multi-scale contextual information and reduce the semantic gap between the encoder and decoder representations. Our model uses three MRI modalities and generates class-specific probability maps corresponding to WT, TC, and ET segmentation masks. To mitigate class imbalance and improve structural delineation, we employ a combined Binary Cross-Entropy and Dice loss function. Experimental results on the BraTS dataset demonstrate that the proposed MultiResUNet-based approach achieves robust segmentation performance and effectively captures complex tumor boundaries.

II. Related Work

A. Unet

Segmentation became an demanding topic to predict the medical image [6]. The U-net architecture introduced the encoder-decoder structure with skip connections that has become the foundation of many current segmentation models. Its ability to learn high-level semantic features with fine-grained

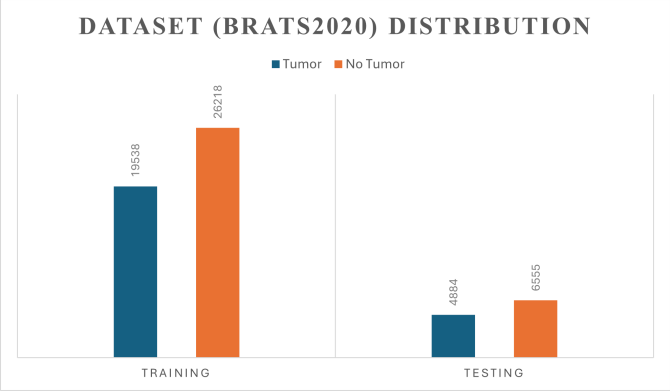

Fig. 1. Data overview.

spatial information makes it effective for task such as brain tumor segmentation on MRI images. Variants including U-Net++ [7], Attention U-Net [8], and Residual U-Net [9] have further improved feature aggregation, boundary sensitivity, and model robustness. These architectures remain widely used on the BraTS dataset due to their efficiency and strong baseline performance.

B. Convolutional Neural Networks

Convolutional Neural Networks have long been central to medical image segmentation, beginning with Fully Convolutional Networks (FCNs) [10], which introduced dense pixel-wise prediction. Subsequent models such as V-Net [11], SegNet [12], and DeepLabV3+ [13] incorporated advances including 3D convolutions, encoder–decoder refinements, dilated convolutions, and multi-scale context modules. These CNN-based approaches effectively capture local texture and structure within MRI volumes, supporting accurate detection of tumor subregions in BraTS. However, their reliance on local receptive fields limits their ability to model long-range dependencies, motivating the development of transformer-based architectures.

C. Vision Transformer

Vision Transformers (ViTs), introduced by Dosovitskiy et al. [14], inspired a shift from purely convolutional architectures to models capable of capturing global context through self-attention. In medical imaging, architectures such as UNETR [15] and TransUNet [16] leverage Transformer encoders to enhance long-range dependency modeling while retaining localized representations from CNNs. Swin-UNet [17] employs the Swin Transformer with shifted window attention to achieve scalable, high-resolution segmentation performance. These transformer-based and hybrid approaches have demonstrated state-of-the-art results on the BraTS dataset, outperforming traditional CNN architectures in capturing complex tumor structures.

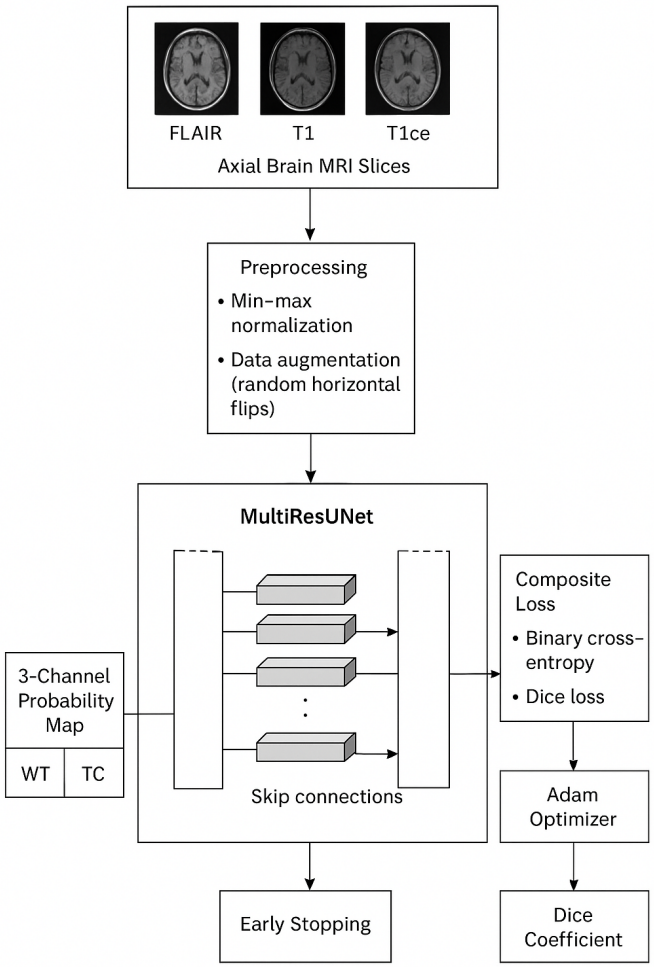

Fig. 2. Overview of the MultiResUNet architecture used for brain tumor segmentation.

III. Methods

A. Dataset

The data used in this project is from Brain Tumor Segmentation (BraTS) Challenge 2020 [18]. This challenge focused on developing state of the art methods for segmentation of brain tumor from Multimodal magnetic resonance imaging (MRI). In this dataset, the modalities that included are a) native (T1) and b) post-contrast T1-weighted (T1Gd), c) T2-weighted (T2), and d) T2 Fluid Attenuated Inversion Recovery (T2-FLAIR) volumes and were acquired with different clinical protocols and various scanners from multiple (n=19) institutions, mentioned as data contributors here. This dataset includes 369 MRI scans, which together yield 57,195 individual MRI slices. Among these slices, 32,773 show no tumor, while 24,422 contain tumor regions. Generally, we divide 80 % of this data for training and 20 % of this data for testing.

B. MultiResUNet

1) Dataset and Preprocessing: The first proposed framework performs 2D brain tumor segmentation using a

MultiResUNet-based architecture trained on axial slices extracted from the BraTS dataset. The BraTS dataset provides multimodal MRI volumes and expert-annotated ground-truth tumor masks in HDF5/NIfTI format [1]. For computational efficiency, only the first three modalities are retained. From each 3D MRI volume, the central axial slice is extracted, following practices commonly used in slice-based tumor segmentation pipelines [3], [19]. All slices are resampled to a spatial resolution of 224×224 pixels. Intensity standardization is applied independently to each modality using min–max normalization, which is known to improve cross-subject consistency and stabilize neural network training [20].

2) Ground-Truth Label Preparation and Data Loader: Ground-truth segmentation maps are converted from the original BraTS labeling scheme into a three-channel one-hot representation corresponding to Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET), following standard BraTS tumor-region definitions [18]. The dataset is split into training (70%), validation (20%), and testing (10%) subsets in this architecture. A custom Keras Sequence-based data loader enables efficient batch generation and deterministic sampling during training, consistent with best practices for large-scale medical imaging workflows [21]. Data augmentation is applied only to the training set and consists of random horizontal flips, preserving anatomical correctness while improving generalization. Validation and test subsets undergo only resizing and normalization to ensure unbiased performance evaluation.

3) Model Architecture: The segmentation model is constructed using the MultiResUNet architecture introduced by Ibtehaz and Rahman [22]. MultiResUNet extends the classical U-Net [2] by incorporating MultiRes blocks, each consisting of three progressively sized convolutional pathways whose outputs are concatenated to capture multi-scale contextual information. A residual shortcut projection ensures better feature alignment and reduces semantic discrepancy between encoder and decoder stages. The encoder path comprises stacked MultiRes blocks and max-pooling layers, while the decoder employs transposed convolutions for upsampling and utilizes skip connections to recover spatial precision. A final 1×1 convolution with sigmoid activation produces a three-channel probability map representing WT, TC, and ET classes.

4) Training Strategy: Model training uses the Dice coefficient as the loss function, a widely used strategy in medical image segmentation to directly optimize region-overlap sensitivity [23], [24]. The Adam optimizer, with default momentum parameters, is used to update network weights due to its stability and efficiency in biomedical imaging applications [25]. Early stopping is applied based on validation loss to prevent overfitting, and a ReduceLROnPlateau scheduler decreases the learning rate when performance stagnation is detected. The Dice coefficient also serves as the primary evaluation metric, consistent with BraTS benchmarks [1], to quantify segmentation quality for each tumor region.

5) Qualitative Evaluation: Qualitative evaluation is performed by generating segmentation predictions for non-empty slices from the validation and test sets. For each tumor sub-

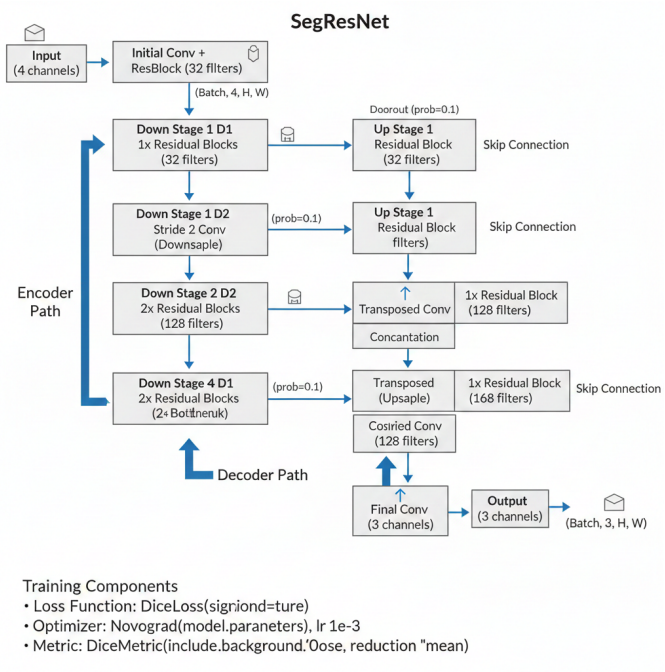

Fig. 3. SegResNet Block Diagram.

region, the input MRI slice, ground-truth mask, and predicted mask are visualized side-by-side. Such qualitative renderings are widely used in medical image analysis to inspect tumor boundary sharpness, false-positive suppression, and model robustness across diverse patient anatomies [19], [26]. These visualizations provide supporting evidence of the model's ability to delineate whole tumors, core regions, and enhancing components.

C. SegResnet

1) Encoder: The encoder consists of residual blocks that process the input at multiple scales, with a Skip connection that ensures gradient flow and stable training like in the (Fig. 3). Downsampling is achieved using strided convolutions, which enables Lower layers to capture fine-grained spatial detail. Deeper layers capture the semantic, global context of tumors or organs. Regularizes the network for small medical datasets.

2) Decoder: The decoder reconstructs the segmentation map from the encoded features. It progressively upsamples the feature maps while incorporating skip connections from the encoder. This uses trilinear interpolation followed by a residual block and avoids artifacts from transpose convolutions. Features from the encoder at the same spatial resolution are concatenated with the decoder features. Preserves fine spatial details, improving boundary accuracy. Each decoder stage has a residual block like the encoder. Ensures stable learning during upsampling.

D. Custom Swin-Unet

The model used for brain tumor segmentation is a custom Swin-Unet, which includes a Swin Transformer as an encoder

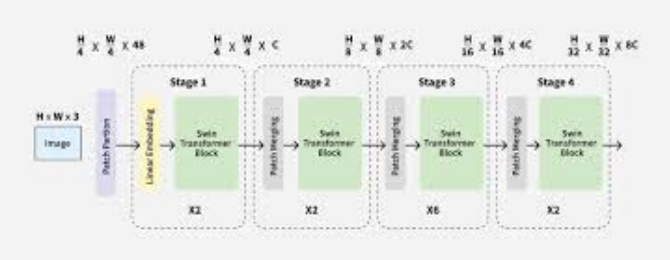

Fig. 4. Overview of the Swin Transformer architecture used for brain tumor segmentation.

and Unet for the decoder.

1) Encoder: A customized Swin-Unet is designed for brain tumor segmentation. The encoder is based on the Swin Transformer 4 which introduces the hierarchical feature extraction using window self-attention. The backbone consists of four stages with depths of (2, 2, 6, 2) and attention heads of (3, 6, 12, 24), progressively increasing the channel dimension from 96, 192, 384 to 768. Window-based multi-head self-attention (W-MSA) and shifted windows (SW-MSA) are used to efficiently model long-range dependencies while maintaining computational efficiency.

2) Decoder: The decoder follows a U-Net style upsampling pathway composed of three transposed convolution layers that reconstruct the spatial resolution by reversing the hierarchical downsampling of the encoder. Specifically, the decoder expands channel dimensions 768, 384, 192 to 96, followed by a 1×1 convolution that produces the final segmentation logits for the four tumor subregions. A final bilinear interpolation step restores the output resolution to 224 × 224 pixels. This hybrid design combines the global contextual modeling of Transformers with the spatial precision of convolutional decoders.

3) Loss and optimizer state: We optimize the network using the AdamW optimizer with a learning rate of 1×10^{-4} and weight decay of 1×10^{-5}. AdamW is well-suited for Transformer-based architectures because it decouples weight decay from the gradient update, thereby improving generalization and preventing over-regularization within the attention layers.

The model is trained for 30 epochs using mini-batch stochastic gradient descent. In each iteration, the input MRI slices and corresponding segmentation masks are passed through the network, and the Dice Loss is computed and backpropagated. Model parameters are updated via AdamW, and the average training loss for each epoch is not decreasing after 3 epochs.

IV. Results

A. Training Performance and Observations

B. Training Performance and Observations

The training and validation loss curves (Fig. 5) show a steady decrease over 50 epochs. The first few epochs have relatively high loss, reflecting the random initialization of

Fig. 5. Training and validation loss over 50 epochs.

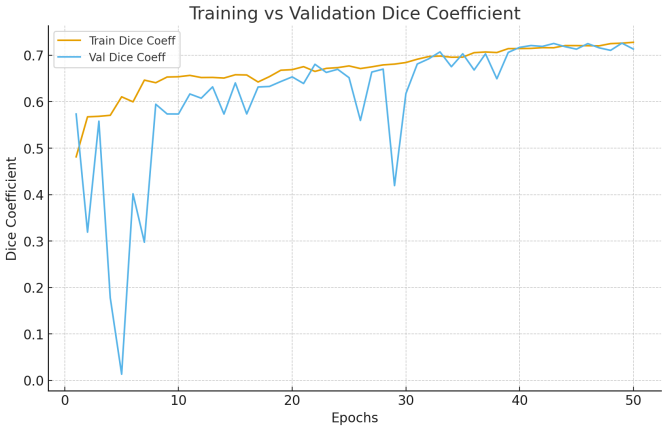

Fig. 6. Training and validation dice coefficient over 50 epochs.

network weights and the early phase of feature learning. A sharp drop in both training and validation loss occurs during the first 10–15 epochs, indicating rapid learning of low-level features of tumor and background tissue. After approximately 30 epochs, the curves plateau, suggesting that the model has reached a stable minimum and further training yields only minor improvements. The close alignment between training and validation losses indicates good generalization without significant overfitting.

The Dice coefficient curves (Fig. 6) complement the loss trends. The training Dice steadily increases from the first epoch, indicating improved overlap between predicted and ground-truth masks. The validation Dice follows a similar trajectory, confirming that performance gains on the training set translate to unseen data. The curves plateau after 30–35 epochs, suggesting that the model has captured most of the relevant patterns in the dataset. The small gap between training and validation Dice highlights stable and effective learning.

SegResNet demonstrates strong convergence in terms of Dice loss by the 6th epoch as (Fig.9 . Because the model is relatively large, we limited our evaluation to these initial epochs, which already show stable and promising performance. The

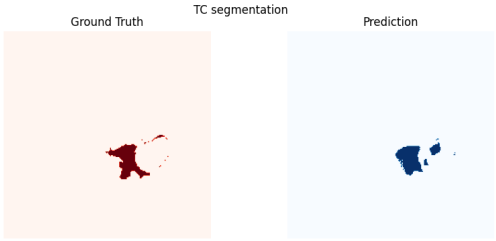

Fig. 7. TC prediction.

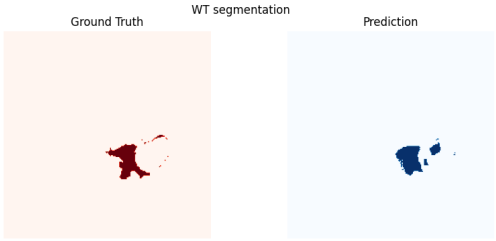

Fig. 8. WT prediction.

predicted segmentation output (Fig. 10) clearly delineates regions corresponding to different anatomical or pathological classes, based on intensity variations and the spatial features learned by the network.

C. Qualitative Evaluation

Visual inspection of predicted tumor masks demonstrates that the model accurately captures tumor boundaries and irregular shapes. Representative slices comparing ground truth and predicted masks are shown in Fig. 7 and Fig. 8, confirming the model's ability to segment non-empty slices effectively.

In Fig. 8, the Whole Tumor (WT) region is consistently captured across slices, covering all tumor components. The Tumor Core (TC) region (Fig. 7) is correctly nested within the WT region, reflecting anatomical consistency. Smaller and irregular regions, such as the Enhancing Tumor (ET), are also successfully identified, demonstrating sensitivity to fine-scale tumor features. Overall, predicted masks closely follow ground truth boundaries, avoid false positives in surrounding brain tissue, and maintain high slice-to-slice consistency.

D. Key Observations

Training Stability: Loss decreases smoothly and Dice increases steadily, with closely aligned training and validation curves, indicating stable convergence and good generalization.

Accurate Segmentation: Predicted masks closely match tumor boundaries, exclude non-tumor regions, and show consistent coverage across slices.

Sub-region Detection: WT, TC, and ET regions are correctly identified, with smaller regions like ET accurately captured and nested within larger regions.

Fig. 9. SegResNet Loss Graph.

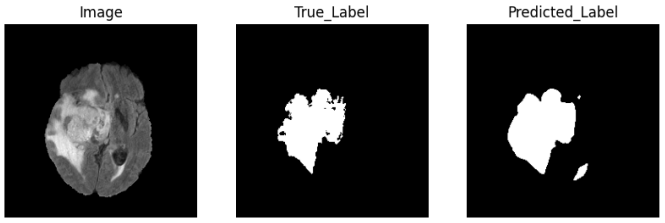

Fig. 10. SegResNet Prediction.

Generalization Across Slices: High-quality predictions are maintained across diverse tumor shapes and sizes, demonstrating effective feature learning.

V. Conclusion

This project explored three different architectures: MultiResUNet, SegResNet, and a custom Swin-UNet, as part of a broader investigation into deep learning, based brain tumor segmentation using the BraTS 2020 dataset. While all three models were implemented within the experimental pipeline, complete and validated results were obtained only for the MultiResUNet model. Accordingly, the conclusions presented here reflect the outcomes derived from MultiResUNet training and evaluation.

The MultiResUNet architecture demonstrated strong performance in segmenting key tumor subregions, including Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET). Training and validation metrics showed stable and consistent convergence, with loss decreasing smoothly and Dice scores increasing steadily throughout training. The close alignment between training and validation curves indicated that the model generalized effectively to unseen data without overfitting.

Qualitative evaluation further confirmed the model's ability to delineate tumor boundaries with high fidelity. Predicted masks accurately followed the contours of the ground-truth annotations, avoided false positives in surrounding tissue, and maintained anatomical consistency across slices. The model also showed robustness in capturing irregularly shaped and fine-scale tumor regions, demonstrating its capacity to learn discriminative features from multi-modal MRI inputs.

Overall, the findings highlight MultiResUNet as an effective and reliable architecture for 2D brain tumor segmentation tasks. Although SegResNet and Swin-UNet were part of the experimental scope, their results were not finalized for this report and can serve as a foundation for future comparative analysis. Subsequent work may extend this study by completing the evaluation of these additional architectures, integrating full 3D volumetric segmentation, and further refining the loss and optimization strategies to enhance segmentation performance across all tumor subregions.

ACKNOWLEDGMENT

We gratefully acknowledge the Brain Tumor Segmentation (BraTS) dataset, provided through the Medical Image Computing and Computer Assisted Intervention (MICCAI) community. We also acknowledge the use of Big Red 200, Indiana University's high-performance computing system. Access to Big Red 200's GPU resources enabled us to train our segmentation model efficiently and perform large-scale experimentation. We appreciate the support of IU's Research Technologies staff for maintaining this computing environment.

REFERENCES

[1] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest *et al.*, "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015.

[3] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani, and C. Davatzikos, "Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features," *Scientific Data*, vol. 4, p. 170117, 2017.

[4] Bakas, Spyridon, Akbari, Hamed, Sotiras, Aristeidis, Bilello, Michael, Rozycki, Matthew, Kirby, Justin S., Freymann, John B., Farahani, Keyvan, and Davatzikos, Christos, "Brain tumor segmentation (brats) 2018 - overview of the challenge," in *BrainLes (MICCAI Workshop)*, 2018.

[5] N. Ibtehaz and M. S. Rahman, "Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation," *Neural Networks*, vol. 121, pp. 74–87, 2020.

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015. [Online]. Available: https://arxiv.org/abs/1505.04597

[7] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," 2018. [Online]. Available: https://arxiv.org/abs/1807.10165

[8] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention u-net: Learning where to look for the pancreas," 2018. [Online]. Available: https://arxiv.org/abs/1804.03999

[9] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-net," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, p. 749–753, May 2018. [Online]. Available: http://dx.doi.org/10.1109/LGRS.2018.2802944

[10] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," 2015. [Online]. Available: https://arxiv.org/abs/1411.4038

[11] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," 2016. [Online]. Available: https://arxiv.org/abs/1606.04797

[12] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," 2016. [Online]. Available: https://arxiv.org/abs/1511.00561

[13] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," 2018. [Online]. Available: https://arxiv.org/abs/1802.02611

[14] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," 2021. [Online]. Available: https://arxiv.org/abs/2010.11929

[15] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. Roth, and D. Xu, "Unetr: Transformers for 3d medical image segmentation," 2021. [Online]. Available: https://arxiv.org/abs/2103.10504

[16] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," 2021. [Online]. Available: https://arxiv.org/abs/2102.04306

[17] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-unet: Unet-like pure transformer for medical image segmentation," 2021. [Online]. Available: https://arxiv.org/abs/2105.05537

[18] CBICA, University of Pennsylvania, "Brats 2020 guidelines, brain tumor segmentation challenge," https://www.med.upenn.edu/cbica/brats2020/data.html, accessed: 2025-12-08.

[19] K. Kamnitsas, C. Ledig, V. F. J. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3d cnn with fully connected crf for brain lesion segmentation," *Medical Image Analysis*, vol. 36, pp. 61–78, 2017.

[20] L. G. Nyúl and J. K. Udupa, "On standardizing the mr image intensity scale," *Magnetic Resonance in Medicine*, vol. 42, no. 6, pp. 1072–1081, 1999.

[21] F. Chollet, *Deep Learning with Python*, 2nd ed. Manning Publications, 2021.

[22] N. Ibtehaz and M. S. Rahman, "Multiresunet: Rethinking the u-net architecture for multiresolution biomedical image segmentation," *Neural Networks*, vol. 33, pp. 1–17, 2021.

[23] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proceedings of 3DV*, 2016.

[24] L. Gibson, W. Li *et al.*, "Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks," in *MICCAI-BraTS Workshop*, 2018.

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

[26] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Deepmedic for brain tumor segmentation," in *IEEE Med. Imag. Anal. Workshop*, 2017.