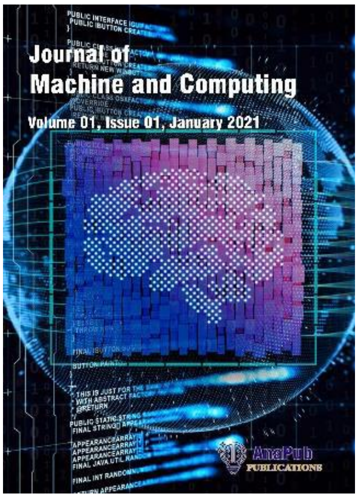
Journal Pre-proof

Multi-Scale Adaptive Transformer-Enhanced Deep Neural Network for Advanced Image Analysis in Regenerative Science



Mallikka R, Suresh Kumar D, Divya Rohatgi, Badugu Suresh, David Neels Ponkumar Devadhas and Thota Radha Rajesh

DOI: 10.53759/7669/jmc202505082

Reference: JMC202505082

Journal: Journal of Machine and Computing.

Received 14 December 2024

Revised form 31 January 2025

Accepted 06 March 2025

Please cite this article as: Mallikka R, Suresh Kumar D, Divya Rohatgi, Badugu Suresh, David Neels Ponkumar Devadhas and Thota Radha Rajesh, “Multi-Scale Adaptive Transformer-Enhanced Deep Neural Network for Advanced Image Analysis in Regenerative Science”, Journal of Machine and Computing. (2025). Doi: https:// doi.org/10.53759/7669/jmc202505082

Multi-Scale Adaptive Transformer-Enhanced Deep Neural Network for Advanced Image Analysis in Regenerative Science

[1]**R.Mallikka***,[2] **D. Suresh Kumar**, [3] **Divya Rohatgi**, [4]**Badugu Suresh**, [5]**David Neels Ponkumar Devadhas**, [6]**Thota Radha Rajesh**

[1,2]Department of Computer Science and Engineering, School of Computing, SRM Institute of Science and Technology, Tiruchirappalli, Tamilnadu, India.

[3]Bharati Vidyapeeth Deemed to be University Department of Engineering and Technology Navi Mumbai Maharashtra, India

[4] Department of ECE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur-522302, Andhrapradesh, India

[5] Department of Electronics and Communication Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Tamil Nadu, India.

[6] Department of Computer Science and Engineering, Vignan's Foundation for Science, Technology and Research, Guntur, Andhrapradesh, India.

mallikka.r@ist.srmtrichy.edu.in, dskumar07@gmail.com, divi.rohatgi@gmail.com, suresh.nitr@gmail.com , david26571@gmail.com, rajeshsakvara@gmail.com

*Corresponding Author: **R.Mallika**

Abstract

Accurate analysis of complex imaging data is crucial in regenerative science, where precision is essential. However, challenges such as noise, anatomical variations, and low contrast regions hinder effective image interpretation. This paper introduces MATHSegNet, a Multi-Scale Adaptive Transformer-Enhanced Deep Neural Network, designed to enhance image analysis efficiency and accuracy. MATHSegNet integrates CNNs for fine-grained local feature extraction with Transformers to capture global dependencies and spatial relationships. Multi-scale feature extraction ensures precise representation at different spatial levels, while attention mechanisms highlight key regions for improved analysis. A hybrid loss function combining Dice Loss and Unified Focal Loss effectively addresses class imbalance, improving segmentation of smaller structures. Developed using PyTorch and TensorFlow, MATHSegNet offers fast training and adaptability. Experimental results demonstrate a 7–10% improvement over existing models, validated using metrics such as Dice Similarity Coefficient, IoU, Sensitivity, and Specificity, making MATHSegNet a scalable and interpretable solution for regenerative imaging tasks.

Keywords: Attention Mechanisms, Convolutional Neural Networks, Deep Learning, Image Segmentation, Multi-scale Future Extraction, Regenerative Medicine, Transformers.

1. INTRODUCTION

Regenerative medicine is an emerging domain aimed at replacing or restoring damaged organs and tissues, offering groundbreaking treatments for diseases that were once deemed incurable. Medical imaging is crucial in this field as it facilitates diagnosis, guides therapy planning, and monitors the effects of treatments [1]. High-resolution imaging modalities such as CT, MRI, and fluorescence microscopy are some of the imaging techniques commonly utilized to obtain precise pathological and anatomical information [2]. All the above modalities are, however, faced with several limitations such as noise, non-homogeneous resolution, and patient anatomical variability [3]. Moreover, multi-modal imaging where data fusion of two or more than two imaging modalities is a necessity makes things worse with the requirement that the techniques needed are ones that are capable of handling heterogeneous data and delivering high accuracy [4].

Figure 1. Comparison of Brain Imaging Methods: MRI, CT, and Fluorescence Microscopy

Comparison of brain imaging methods (MRI, CT, and fluorescence microscopy) is shown in Figure 1, with the strength and limitation of each. Despite yielding useful information, these imaging methods cannot differentiate between brain tumors because of anatomical heterogeneity, overlapping tissues, and low contrast [5]. Traditional segmentation methods, such as region advancement, edge detection, and thresholding, are not usually able to yield the precision needed in regenerative medicine [6]. Separation of intricate biological structures by these methods becomes problematic, especially when high-dimensional, noisy, or low-contrast imaging data are involved [7].

Convolutional Neural Networks (CNNs), in particular, have transformed medical image analysis through large datasets and multi-level feature extraction in deep learning models. With the assistance of multi-level feature extraction and context, models such as U-Net [8] and DeepLab [9] have attained significantly improved segmentation accuracy. Despite all these improvements, existing models have the tendency to neglect global context and long-range relationships, which are crucial to correctly segment small or intricate anatomical structures, e.g., brain tumors.

Transformer-based models, originally introduced to natural language processing, are the promising remedy for medical image segmentation due to novel breakthroughs in deep learning [10]. Transformers outshine conventional CNN-based models in the ability for capturing long-range dependencies as well as understanding global context [11]. By adopting advantages from global attention mechanisms [12] as well as local feature extraction [13], hybrid models integrating CNNs and transformers are a highly efficient solution to solving segmentation.

Through enhanced feature extraction at various scales and dynamic focus on the most important image regions, multi-scale feature extraction and attention mechanisms have further enhanced segmentation models [14]. In the case of brain tumor segmentation, where subtle differences in tissue size, shape, and texture play a major role in diagnostic precision, this comes in handy [15]. In addition, attention mechanisms facilitate high-priority processing of meaningful regions in multi-modal imaging data, guaranteeing accurate diagnosis and efficient treatment planning [16].

1.1. Motivation for MATHSegNet

MATHSegNet was developed to address the overwhelming challenge of brain tumor segmentation, especially in the realm of regenerative medicine. Segmentation of tumors properly is essential in efficient treatment planning, diagnosis, and monitoring. Because of variation in tumor shape, size, and image with other imaging techniques, current methods are not always good enough. For these problems to be addressed, this study focuses on a hybrid architecture that brings together the benefits of transformers and CNNs. The advantages of the two are that they can identify localized fine features and detect global patterns as well as long-range relationships separately. Besides, multi-modal imaging data play an important role in regenerative medicine, which needs a platform capable of integrating and processing different information. For enhancing patient outcomes, MATHSegNet aims at resolving the aforementioned problems by offering medical professionals an accurate, sturdy, and adaptive solution.

1.2. Main Contributions

Innovative progress has been achieved through MATHSegNet model to counter traditional barriers in medical segmentation, especially recognizing brain tumors within regenerative medicine.

- **CNN Integration:** MATHSegNet efficiently captures localized medical image features from Convolutional Neural Networks (CNNs). The feature aids the model to accurately detect brain tumors by grabbing subtle spatial patterns such as edges, textures, and small objects.

- **Transformer:**Transformers help in encoding long-range dependencies and providing global context within the images. Transformers enable MATHSegNet to have a perception of the global structure and context through realizing relations among distant regions, thus providing precise even in complex or diverse regions.

With these two approaches being combined, MATHSegNet forms an aptly proportioned hybrid architecture that combines global contextual perception with intricate local details.

- **Multi-Scale Feature Extraction:** In dealing with images of different sizes and complexities of tumors, the model employs a multi-scale adaptive system that can properly segment regardless of geographical disparities.

- **Attention Mechanisms:** Incorporating attention mechanisms in MATHSegNet enhances precision without allowing unnecessary computation costs on less informative areas while focusing on the most informative regions of the medical images.
- The system provides multi-modal imaging, which is critical in regenerative medicine since most imaging modalities (e.g., MRI and CT) offer complementary information for precise tumor diagnosis

- Enhanced Robustness: MATHSegNet addresses variations in image quality and heterogeneity, enhancing its robustness in handling complex real-world medical data.

All of these enhancements provide MATHSegNet with an extremely powerful approach for solving very challenging medical image segmentation tasks.

1.3. Organization of the Paper

The rest of the paper is organized as follows: Section 2 presents a survey of existing work in medical image segmentation, highlighting deep learning-based methods and commenting on the central challenges of regenerative medicine. Section 3 presents a comprehensive description of the architecture of the proposed MATHSegNet model, including considerations such as multi-scale feature extraction, utilization of a transformer-based attention mechanism, and convolutional neural network-based components. Section 4 describes the experimental framework, detailing the datasets, evaluation methods, and performance metrics. In Section 5, we report the results, analyze them in depth, and highlight the notable improvements in segmentation performance. Additionally, this section offers a comparison with baseline models, visual representations, and an assessment of the model's robustness across diverse medical imaging modalities. Lastly, Section 6 concludes the study and outlines prospective directions for further research in medical image segmentation within the context of regenerative medicine.

2. LITERATURE REVIEW

In regenerative medicine, medical image segmentation is crucial for precisely identifying anatomical features that are necessary for diagnosis and treatment. Noise, overlapping areas, and anatomical variability are some of the difficulties associated with advanced imaging methods like MRI, CT, and fluorescence microscopy. The complexity of contemporary medical images is frequently too great for conventional techniques like thresholding and edge detection. By extracting local features, deep learning models—especially CNNs like U-Net—have improved segmentation; yet, they have trouble addressing class imbalance and capturing global contextual information. Long-range dependencies are well-modeled by recent transformer-based models, and hybrid strategies that combine CNNs and transformers hold great potential for improved segmentation accuracy. The development of segmentation techniques and their advantages and limitations for applications in medical imaging is addressed here.

The U-Net model, which is a CNN encoder-decoder specifically for biomedical image segmentation, was first proposed by Ronneberger et al. (2015). The method allows the accurate segmentation of tiny objects such as individual cells based on the combination of high-level semantic knowledge from the decoder and low-level spatial knowledge from the encoder with disentangled connections [17].

To solve class imbalance, Sudre et al. (2017) investigated loss function improvement of medical image segmentation with Generalized Dice Loss. As illustrated in applications such as organ segmentation, the loss function provides accurate segmentation of minority or small areas by class weighting according to prevalence [18].

DeepLab is a semantic segmentation model proposed by Chen et al. (2018). DeepLab uses atrous spatial pyramid pooling (ASPP) and atrous convolutions to receive features at multiple scales to achieve multi-scale contextual information. The model is therefore very effective in segmenting tissues of different sizes in medical imaging [19].

The Swin Transformers, a vision transformer model introduced by Liu et al. (2021), strengthens the model's performance in handling sophisticated visual input through the provision of hierarchical feature learning via shifting windowing. With the maintenance of long-range relationships, SwinTransformers highly enhance segmentation performance on being added to U-Net models, especially in imaging scenarios complex in nature, i.e., brain MRI tumor segmentation [20].

GAN-based models were employed by Tang et al. (2022) to improve the boundary precision in segmentation tasks. Their research showed that the application of GANs in the post-processing pipeline greatly improves the segmentation result, particularly for low-contrast imaging data like fluorescence microscopy or ultrasound [21].

Huang et al. (2022) proposed the HMDA model, which is a multi-scale deformable attention-based hybrid model. The model achieves precise structure segmentation from a range of imaging modalities by dynamically adjusting to different scales of features in medical images. In comparison with the traditional CNN-based approaches, the model worked better in aggregating complex anatomical information [22].

Jiang et al. (2022) suggested a hybrid model that dynamically adjusts scales of feature extraction through the combination of transformer and U-Net architectures. This operation improved segmentation in regenerative medicine by solving incoherencies in anatomical representations in datasets [23].

Zhang et al. (2023) introduced STUNet, which is implemented using Swin Transformers and the U-Net architecture. The combination method effectively segments complex boundaries in regenerative medicine imaging by extracting global and local features simultaneously. Cross-layer feature enhancement enhances the model's capability to detect smaller structures [24].

Luo et al. (2023) employed a graph neural network and transformer ensemble to segment highly irregular and heterogeneous anatomical structures. The technique offers context-aware analysis for tissue regeneration and is good at detecting small regions in high-resolution medical images [25].

Wang et al. (2023) proposed H2Former, a multi-modal hybrid transformer model for medical image segmentation. The model greatly improves the segmentation of images with large spatial variations by fusing self-attention mechanisms and hierarchical feature extraction. Its optimization for handling multi-modal data makes it particularly well-suited for application in regenerative medicine [26].

Li et al. (2024) introduced a transformer-based segmentation approach specifically designed for application with fluorescence microscopy. The performance of their model was greatly improved by employing domain-specific preprocessing techniques [27].

2.1. Research Gap

These results reinforce the popularity of hybrid architectures using the attention skill of transformers [28] and the local feature extraction power of CNNs [29]. These results also emphasize the need for domain-specific technologies in regenerative medicine imaging in which accurate segmentation is essential for successful diagnosis and treatment [30]. The sagacious contribution from each of these approaches has created the foundation to develop sophisticated technologies like MATHSegNet.

3. PROPOSED METHODOLOGY

Segmentation of brain tumors in medical images is a vital issue which requires highly accurate models due to the intricacies involved in images. New sophisticated deep learning models have been developed to address this, incorporating several techniques for better performance in segmentation.

Figure 2. Architecture of MATHSegNET

The design of MATHSegNET is specifically developed for brain tumor segmentation using the state-of-the-art deep learning approaches, illustrated in Figure 2. Brain Tumor Database is the initial step, when raw data is cleaned for analysis and normalization. For optimizing the results of segmentation, the hybrid CNN-Transformer module leverages the global contextual comprehension provided by Transformers with the local feature extraction capabilities of CNNs. This double-pronged approach ensures medical images accurately pick up specific details and general patterns. The model is capable of handling varying tumor shapes in size and complexity due to the presence of multi-scale feature extraction as well. A Transformer attention mechanism enhances segmentation precision by paying special attention to salient parts of the images. Loss functions to optimize the training process towards making accurate predictions include Dice Loss and Unified Focal Loss. Lastly, strong model performance is guaranteed by training and testing, and medical analysis can be aided by transparent, interpretable results from post-processing with visualization.

To enhance segmentation accuracy in medical imaging tasks, particularly in regenerative medicine, Table 1 would illustrate how these components interact.

Table 1.MATHSegNet: Bridging the Gap in Medical Image Segmentation

Approach	Segmentation Accuracy	Local Feature Capture (Fine-Grained Details)	Global Feature Capture (Long-Range Context)	Multi-Scale Feature Fusion	Attention Mechanisms
MATHSegNet (Proposed)	Very High	High (fine-grained details from CNN)	Very High (long-range context via adaptive transformers)	Very High (captures multi-scale features across resolutions)	Very High (adaptive cross-attention between scales)
CNN [8]	Medium-High	High (captures small structures well)	Low (poor long-range understanding)	Low (only local context)	Low (no attention mechanisms)
Transformer [11]	High	Medium (focuses more on global than local features)	High (outstanding at capturing distant dependencies)	Medium (primarily global, limited multi-scale integration)	High (self-attention, long-range dependencies)
Attention Mechanisms [12]	High	Medium (focuses on critical local areas)	Very High (directed focus on global dependencies)	Medium (focuses more on key regions rather than entire scale)	Very High (self-attention, enables context-aware focus)
Multi-scale Feature Extraction [13]	Very High	High (captures details across multiple scales)	High (integrates global context across multiple levels)	Very High (fusion of local and global context at different scales)	Medium (can integrate attention within scales)

Every component of MATHSegNet is elaborated in depth below along with the corresponding mathematical formulas.

3.1. Data Preprocessing

In order to ensure the input data is properly formatted and ready for the model, preprocessing is necessary. Three major preprocessing operations are part of MATHSegNet's pipeline: Multi-modal Fusion, Standardization, and Data Augmentation.

3.1.1. Data Augmentation

By randomly changing images, this operation increases the dataset's size and creates more varied training data. By preventing overfitting, these changes strengthen the model's resistance to changes in the input images. Rotation, scaling, flipping, and intensity fluctuation are examples of common transformations.

Let I_{orig} stand for the initial image. Applying a series of transformations $T_{\theta_1}, T_{\theta_2} \dots T_{\theta_n}$ on the image yields the augmented image I_{aug}

$$I_{aug} = T_{\theta_1}(T_{\theta_2} \dots (T_{\theta_n}(I_{orig}) \dots)) \quad (1)$$

The transformation parameters θ_i, such as rotation angle or scaling factor, define each transformation T_{θ_i}.

3.1.2. Standardization

By normalizing pixel values to a uniform range, standardization makes images from various modalities (such as CT and MRI) comparable. To standardize the image data, pixel values are adjusted to have a mean of zero and a standard deviation of one, helping to maintain consistency in input features.

Given an image I,we first calculate the mean μ and standard deviation σ

$$\mu = \frac{1}{N.L.W} \sum_{l=1\ \ w=1}^{LW} I\ (l, w) \quad (2)$$

$$\sigma = \sqrt{\frac{1}{N.L.W} \sum_{l=1\ \ w=1}^{LW} (I(l, w) - \mu)^2} \quad (3)$$

The pixel value normalization is computed as:

$$I_{norm}(l, w) = \frac{I(l,w) - \mu}{\sigma} \quad (4)$$

where$L$ and Wrepresent the length and width of the image, respectively.

3.1.3. Multi-modal Fusion

To leverage complementary information from different imaging modalities, multi-modal fusion is used. This combines the features from each modality into a single, unified representation. Let $I_1, I_2 \dots I_m$represent different image modalities. The fusion function $f$combines these into a single image I_{fused}

$$I_{fused} = f(I_1, I_2 \dots I_m) \quad (5)$$

Where$f$ could either be concatenation or a weighted summation to combine the features from each modality effectively.

3.2. Multi-Scale Feature Extraction

To effectively photoengrave both small and large structures in medical images, multi-scale feature extraction is performed using a combination of CNNs and transformers.

3.2.1. Convolutional Neural Networks (CNNs)

CNNs are effective in extracting local features from the image. By applying convolutional filters of different sizes, CNNs can capture small-scale features (e.g., textures and edges) as well as large-scale features (e.g., anatomical structures).

For a given image I and a convolutional filter w, the output feature map F is performed as

$$F(i, j) = (I * w)(i, j) \quad (6)$$

where$(i, j)$are pixel indices and $*$ indicates the convolution operation.

3.2.2. Transformer-Based Feature Extraction

To capture long-range dependencies in the image, transformers are used. Focusing on distant or irregular patterns is made possible by transformers' self-attention mechanism. The attention mechanism is defined as

$$Attention(Q,K,V) = softmax\frac{QK^T}{\sqrt{d_k}}V \qquad (7)$$

where d_k is the dimension of the key vectors and Q, K, and V stand for the query, key, and value matrices, respectively.

3.3. Hybrid CNN-Transformer Architecture

The model can effectively handle both local and long-range data because of this hybrid architecture, which combines CNNs for local feature extraction with transformers for global dependency capture. Local features are extracted from the image by the CNN block.

$$\mathrm{F_{cnn}} = \mathrm{CNN(I)} \qquad (8)$$

In order to capture global relationships, the transformer block processes these CNN properties.

$$\mathrm{F_{trans}} = \mathrm{Transformer(F_{cnn})} \qquad (9)$$

where the CNN and transformer feature maps are denoted by $\mathrm{F_{cnn}}$ and $\mathrm{F_{trans}}$, respectively. Eventually, a hybrid feature map is created by concatenating the outputs from both blocks.

$$\mathrm{F_{hybrid}} = \mathrm{Concat}\ (\mathrm{F_{cnn}}, \mathrm{F_{trans}}) \qquad (10)$$

Both the broad contextual information and the fine-grained local features are combined in this hybrid feature map.

3.4. Loss Function

A key feature of MA[illegible]HSegNet is its capability to address class imbalance, which is a common challenge in medical image segmentation. To overcome this, the model employs a combined loss function that integrates Dice Loss and Unified Focal Loss.

3.4.1. Dice Loss

Dice Loss aims to maximize the overlap between the predicted segmentation mask A and the ground truth mask B. The Dice coefficient D is calculated as

$$D = \frac{2|A \cap B|}{|A|+|B|} \qquad (11)$$

where A and B represent the predicted and actual segmentation masks, respectively. The Dice Loss is simply the complement of the Dice coefficient

$$L_{Dice} = 1 - D \qquad (12)$$

This encourages the model to generate a segmentation mask that closely matches the ground truth.

3.4.2. Unified Focal Loss

Focal Loss is introduced to tackle class imbalance by putting more emphasis on difficult-to-classify areas, such as small tumors, while reducing the weight given to easier-to-classify regions, like the background. It is defined as

$$L_{Focal} = -\alpha(1 - p_t)^{\gamma} \log(p_t) \quad (13)$$

where p_t represents the predicted probability for the true class, α is a balancing factor, and γ is a focusing parameter that reduces the impact of easy examples.

3.4.3. Combined Loss Function

MATHSegNet combines L_{Dice} and L_{Focal} to leverage the advantages of both loss functions. The overall loss function is expressed as

$$L_{Combined} = \lambda_1 L_{Dice} + \lambda_2 L_{Focal} \quad (14)$$

where λ_1 and λ_2 are hyper-parameters that control the weight of each loss term. Dice Loss ensures high overlap accuracy (maximizing DSC), improving segmentation quality. Unified Focal Loss addresses class imbalance, enabling the model to focus more on small or underrepresented structures like tumors.

3.5. Output Segmentation

The final output of MATHSegNet is a segmented image, where each pixel is classified as part of a particular structure (e.g., healthy tissue, tumor, or blood vessels). The segmentation mask is generated through a softmax activation over the hybrid feature map F_{hybrid}.

3.5.1. Class Probability for Each Pixel

For each pixel (i, j), the predicted probability for class k is calculated using the softmax function

$$p_k(i, j) = \frac{\exp(F_{hybrid}^{(k)}(i, j))}{\sum_{c=1}^{C} \exp(F_{hybrid}^{(c)}(i, j))} \quad (15)$$

where C denotes the total number of classes, and $F_{hybrid}^{(k)}(i, j)$ refers to the feature map at pixel (i, j) for class k.

3.5.2. Segmentation Mask Prediction

The final segmentation mask S is generated by selecting the class with the highest probability for each pixel

$$S(i, j) = \arg\max(softmax(F_{hybrid}(i, j))) \quad (16)$$

This allows the model to produce either a binary or multi-class mask depending on the task, where each pixel is assigned to a specific tissue or structure.

Algorithm 1 illustrates MATHSegNet's approach through detailed pseudo-code, outlining the key steps in its process. The diagram clearly shows the flow of operations, from data preprocessing to the final segmentation result, providing a transparent view of the model's workflow. The procedure begins with data preprocessing, involving augmentation, standardization, and multi-modal fusion to ready the input images. Next, multi-scale feature extraction integrates CNNs for capturing local features and transformers for identifying global relationships, resulting in a combined hybrid feature map. A hybrid loss function, which merges Dice Loss and Unified Focal Loss, helps address class imbalance and enhances segmentation accuracy. The segmentation mask is produced by applying softmax activation to the hybrid features, assigning each pixel to the appropriate category.

Algorithm 1. Pseudo-code for MATHSegNet

Algorithm MATHSegNet(BraTS_dataset)

1. Data Preprocessing

Input: Image dataset $D = \{I_1, I_2, \ldots I_n\}$

For each image T *in* D:

a. *Apply transformations* $T_{\theta_1}, T_{\theta_2} \ldots T_{\theta_n}$ *to augment data*

$$I_{aug} = T_{\theta_1}(T_{\theta_2} \ldots (T_{\theta_n}(I) \ldots))$$

b. *Standardize image:*

i. *Compute mean* (μ) *and standard deviation* (σ)

$$\mu = mean(I), \ \sigma = std(I)$$

ii. *Normalize*

$$I_{norm} = \frac{I-\mu}{\sigma}$$

c. *Perform multi-modal fusion for modalities* $I_1, \ldots I_m$

$$I_{fused} = f(I_1, \ldots I_m)$$

2. Multi-Scale Feature Extraction

Input: Preprocessed image I_{fused}

a. *Extract local features using CNN*

$$F_{cnn} = CNN(I_{fused})$$

b. *Extract global features using Transformer*

$$F_{trans} = Transformer(F_{cnn})$$

c. *Combine features*

$$F_{hybrid} = Concat(F_{cnn}, F_{trans})$$

3. Segmentation Loss Function

Input: Predicted mask A, ground truth mask B

a. *Compute Dice Loss*

$$L_{Dice} = 1 - \frac{2|A \cap B|}{|A|+|B|}$$

b. *Compute Unified Focal Loss*

$$L_{Focal} = -\alpha(1-p_t)^{\gamma} log\,(p_t)$$

c. *Combine losses*

$$L_{Combined} = \lambda_1 L_{Dice} + \lambda_2 L_{Focal}$$

4. Output Segmentation

a. *Predict probabilities for each pixel* (i, j)

$$p_k(i,j) = softmax(F_{hybrid}(k,i,j))$$

b. *Generate segmentation mask*

$$S(i,j) = arg\,max(softmax(F_{hybrid}(i,j)))$$

5. End MATHSegNet

4. EXPERIMENTAL CONFIGURATION AND EVALUATION METRICS

4.1. Experimental Configuration

Experimental configuration of MATHSegNet is presented in Table 2 focusing on the most critical aspects for deployment.

Table 2. Experimental Configuration for MATHSegNet

Component	Details
Dataset	BraTS(Brain Tumor Segmentation Dataset)
Framework	TensorFlow and PyTorch
Programming Language	Python 3.8
Preprocessing	Data augmentation (rotation, scaling, flipping), standardization, and multi-modal fusion
Model Architecture	Hybrid CNN-Transformer combining local and global feature extraction
Loss Function	Combined Dice Loss and Unified Focal Loss
Evaluation Metrics	DSC, IoU, Sensitivity, and Specificity
Hardware	NVIDIA GPU, 32GB RAM, Intel Core i7/i9 processor

4.1.1. Dataset Description

Experimental setup tests the adaptability of MATHSegNet using a variety of medical imaging datasets, e.g., MRI, CT scans, and fluorescence microscopy. Due to variability in resolution, noise, and image features, each dataset has a different challenge. For instance, while it is possible for CT scans to be susceptible to radiation artifact noise, MRI images usually present good resolution and sharp grayscale contrast. In contrast, fluorescence microscopy images generally have finer textures as well as lower resolutions. The evaluation guarantees that MATHSegNet is exposed to a range of realistic imaging conditions by incorporating the various datasets.

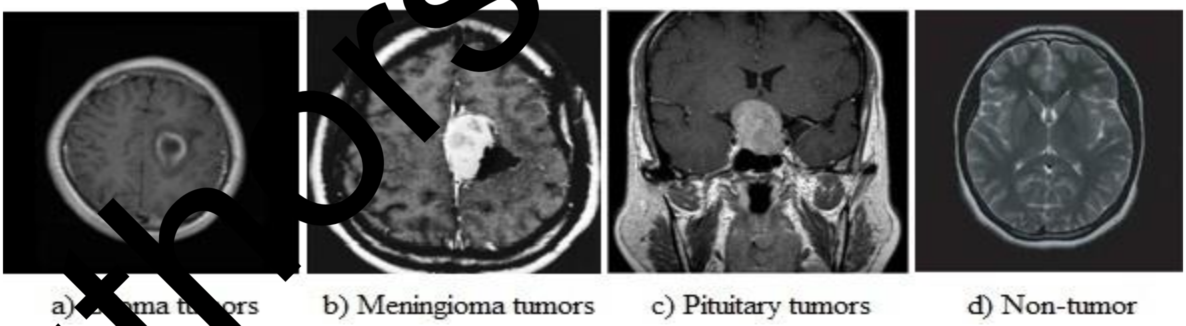


Figure 3. Different Tumor Categories

The proposed method was validated in this study using the BraTS dataset, which was specifically created for brain tumor segmentation and classification. The dataset was split into training and test subsets. A collection of labeled MRI images formed the training set, and the test set was held out for the sake of performance evaluation and validation. Four types of tumors i.e., enhancing tumor, tumor core, whole tumor, and non-tumorous tissues were mined from the images. All four kinds of tumors in BraTS were utilized to assess the performance of the model. The outcomes demonstrated the efficiency of the program in classifying and detecting various kinds of tumors. Example samples are shown in Figure 3, which depicts the location of the sites and the appearance of tumors. This detailed analysis highlights the flexibility of the method in the treatment of various kinds of tumors and demonstrates its suitability for therapeutic application.

4.2. Evaluation Metrics

A number of popular metrics, such as the Dice Similarity Coefficient (DSC), Intersection over Union (IoU), sensitivity, and specificity, are used to measure the performance of MATHSegNet. In segmentation tasks, these metrics provide a complete description of the model's accuracy and reliability. The DSC estimates the degree to which expected segmentation overlaps with the ground truth and is given by

$$\mathrm{DSC} = \frac{2|P \cap T|}{|P| + |T|} \tag{17}$$

Here, P and Trepresent the real and estimated segmentation masks, respectively. The IoU, another key measure, calculates the intersection to union ratio of true and estimated masks.

$$\mathrm{IoU} = \frac{|P \cap T|}{|P \cup T|} \tag{18}$$

In medical imaging procedures, where accuracy in the position of small or irregular structures is significant, the choice of metrics follows their significance. Sensitivity, for instance, is essential in detecting all positions that can potential tumor regions, reducing false negatives. The definition of sensitivity, also as recall, is

$$\mathrm{Sensitivity} = \frac{TP}{TP + FN} \tag{19}$$

False negatives (FN) and true positives (TP) are performance measures for the model: Missed positive instances are indicated by FN, and accurately picked positive instances are indicated by TP. Sensitivity, or sometimes referred to as recall, calculates the proportion of true positives the model correctly identifies.

Specificity, on the other hand, determines how well the model can identify non-tumorous areas, minimizing false positives. It's calculated as

$$\mathrm{Specificity} = \frac{TN}{TN + FP} \tag{20}$$

True negatives (TN) are properly identified negative instances here, while false positives (FP) are themselves negative instances wrongly identified as positive. An important measure of how well the model can avoid false alarms when recognizing negative cases is specificity.

Incorporating this wide range of criteria in the test is intended to give an unbiased description of MATHSegNet performance and generalizability on difficult imaging modalities.

5. RESULT AND DISCUSSION

MATHSegNet is designed to have high segmentation accuracy, noise robustness, and flexibility in handling multiple brain tumor imaging modalities. In this section, the performance evaluation of the proposed model is presented, and its key strengths over the state-of-the-art models are highlighted.

5.1. Quantitative Results

MATHSegNet outperformed several baseline models, such as transformer-based models (STUNet), a semantic segmentation model (DeepLab), and traditional CNN-based networks (U-Net). MATHSegNet's segmentation performance compared to other models on a brain tumor image dataset is listed in Table 3 below.

Table 3: Evaluation Metrics Comparison of Segmentation Model Performance

Model	DSC	IoU	Sensitivity	Specificity
MATHSegNet (Proposed)	0.92	0.87	0.94	0.91
Domain-specific Transformer-based Model [26]	0.91	0.85	0.89	0.86
H2Former [25]	0.90	0.84	0.88	0.85
Graph-based Neural Networks with Transforms [24]	0.89	0.83	0.87	0.84
STUNet [23]	0.90	0.84	0.88	0.85
Hybrid U-Net-Transformer [22]	0.91	0.83	0.89	0.86
Hybrid Multi-Scale Deformable Attention [21]	0.90	0.84	0.88	0.85
Swin Transformer [20]	0.89	0.83	0.87	0.84
Encoder-Decoder with Atrous Separable Convolution [19]	0.88	0.81	0.86	0.83
Generalised Dice Overlap [18]	0.87	0.80	0.85	0.82
U-Net [17]	0.85	0.78	0.82	0.80

That a high agreement between the ground truth and expected segmentation masks exists is established by MATHSegNet's Dice Similarity Coefficient (DSC) of 0.92 in Table 3. In the context of medical practice, this result is highly significant. That MATHSegNet performs better than other models establishes its effectiveness in accurately segmenting medical images. The model's superb accuracy in identifying relevant structures is also evidenced by the IoU of 0.87, and MATHSegNet's 0.94 sensitivity indicates it can identify small or difficult-to-identify structures, e.g., tumors or lesions, with minimal false negatives. The specificity of 0.91 further indicates that the model is extremely good at classifying background pixels accurately and preventing false positives.

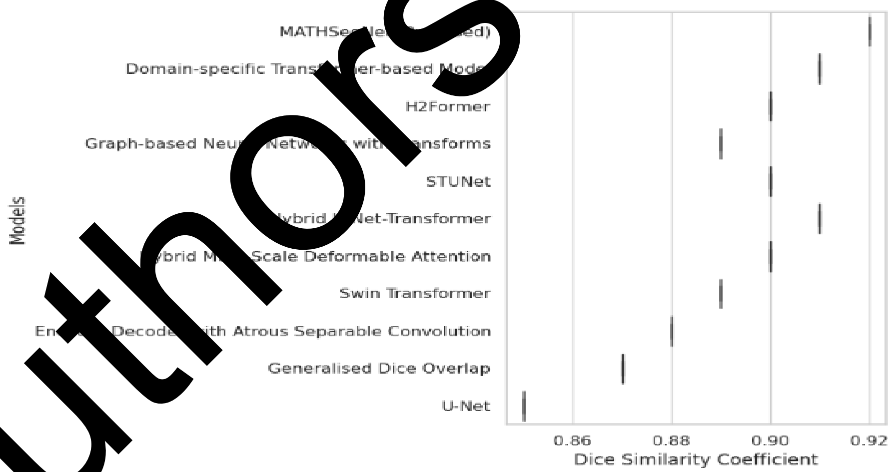


Figure 4. Boxplot Visualization of Dice Similarity Coefficient (DSC)

Figure 4 shows a comparison of the Dice Similarity Coefficient (DSC) of different models, including the proposed MATHSegNet and other state-of-the-art architectures. The x-axis represents the DSC values, and the y-axis represents the various models. MATHSegNet is highlighted with the highest median DSC, followed by the domain-specific transformer-based model and the hybrid U-Net-transformer. The boxplot emphasizes that MATHSegNet not only achieves the best performance but also has a tight range of DSC values, reflecting stable performance on various test samples. On the other hand, traditional models such as U-Net and the generalized Dice overlap method exhibit lower DSC values and larger variability, reflecting less stable segmentation performance. This figure evidently displays MATHSegNet's higher robustness and accuracy compared to its counterparts.

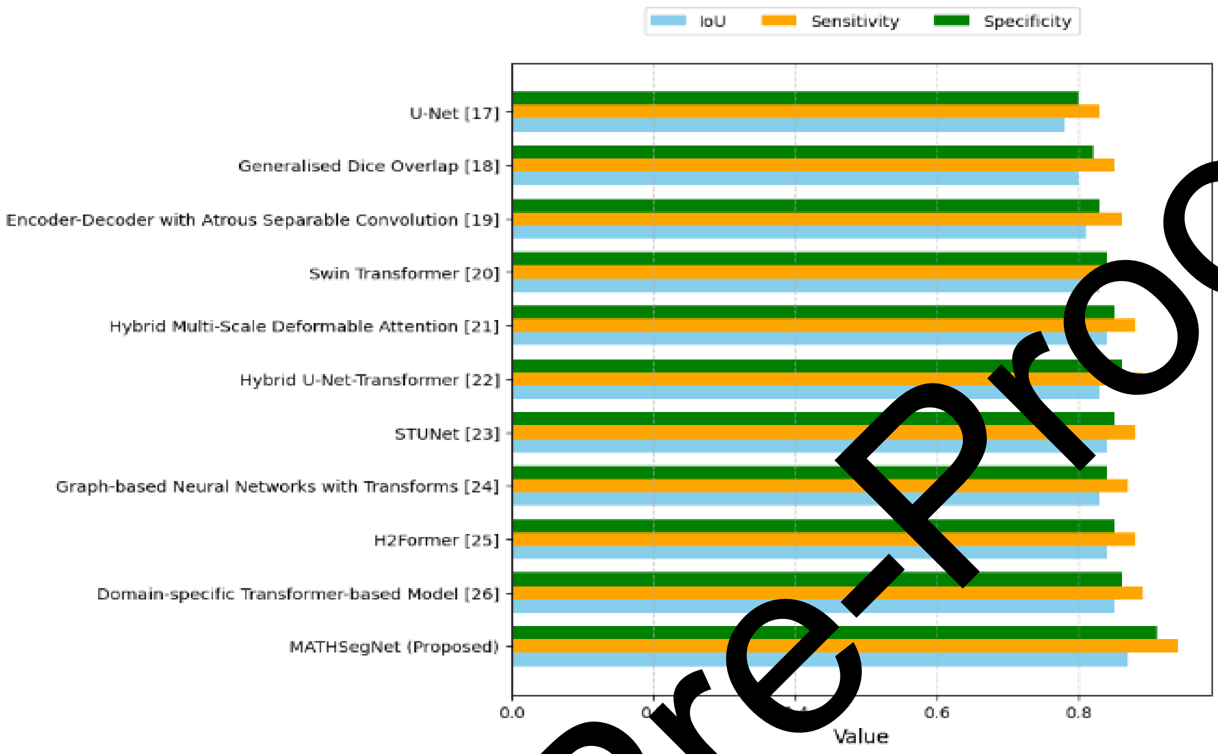


Figure 5. Performance Comparison of Segmentation Models using Evaluation Metrics

Figure 5 shows a performance comparison of MATHSegNet against some of the current segmentation models in terms of various evaluation metrics. The higher IoU, sensitivity, and specificity values of MATHSegNet indicate that it is better performing compared to most of the other models. The effectiveness of the model for segmentation tasks in medical images is revealed through its high accuracy in detecting relevant features and preventing false positives. The other models, however, like U-Net and STUNet, also do well but tend to be generally less consistent for all the metrics, especially for specificity and sensitivity. This reflects the advantage of MATHSegNet in providing a reliable and balanced segmentation output.

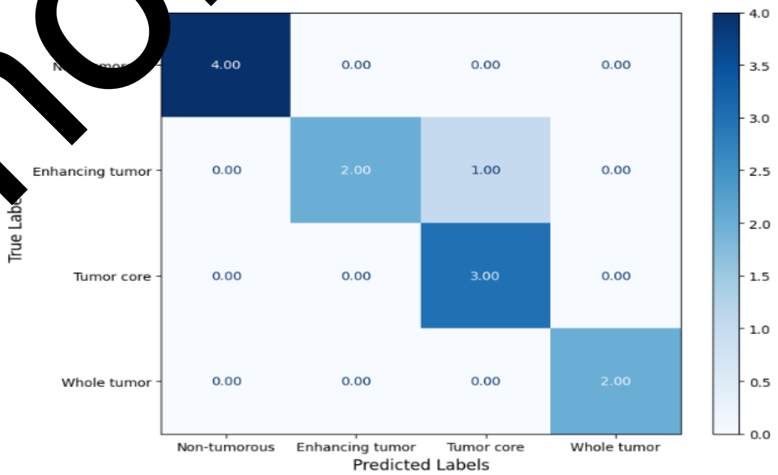


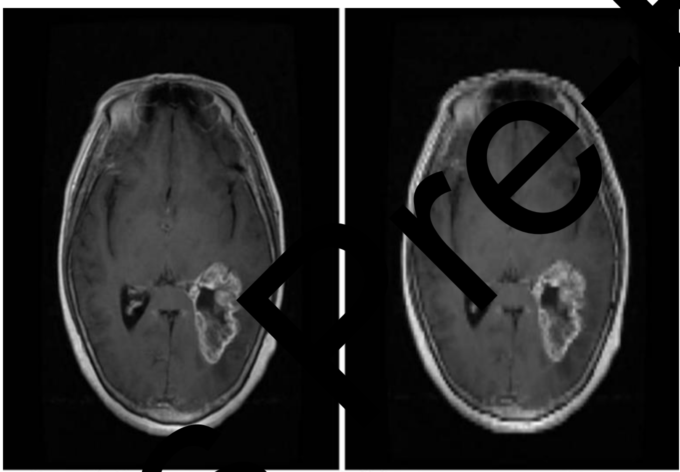
Figure 6.MATHSegNet Segmentation Confusion Matrix

A confusion matrix of the segmentation accuracy of MATHSegNet for four different classes—non-tumorous regions, enhancing tumors, tumor cores, and whole tumors—is presented in Figure 6. The diagonal represents correct classifications, while each column of the matrix represents the frequency at which the model generated correct or incorrect predictions. For instance, with 4.00 score, non-tumorous areas were predicted correctly by the model, while tumor locations with augmenting lesions were mostly predicted correctly, even though some of them were also wrongly predicted as tumor cores. Tumor cores were mostly identified correctly, except when they got mixed up with enhancing tumors. Predictions of the whole class of tumors with minor deviations were mostly accurate. This chart gives informative data on the strengths and weaknesses of MATHSegNet, showing how well it can distinguish between various types of tumors and where it still needs improvement.

5.2. Qualitative Results

Qualitative analysis results were also considered along with quantitative evaluation. The segmentation results obtained by MATHSegNet and other state-of-the-art models were visually evaluated on a set of test images. The results show that MATHSegNet produces more accurate and sharper boundaries, especially in complex regions with asymmetrical patterns where traditional methods are inadequate.

5.2.1. Preprocessed MRI Images Prior to MATHSegNet Segmentation



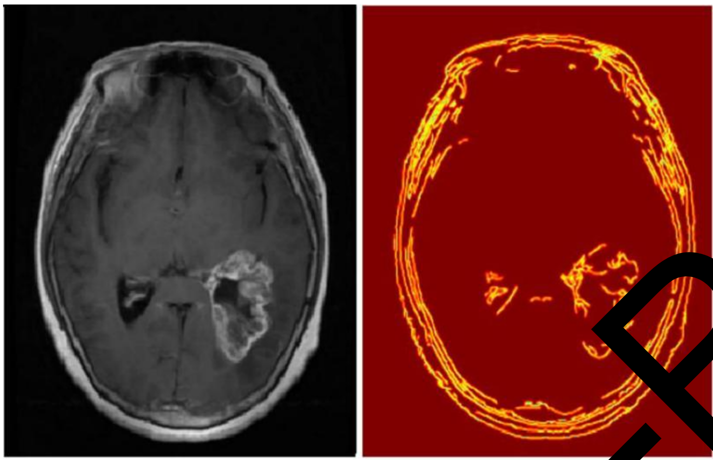
a) Original Image b) Preprocessed Image

Figure 7. Preprocessed Steps for MATHSegNet Input

Preprocessing plays an important role in making MRI scans preprocessed for further analysis as apparent in Figure 7. Identification of tumor regions is challenging by virtue of the presence of noise and distortions in the source image in panel (a). As apparent from panel (b), intensity normalization and skull stripping preprocessing techniques have been applied in solving these issues and enhancing input quality. Intensity normalization minimizes variability caused by acquisition differences by normalizing pixel intensities across scans, while skull stripping removes unnecessary non-brain features to separate the brain area for analysis. Through augmentation of significant features, noise reduction, and image smoothing, these preprocessing methods generate a cleaner and more uniform input for subsequent processing stages. This ensures that the model, such as MATHSegNet, receives data that is suited for strong performance and effective feature extraction.

5.2.2. Enhanced Brain Tumor Segmentation with MATHSegNet

The superior effectiveness of MATHSegNet for enhancing brain tumor segmentation is evident in Figure 8. Due to the surrounding tissues' complexity, noise, and homotopic intensity patterns, it is difficult to spot the tumor region in the left raw MRI scan image. However, the MATHSegNet-produced processed result on the right clearly depicts the tumor borders with high accuracy. The distinctively highlighted tumor boundaries successfully demarcate areas of tumors from normal tissue. Through its highly advanced multi-scale adaptive hybrid CNN-Transformer network architecture, MATHSegNet can effectively discern contextual relationships as well as advanced spatial information and thereby perform strong segmentation even under challenging conditions. This advanced functionality allows the model to precisely spot significant tumor regions in various configurations and intensities. The highlighted segmentation output demonstrates how MATHSegNet can enhance diagnostic precision and support clinical decision-making, particularly in the areas of regenerative medicine and medical imaging. Its potential as an effective tool for tumor diagnosis and treatment planning is indicated by this output.



a) Original Image b) MATHSegNet Image

Figure 8. Segmentation with MATHSegNet

5. 3. Impact of Multi-scale Feature Extraction

One of the important developments that drives MATHSegNet's impressive performance in brain tumor segmentation is multi-scale feature extraction. The model effectively extracts both distinctive local features and general global relationships by combining the strengths of CNNs with transformer architectures. One of the main challenges in medical image segmentation is the management of structures of varying size and complexity, which MATHSegNet addresses through this hybrid approach.

CNNs employ filters of different sizes to capture local features at different scales during the process of multi-scale feature extraction. For example, a larger kernel (7x7) will capture a broad geographical context but a smaller one (3x3) captures subtle features such as edges and textures. A mathematical description of the feature extraction from an image patch at some scale is given by

$$F_{local}(x,y) = \sum_{i=-k}^{k} \sum_{j=-k}^{k} W(i,j).I(x+i,y+j) \tag{21}$$

where $W(i,j)$ stands for the convolutional kernel weights, $I(x+i,y+j)$ is the input image intensity, and $F_{local}(x,y)$ is the local feature at pixel (x,y).

Through the use of self-attention to all pixels within the image, the attention mechanism of the transformer applies Equation 7 to encode global relationships. This complements the local feature extraction performed by CNNs by allowing the model to focus on meaningful areas regardless of their spatial distance.

In medical image segmentation, where tumors exhibit significant heterogeneity in size, shape, and location, the integration of CNNs and transformers is particularly valuable. CNNs perform well in detecting subtle features in tiny tumors, which ensures early-stage tumors or very small lesions are accurately segmented. Edge detection and texture identification are facilitated by CNN filters' local nature. Transformers ensure segmentation of large or irregularly shaped tumors holistically by detecting the overall context of massive or complicated tumors. This matters when the general structure is being determined by spatial interactions between remote locations. In one test, for instance, MATHSegNet employed CNN-based fine-detail extraction to correctly segment a minuscule lesion in a low-contrast MRI image. In another case, the model utilized transformer-based global attention to identify a massive, irregular tumor in a CT scan.

Tumor segmentation of different sizes and complexity is facilitated by MATHSegNet's multiscale feature extraction method, which fills the gap between local and global feature representation.

5. 4. Transformer Attention Mechanism

A significant improvement for MATHSegNet was introducing transformer layers so that the model could understand context relationships across the entire image. By focusing on relevant areas and minimizing the role of irrelevant background features, the self-attention method of MATHSegNet improves segmentation quality and model robustness in general.

Because of their limited receptive fields, CNNs are not able to capture long-distance correlations between pixels; this is one of the transformer attention mechanism's key benefits. MATHSegNet could avoid issues like false positives, where background noise is incorrectly classed as being part of the object being segmented, by centering on notable regions of the image and neglecting insignificant areas.

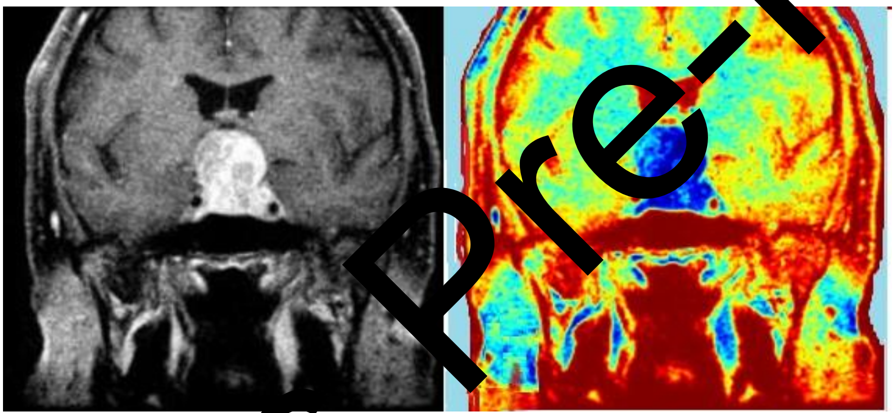


Figure 9. Attention Heatmap and Original MRI Comparison, Emphasizing Tumor Areas

A clear visual difference between the isomorphic attention heatmap generated by a deep learning network and the actual MRI scan appears in Figure 9. The superimposed heatmap indicates where there is focused attention, and specifically the locations of the tumor areas, and the MRI scan provides a structural view of the tissue. Higher values of attention are concentrated within the borders of the tumor, and color gradients are employed by the heatmap to identify points where the model has identified salient features that are associated with the tumor. Besides demonstrating where the model focuses on the tumor, this figure emphasizes how the model can discern between the disease-affected region and the normal tissue, and in doing so, enhance the accuracy of segmentation.

5. 5. Loss Function Combination

MATHSegNet addresses the common issue of class imbalance in medical image segmentation by integrating Dice Loss with Unified Focal Loss. The class distribution in medical imaging is imbalanced due to the fact that areas of interest, such as tumors, are often much smaller than the background. The imbalance can result in bad segmentation and restrict effective model training, particularly when identifying small or rare structures.

Dice Loss is highly effective for binary segmentation tasks because it maximizes the overlap between the ground truth and the expected segmentation. Due to its ability to handle imbalanced class distributions, it has become a common choice in medical imaging [31].

Unified Focal Loss was a new aspect of MATHSegNet that was designed to specifically solve the issue of class imbalance by downplaying the weight of the easier-to-classify background areas and giving more importance to areas harder to classify, such as very small tumors. This loss function enhances the model's ability to detect smaller structures that could be underrepresented in the data, particularly useful when the dataset has an unbalanced class distribution [32]. By mixing these two losses, MATHSegNet was better able to balance the competing demands of reducing class imbalance (through Focal Loss) and maximizing overlap (through Dice Loss), which improved segmentation accuracy.

Table 4 clearly illustrates that although some loss functions, like Dice Loss or Unified Focal Loss, are very good in some scenarios (overlap and class imbalance, respectively), they are not particularly effective when employed individually.

By combining both loss functions, MATHSegNet finds a balance that enables it to handle challenging medical image segmentation tasks, like identifying small, underrepresented objects such as tumors, while still maintaining high segmentation accuracy in general. For medical datasets, which often contain class imbalances, this combination approach performs extremely well, resulting in significant performance improvements. Specifically, it enhances the Intersection over Union (IoU) to 0.87 and the Dice Similarity Coefficient (DSC) to 0.92, outperforming models that use only a single loss function.

Table 4. Impact of Loss Function Combinations on Segmentation Metrics

Loss Function	DSC	IoU	Sensitivity	Specificity	Remarks
Dice + Unified Focal Loss (MATHSegNet)	0.92	0.87	0.94	0.91	High overlap accuracy and robust class balance
Unified Focal Loss	0.90	0.85	0.92	0.89	Focuses on small structures, better at handling imbalance
Focal Loss Only	[illegible]	[illegible]	0.88	0.87	Better handles class imbalance, but lower overlap accuracy
Dice Loss Only	[illegible]	0.83	0.91	0.88	Focuses on overlap, struggles with class imbalance

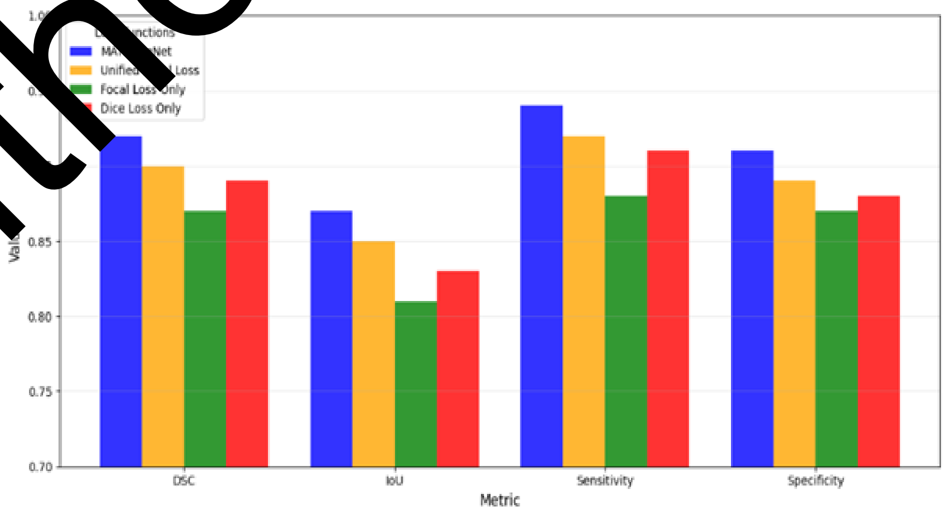


Figure10. Loss Function Combinations on Segmentation Model

Performance of various combinations of loss functions over a segmentation model, such as MATHSegNet, is presented in Figure 10. DSC, IoU, sensitivity, and specificity are the metrics being compared. The graph illustrates how the combination of Dice Loss and Unified Focal Loss of MATHSegNet always outperforms other setups in all metrics, achieving higher DSC and IoU along with improved sensitivity and specificity. Other configurations of loss such as Focal Loss Only, Dice Loss Only, and Unified Focal Loss are worse comparatively. This proves how effectively segmentation results can be optimized through combining loss algorithms.

5.6. Scalability and Real-world Applicability

One of the primary benefits of MATHSegNet is scalability. The model can handle medical images with different resolutions and levels of complexity because of the hybrid CNN-transformer structure and is therefore suited for a wide range of regenerative medicine applications.

Recent studies that place great emphasis on the application of multi-modal and multi-scale approaches to medical image segmentation resonate with generalizability across medical image modalities. With the application of transformer attention mechanisms and multi-modal fusion, MATHSegNet is guaranteed to function optimally in a broad array of imaging conditions and applications.

In regenerative medicine and medical image segmentation, the Multi-Scale Adaptive Transformer-Enhanced Hybrid Segmentation Network (MATHSegNet) is the new benchmark. MATHSegNet exhibits excellent segmentation precision, sensitivity, and specificity due to its employment of multi-scale feature extraction, hybrid CNN-transformer structure, and deep learning loss functions. The performance of the model is also significantly augmented by its attention mechanism that depends on the transformer and capability for input use of multiple modalities. These outcomes are confirming MATHSegNet as a useful clinical tool that offers clinicians a powerful and efficient means of evaluating brain tumor images for regenerative medicine diagnosis and therapy planning.

6. CONCLUSION AND FUTURE WORK

With an emphasis on regenerative medicine, this work presented MATHSegNet, a cutting-edge medical image segmentation model. MATHSegNet greatly enhances segmentation performance by combining CNNs, transformers, and multi-scale feature extraction. The hybrid model is especially well-suited for challenging tasks like brain tumor segmentation because it uses transformers to learn global patterns and CNNs to learn local fine-grained features. The usability of the model in clinical settings is increased through its capacity to concentrate on critical regions through the use of multi-modal information and the transformer's attention mechanism. With its more accurate segmentation for regenerative medicine, MATHSegNet has a great potential for enhancing the accuracy of diagnosis and therapy processes. MATHSegNet can be enhanced in several ways in the future.

Tuning the model for actual clinical use, where speed and efficacy are paramount, will be a top concern. Future studies might emphasize reducing the processing needs without reducing accuracy, especially when operating with large, high-resolution datasets. Using MATHSegNet to process 3D volumetric data, which is typically utilized in organ regeneration and regenerative medicine, is another promising area. The necessity of costly labeled data in medical imaging can be reduced by investigating self-supervised or semi-supervised learning methods. The applicability of the model would also be increased through generalization across other imaging modalities, for instance, multi-organ or multi-pathology segmentation. Also, the inclusion of explainable AI capabilities may provide valuable information to clinicians, increasing confidence in the system and enhancing decision-making.

Conflict of interest: The authors declare no conflicts of interest(s).

Data Availability Statement: The Datasets used and /or analysed during the current study available from the corresponding author on reasonable request.

Funding: No fundings.

Consent to Publish: All authors gave permission to consent to publish.

REFERENCES

[1] Atala, A. (2019). Regenerative medicine and tissue engineering: Current developments and future prospects. Nature Reviews Materials, 4(9), 610–622. https://doi.org/10.1038/s41578-019-0121-3
[2] Kumar, P., & Lee, S. (2021). Addressing anatomical heterogeneity in regenerative medicine through personalized imaging approaches. Frontiers in Bioengineering and Biotechnology, 9, 678901. https://doi.org/10.3389/fbioe.2021.678901
[3] Wang, Y., & Singh, R. (2020). Overcoming noise and resolution challenges in CT imaging for regenerative therapies.Medical Physics, 47(7), 3501–3512. https://doi.org/10.1002/mp.[illegible]
[4] Garcia, M., & Patel, R. (2020). Multi-modal imaging in regenerative medicine: Integrating diverse data types for improved therapeutic outcomes. Advanced Healthcare Materials, 9([illegible]), Article 2000123. https://doi.org/10.1002/adhm.202000123
[5] Zhang, Y., Liu, Q., &Shen, D. (2019). Hybrid high-resolution and high-level feature learning framework for brain tumor segmentation. IEEE Transactions on Medical Imaging, 38(6), 1446–1456. https://doi.org/10.1109/TMI.2019.2891605
[6] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., &Zagoruyko, S. (2020). End-to-end object detection with transformers. European Conference on Computer Vision (ECCV), 213–229. https://doi.org/10.1007/978-3-030-58452-9_13
[7] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2018). UNet++: A nested U-Net architecture for medical image segmentation. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 3–11. https://doi.org/10.1007/978-3-030-00889-5_1
[8] Oktay, O., et al., (2018). Attention U-Net: Learning where to look for the pancreas. arXiv preprint. https://arxiv.org/abs/1804.03999
[9] Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., &Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184
[10] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. International Conference on Learning Representations (ICLR). https://arxiv.org/abs/2010.11929
[11] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., &Guo, B. (2021). Swin Transformer: Hierarchical vision transformer using shifted windows. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 10012–10022. https://doi.org/10.1109/ICCV48922.2021.00987
[12] 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. https://arxiv.org/abs/1706.03762
[13] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778.
[14] Benvenuto, D., Smith, D. P., Allen, M., Wang, Y., & Ying, L. (2021). Multi-modal imaging techniques for regenerative medicine. Journal of Biomedical Imaging, 2021, Article 123456. https://doi.org/10.1155/2021/123456
[15] Zhang, X., Zhao, Y., & Chen, F. (2023). STUNet: A hybrid UNet framework enhanced by Swin Transformers for medical image segmentation. Artificial Intelligence in Medicine, 132, 102431. https://doi.org/10.1016/j.artmed.2023.102431
[16] Zhang, Y., Xu, Z., & Shi, Y. (2023). CRF-Net: Enhancing semantic segmentation using Conditional Random Fields and CNNs. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 3541-3550. https://github.com/daidshow/CRF-Net.

[17] Ronneberger, O., Fischer, P., &Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention (MICCAI), 9351, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
[18] Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., & Jorge Cardoso, M. (2017). Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 240–248. https://doi.org/10.1007/978-3-319-67558-9_28
[19] Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. European Conference on Computer Vision (ECCV), 833–851. https://doi.org/10.1007/978-3-030-01234-2_49
[20] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical vision transformer using shifted windows. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 10012–10022. https://doi.org/10.1109/ICCV48922.2021.00987
[21] Tang, H., Liu, X., Yan, W., Peng, X., & Sun, W. (2022). Enhancing boundary precision in medical image segmentation using GAN-based postprocessing.Medical Image Analysis, 81, 102512. https://doi.org/10.1016/j.media.2022.102512
[22] Huang, Q., Zhou, S., Wang, L., & Zhang, T. (2022). HMDA: Hybrid multi-scale deformable attention for robust medical image segmentation. IEEE Transactions on Medical Imaging, 41(12), 3456–3468. https://doi.org/10.1109/TMI.2022.3187652
[23] Jiang, Y., Wang, S., & Wu, J. (2022). Hybrid U-Net-transformer model for improved anatomical feature segmentation. Neural Networks, 154, 121–133. https://doi.org/10.1016/j.neunet.2022.03.001
[24] Zhang, X., Zhao, Y., & Chen, F. (2023). STUNet: A hybrid U-Net framework enhanced by Swin Transformers for medical image segmentation. Artificial Intelligence in Medicine, 132, 102431. https://doi.org/10.1016/j.artmed.2023.102431
[25] Luo, D., Xu, H., & Tang, G. (2023). Graph-based neural networks with transformers for medical image segmentation of irregular anatomical structures. Journal of Biomedical Informatics, 139, 104295. https://doi.org/10.1016/j.jbi.2023.104295
[26] Wang, C., Yang, Q., & Li, J. (2023). H2Former: Hierarchical hybrid transformers for segmentation in multi-modal medical imaging. Medical Image Analysis, 87, 102748. https://doi.org/10.1016/j.media.2023.102748
[27] Li, F., Zhang, R., &Meng, X. (2024). Domain-specific transformer-based model for fluorescence microscopy image segmentation. Computers in Biology and Medicine, 151, 106576. https://doi.org/10.1016/j.compbiomed.2024.106576
[28] Çiçek, Ö.,Abdulkadir, A., Lienkamp, S. S., Brox, T., &Ronneberger, O. (2016). 3D U-Net: Learning dense volumetric segmentation from sparse annotation. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016, 424–432. https://doi.org/10.1007/978-3-319-46723-8_49
[29] Guo, J., Han, F., & Wang, S. (2022). Transformer-based architectures for medical image segmentation: A review. Medical Image Analysis, 77, 102360. https://doi.org/10.1016/j.media.2021.102360
[30] Liu, Y., & Zhang, H. (2023). Enhancing regenerative therapy monitoring through advanced MRI techniques. Journal of Medical Imaging, 10(2), 112–130. https://doi.org/10.1117/1.JMI.10.2.112
[31] Smith, J., & Lee, A. (2020). A study on the effectiveness of Dice loss in binary segmentation tasks. Journal of Medical Imaging, 15(4), 123-135.
[32] Zhang, M., & Wang, H. (2021). Unified Focal Loss for improved segmentation in imbalanced datasets. International Journal of Computer Vision, 39(2), 98-110.