

# Improved Mask RCNN Network for Brain Tumor MRI Image Instance Segmentation Analysis

# Ming Him Foun\*

School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland

# ABSTRACT

In the pursuit of advancing the precision of brain tumor MRI image detection and segmentation for the purpose of fulfilling the requirements of automated medical analysis, this research introduces an enhanced Mask R-CNN method specifically tailored for high-precision brain tumor instance segmentation. The augmentation involves the incorporation of the Convolutional Block Attention Module (CBAM) hybrid attention mechanism, aimed at improving the model's feature extraction capabilities and adaptively reinforcing its responsiveness to critical features. This enhancement facilitates a more precise capture of key tumor information. Furthermore, the integration of the Bi-directional Feature Pyramid Network (BiFPN) feature fusion technology ensures the model's ability to accurately segment brain tumors of diverse sizes and shapes, thereby enhancing its capacity to identify and segment multi-scale targets. Through a series of rigorous experimental validations, the proposed model demonstrates notable improvements. The precision of the model attains 90.79%, marking a 0.67% enhancement compared to the original model. Similarly, the recall achieves 91.44%, indicating a 0.79% improvement, while the mean Average Precision (mAP) reaches 95.12%, reflecting a substantial increase of 1.88%. Beyond achieving accurate segmentation of brain tumor MRI images, the proposed method excels in precisely calculating the tumor's area and diameter. Consequently, these findings furnish valuable reference data for medical research and diagnosis, underscoring the potential clinical significance of the developed methodology.

Keywords: Magnetic resonance imaging; Brain tumor; Fusion operations; Segmentation; Feature pyramid network

# INTRODUCTION

As a common nervous system disease, the clinical diagnosis and treatment of brain tumor have received extensive attention [1]. In clinical practice, doctors often need to use medical imaging to assess the location, shape, size, and relationship of brain tumors to surrounding tissues [2]. Magnetic Resonance Imaging (MRI), as a non-invasive, high-resolution medical imaging technology, is widely used in the diagnosis and monitoring of brain tumors because of its good contrast to soft tissue. The advantages of MRI are its non-invasiveness, absence of ionizing radiation and high sensitivity to soft tissue, especially for imaging brain tumors and surrounding tissues [3]. However, due to the diverse size, shape, and structure of brain tumors, as well as the blurring of boundaries with neighboring tissues, it has become extremely challenging to accurately and automatically segment brain tumors from MRI images.

However, manually analyzing brain tumor areas in MRI images is a time-consuming and subjective task that is susceptible to individual physician differences and fatigue levels, as well as possible inaccuracies [4]. In order to solve this problem, Computer Aided Diagnosis (CAD) technology came into being [5]. Among them, brain tumor detection and segmentation methods based on machine learning and image processing have developed rapidly in recent years and become a research hotspot. Anand, et al., adopted a multimodal strategy integrating machine learning with medical assistance [6]. They used geometric mean filters to remove image noise and ued the fuzzy c-means algorithm to partition the images into smaller blocks for more precise localization of brain tumors. On the other hand, Zhan, et al., proposed an approach that combines semi-supervised learning principles, image characteristics, and clinical prior knowledge [7]. Particularly effective when labeled data is scarce, they enhanced segmentation through co-training with multiple classifiers. Additionally, Thayumanavan, et al., designed a specialized median filtering technique optimized for skull separation in MRI [8]. This method can identify abnormal brain tissue even in low-contrast conditions and accurately locate the boundaries of diseased tissue,

Correspondence to: Ming Him Foun, School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland, E-mail: founm@tcd.ie

Received: 29-Jun-2024, Manuscript No. JMDM-24-32540; Editor assigned: 02-Jul-2024, PreQC No. JMDM-24-32540 (PQ); Reviewed: 18-Jul-2024, QC No. JMDM-24-32540; Revised: 26-Jul-2024, Manuscript No. JMDM-24-32540 (R); Published: 02-Aug-2024, DOI: 10.35248/2168-9784.24.13.482

Citation: Foun MH (2024) Improved Mask RCNN Network for Brain Tumor MRI Image Instance Segmentation Analysis. J Med Diagn Meth. 13:482.

**Copyright:** © 2024 Foun MH. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

#### Foun MH

enabling further analysis of texture and morphological features of brain tumors. Alam, et al., developed a Template-Based K-Means and Improved Fuzzy C-Means (TKFCM) algorithm model for human brain tumor detection in MRI images [9].

In recent years, deep learning techniques have demonstrated immense potential in MRI brain tumor segmentation. Specifically, the application of Convolutional Neural Networks (CNNs) has made the identification and segmentation of brain tumors from complex brain MRI images more accurate and efficient [10,11]. These networks can automatically learn tumor features from large training datasets and apply them to new, unknown images, achieving high segmentation accuracy. Researchers have proposed various network architectures and training strategies tailored to different types of brain tumors. For instance, Deng, et al., introduced a novel brain tumor segmentation method that integrates unquantifiable local features into the Fully Convolutional Neural Network (FCNN) for fine-grained boundary segmentation [12]. Compared to traditional MRI brain tumor segmentation methods, this approach significantly improves segmentation accuracy and stability. Zheng, et al., improved the segmentation performance of the U-Net network by introducing Hybrid Dilated Convolution (HDC) modules and enhanced connectivity between modules in two sequential networks, achieving higher-precision semantic segmentation of brain tumors [13]. Ranjbarzadeh, et al., designed a flexible and efficient brain tumor segmentation system that significantly reduces computation time through local image preprocessing [14]. They also improved tumor segmentation accuracy by using a Cascade Convolutional Neural Network (C-CNN) and the Distance-Wise Attention (DWA) mechanism. Zhao, et al., integrated Fully Convolutional Neural Networks (FC-NNs) and Conditional Random Fields (CRFs) into a unified framework, developing a novel brain tumor segmentation method that achieves brain tumor segmentation results with both appearance and spatial consistency [15].

Despite significant technological advancements in the aforementioned methods, brain tumor segmentation tasks still face numerous challenges. Firstly, there is a noticeable variation in the quality of MRI images. Differences in equipment or variations in patient positioning can result in substantial changes in image contrast, resolution, and noise levels. This necessitates algorithms with strong adaptability and robustness to handle various image quality variations. Secondly, the heterogeneity of brain tumors presents segmentation challenges. Brain tumors can be benign or malignant, primary or recurrent, and exhibit significant variations in morphology, size, and growth patterns. Furthermore, tumors may be located deep within critical brain functional areas, making precise boundary detection an important task. Addressing these challenges, this paper proposes an efficient instance segmentation method tailored for brain tumor MRI images. On one hand, it optimizes brain tumor MRI images to ensure their efficiency and accuracy in detection and segmentation. Moreover, this method goes beyond meeting basic segmentation requirements; it also provides substantial diagnostic tools for medical professionals. This includes in-depth analysis of pathological features on MRI images, precise estimation of tumor size, and calculation of total tumor area, thereby offering comprehensive data support for medical diagnostic decisionmaking.

# MATERIALS AND METHODS

## Data sets collection and processing

In this study, the publicly available brain MRI image dataset from the Kaggle international competition platform was selected to provide a data foundation for in-depth research on the detection and segmentation of brain MRI images [16]. This dataset comprises 1,280 brain MRI images categorized into two major groups: Healthy and brain tumor cases. It is worth noting that these images are not only sourced from medical imaging centers worldwide, ensuring geographic diversity, but also encompass various types of brain tumors, including benign, malignant, primary, and secondary tumors. To facilitate the analysis and application of images by researchers, all MRI images are stored in the widely used JPG format. Additionally, for more accurate and specific image interpretation, each image is accompanied by a JSON label file that provides detailed annotations of key information within the image, such as tumor location, size, shape, and other relevant details (Figure 1).

#### Image augmentation

Data augmentation is a method of artificially increasing the training data volume by applying various image transformation techniques, with the aim of enhancing the model's generalization ability and reducing the risk of overfitting [17]. In brain MRI tumor images, obtaining high-quality tumor image data can be challenging, making data augmentation techniques particularly important. In this study, four primary data augmentation strategies were selected to enhance the robustness of model training, including rotation, mirroring, brightness, and noise as shown in Figure 2 [18,19]. Specifically, by applying rotations at different angles to the original images, tumor imaging features from various perspectives were stimulated, thus increasing data diversity. Horizontal mirroring of images generated slightly different new images from the original ones, aiding the model in capturing symmetric features more effectively. Altering the brightness of images can simulate different imaging conditions, such as lighting and exposure times. This approach allows the model to better adapt to variations in image brightness that occur in the real world due to different scanning devices or conditions. Adding random noise to images can simulate random perturbations during the image capture process, such as electronic noise from equipment or other factors, thereby improving the model's robustness to noise (Figure 2).

# Mask RCNN brain tumor MRI image instance segmentation model

Mask Region-based Convolutional Neural Networks (R-CNN) is a deep learning architecture proposed by the Facebook AI Research team for object detection and instance segmentation tasks [20]. It is an extension of Faster R-CNN, with an additional branch for predicting segmentation masks, enabling it to simultaneously output object bounding boxes and segmentation masks [21]. The overall structure includes a backbone network, a Region Proposal Network (RPN), and two prediction branches. The backbone e network typically utilizes deep convolutional neural networks such as ResNet or VGG to extract features from input brain tumor MRI images. To achieve multi-scale feature extraction and enhance feature representativeness, the backbone network is often combined with a Feature Pyramid Network (FPN). FPN utilizes feature maps from shallow to deep layers to ensure good performance in handling tumors of various scales. The Region Proposal Network (RPN) operates on the extracted feature maps to predict and generate potential tumor bounding box proposals. These proposals are based on global image features and incorporate local details to ensure accuracy. Once these proposals are obtained, Mask R-CNN uses RoIAlign technology to precisely extract fixed-size features from the feature maps of the backbone network. These features are then fed into two prediction branches.

One branch predicts the position and class of the tumor bounding box, while the other branch generates the segmentation mask of the tumor (Figure 3).





Figure 2: Brain tumor MRI Image Data Augmentation. (A): Rotate; (B): Mirror; (C): Light; (D): Noise.



# Improved brain tumor MRI image instance segmentation model

In the task of brain tumor MRI image instance segmentation, the traditional Mask R-CNN model has certain limitations. Firstly, brain tumors in MRI images can exhibit various shapes and sizes, and the surrounding brain tissue structures are complex, which increases the difficulty of segmentation. These detailed features may not receive sufficient attention in the traditional ResNet backbone network. To address this issue, an attention mechanism was incorporated into the backbone network. The attention mechanism allows the model to focus more on the detailed features of the brain tumor, enhancing the model's ability to extract features from these abnormal regions, thereby significantly reducing interference from background noise and other unrelated brain tissues [22]. Secondly, the original Mask RCNN utilizes FPN (Feature Pyramid Network), which may not capture all important contextual information in certain complex scenarios. Brain tumor instance segmentation requires not only the recognition of large, prominent tumor masses but also consideration of the edges and finer tumor details. To effectively address this challenge, the Bidirectional Feature Pyramid Network (BiFPN) was adopted as an alternative to the traditional FPN [23]. BiFPN, through bidirectional feature fusion, effectively captures brain tumor information at different scales, ensuring that the model can comprehensively and accurately capture and segment brain tumors.

CBAM attention mechanism: The Convolutional Block Attention Module (CBAM) is an attention mechanism introduced into convolutional neural networks to enhance the model's sensitivity to input features [24,25]. This attention mechanism does not simply amplify certain features; instead, it allows the model to automatically determine which features are more important through specific strategies and computations, assigning them higher weights. This ensures that when the model processes brain tumor MRI images, it can focus more on the decisive and critical features, thereby improving the overall model performance. CBAM consists of two main components: Channel Attention Module (CAM) and Spatial Attention Module (SAM). In the task of brain tumor MRI image instance segmentation, determining the subtle pathological conditions of brain tumors requires the model to deeply consider the local information in convolutional features. However, accurately locating brain tumors relies more on spatial information. By enhancing the model's

attention to both local and spatial information, reducing the weight assigned to irrelevant features, CBAM ultimately improves the accuracy of brain tumor MRI image instance segmentation. Its overall structure is illustrated in Figure 4.

In the research of brain tumor MRI image instance segmentation, the selection of input feature channels is important. Effective channel selection helps the model focus on features closely related to the tumor while excluding redundant or irrelevant information [26]. Channel attention primarily determines which channels are more important by capturing different feature descriptions through global average pooling and global max pooling, and assigns a weight to each channel based on this information. The specific operation is illustrated in Figure 5. For the input feature map (with height (H), width (W), and channels (C)), it first obtains two 1 × 1 × C channel mappings using global average pooling and global max pooling. Then, it computes the weights for each channel through a shared Multi-Layer Perceptron (MLP) consisting of two layers. Finally, an elementwise sum is performed, and the result goes through a Sigmoid activation function to generate the new channel feature Mc. In this way, the model adaptively weights each channel, emphasizing features most relevant to brain tumor detection and recognition while ignoring channels with low discriminative power or weak association with the target (Figure 5).

The calculation formula is represented as Equation 1, where  $\sigma$  denotes the Sigmoid activation function, MLP stands for multi-layer perceptron, Avgpool represents the average pooling operation, and Maxpool represents the maximum pooling operation.

# $M_{C}(F) = \sigma(MLP(AvgPool(F))) + MLP(MaxPool(F)).....(1)$

Following the application of channel attention processing, a subsequent utilization of spatial attention is implemented to further accentuate pivotal spatial positions within the input feature map. Spatial attention serves the primary purpose of determining the regions of heightened importance within the input feature map, typically accomplished through the employment of a small convolutional network for calculating weights associated with each spatial position. The specific operation is illustrated in Figure 6. Initially, for the feature map derived from channel attention (possessing dimensions of height H', width W', and C' channels), two H' × W' × 1 channel mappings are acquired through global average pooling and global max pooling, with the resulting

mappings concatenated. Subsequently, a  $7 \times 7$  convolutional operation followed by a Sigmoid activation function produces the new spatial feature denoted as Ms. This method enhances the model's ability to pinpoint regions pertinent to brain tumors with increased precision, thereby markedly mitigating interference from irrelevant areas and augmenting the overall accuracy of the model's detection capabilities (Figure 6).

The calculation formula is represented as Equation 2, where denotes the convolution operation, and  $7 \times 7$  is the size of the convolution kernel.

$$M_{s}(F) = \sigma(f^{\gamma \times \gamma}([AvgPool(F); MaxPool(F)])) \dots (2)$$

Integration of attention modules with the backbone network: In this study, CBAM was incorporated into the backbone network of Mask R-CNN to enhance its instance segmentation capabilities for brain tumor MRI images. The specific implementation method is illustrated in Figure 7. First, CBAM's channel attention helps the model identify which channel features are important for detecting brain tumor regions, such as edges, textures, and shapes, thereby enhancing the influence of these channels. Secondly, spatial attention allows the model to focus more on specific locations of the brain tumor, such as the tumor core, tumor boundaries, or other abnormal structural regions, improving segmentation accuracy. Additionally, due to CBAM's adaptability, it provides consistent performance improvements to the model across different levels and contrasts of MRI images.

Overall, integrating CBAM into the Mask R-CNN backbone network not only enhances segmentation accuracy for brain tumors but also improves the model's robustness when dealing with various brain structures and abnormal conditions (Figure 7).





Figure 5: Channel attention implementation method.





Bidirectional feature fusion network: In the field of computer vision, feature fusion is a critical factor in determining model performance. The original Mask R-CNN achieves feature multiscale fusion through FPN (Feature Pyramid Network). FPN considers the combination of deep and shallow feature maps, allowing the model to effectively detect and segment objects in images. However, FPN has some limitations. While shallow feature maps contain rich detail information and deep feature maps provide high-level abstractions of the image, FPN fails to fully integrate the advantages of both. To overcome these limitations and better fuse features from different layers, this study introduces the BiFPN [27]. Unlike FPN, BiFPN, while maintaining top-down and bottom-up connections, introduces cross-layer connections between feature maps. Figure 8 illustrates the sub-module structures of FPN and BiFPN. This structure enhances feature integration capability and improves computational efficiency with fewer parameters. More importantly, the design of BiFPN allows for multiple bidirectional fusions, ensuring comprehensive feature integration across different scales [28]. Each bidirectional path is treated as a feature network layer and is stacked repeatedly to achieve deeper scale fusion (Figure 8).

For the instance segmentation of brain tumor MRI images, preserving both image details and high-level information is important. Brain tumor structures may exist at different scales in MRI images, and the bidirectional fusion mechanism of BiFPN can help capture the detailed structures and locations of these tumors. This means that BiFPN effectively combines high-level semantic information with shallow detail information, providing more accurate features for brain tumor instance segmentation in MRI images. The proposed Bidirectional Feature Pyramid Network (BiFPN) introduces an optimization strategy with crossscale connections to further enhance feature fusion. Specifically, to improve computational efficiency and feature quality, three key improvements were made to the traditional FPN structure. First, during the upsampling and fusion of shallow features, nodes that only served as single inputs and did not contribute to feature fusion were removed. This approach retained only the intermediate feature layers that genuinely contributed to feature fusion, thereby reducing the consumption of computational resources. Additionally, during the downsampling process, when fusing shallow features with deep features, an extra connection to the output nodes at the same level but containing deep information was added. To ensure comprehensive feature exploration across scales, a single bidirectional fusion was not deemed sufficient. Instead, the bidirectional fusion paths were utilized as a feature extraction layer and stacked three times to achieve deeper scale fusion. Specifically, given five feature layers C1, C2, C3, C4, and C5, after three repetitions of bidirectional fusion operations, a three-layer deep structure P1, P2, P3, P4, and P5 was obtained, achieving high-level scale fusion of features. The specific BiFPN structure is illustrated in Figure 9. This strategy captures all features from fine to macroscopic in brain tumor MRI image instance segmentation, significantly enhancing segmentation accuracy and stability (Figure 9).

The loss function for brain tumor MRI image instance segmentation: To achieve optimal performance in brain tumor MRI image instance segmentation, a thorough investigation and adjustment of the loss function is crucial. In this study, by combining classification loss  $L_{cds}$ , regression loss  $L_{reg}$  and segmentation loss  $L_{mask}$ , this study ensure that the model excels in both tumor recognition, localization, and precise segmentation. The specific implementation is as follows in the following formulas.

$$L_{cls} = -\sum_{i} \left[ p_{i} \log\left(\hat{p}_{i}\right) + (1 - p_{i}) \log\left(1 - \hat{p}_{i}\right) \right] \dots (3)$$

$$L_{reg} = \sum_{i} smooth_{L1} \left(t_{i} - \hat{t}_{i}\right) \dots (4)$$

$$L_{mask} = -\sum_{j} \left[ m_{j} \log\left(\hat{m}_{j}\right) + (1 - m_{j}) \log\left(1 - \hat{m}_{j}\right) \right] \dots (5)$$

$$L = \lambda_{cls} L_{cls} + \lambda_{reg} L_{reg} + \lambda_{mask} L_{mask} \qquad (6)$$

Formula 3 represents the classification loss, which aims to ensure that the model can accurately distinguish different categories, such as tumor and non-tumor regions. In this formula,  $p_i$  represents the true class label, and  $p_i$  is the predicted probability by the model. Formula 4 is the bounding box regression loss, which measures the difference between the predicted bounding boxes and the true bounding boxes to accurately locate the tumor in the MRI image.  $t_i$  represents the coordinates of the true bounding box, and  $t_i$  represents the coordinates of the model's predicted box. Formula 5 is the instance segmentation loss, which ensures that the predicted segmentation mask closely matches the actual tumor shape.  $m_i$  represents the true labels for each pixel, and  $m_i$  represents the predicted pixel probabilities by the model. Formula 6 is the overall loss function of the model, which is used to determine the performance of tumor recognition, localization, and segmentation.  $\lambda_{cls}$ ,  $\lambda_{reg}$  and  $\lambda_{mask}$  is weight coefficients that are

adjusted based on different tasks.

#### Experimental environment and hyperparameter settings

The instance segmentation networks were all constructed, trained, and tested using the PyTorch deep learning framework, ensuring efficient and stable overall performance. The experimental environment was set up on a Windows 10 operating system, equipped with an Intel (R) Core (TM) i9-11900KF central processing unit and 2 NVIDIA GeForce RTX 2080Ti graphics processing units. All coding work was done in a Python 3.8 environment, and to accelerate the model training process, CUDA 11.3 and cuDNN 8.4.1 were configured. The hyperparameters for model training are shown in Table 1, with a total of 180 epochs trained to achieve maximum convergence of the models (Table 1).





Parameter	Value
Optimizer	SGD
Learning rate	0.004
Learning rate decay	(25, 60, 100, 130)
Batch size	32
Label smoothing	0.005
RPN anchor scales	(16, 32, 64, 128, 256)

Table 1: Training parameters and values.

#### **Evaluation metrics**

To ensure comprehensive validation of the models proposed in this study, several commonly used performance evaluation metrics were chosen, including Precision, Recall, F1 Score, and mAP (mean Average Precision). These metrics provide a comprehensive and accurate assessment of the model's performance from various perspectives. The calculation formulas for these metrics are as follows:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$mAP_{(seg)} = \frac{\sum_{i=1}^{N} AP_i}{N}$$

$$(10)$$

In the formulas, *TP*, *FP* and *FN* represent the number of correctly detected tumors, the number of incorrectly detected tumors, and the number of tumors that were not correctly detected, respectively. Precision indicates the probability of true tumors among the detected tumor regions. Recall represents how many tumor targets in the image were detected. F1 Score is the harmonic mean of Precision and Recall, providing a single performance evaluation value that combines information from both metrics. A higher F1 Score indicates better model performance, with a maximum value of 1. mAP provides the average precision across various categories and serves as an intuitive indicator of the overall model performance, considering multiple classes simultaneously.

# **RESULTS AND DISCUSSION**

#### Model accuracy

Model accuracy is an important metric for evaluating its performance, typically analyzed by observing the loss values during the training process. To optimize the model and expedite its convergence, a transfer learning strategy was employed using pre-trained weights from the COCO dataset as initial weights and freezing the model's fully connected layers for the first 15 epochs. The model training process is illustrated in Figure 10. Over the course of the 180 training epochs, the model's loss values were relatively high during the initial 10 epochs. However, as training progressed, a gradual decrease in loss values and a corresponding improvement in the model's mAP (mean Average Precision) values were observed. It's worth noting that during the first 50 epochs, the model's mAP exhibited some oscillations. This could be attributed to the initially high learning

rate. Nevertheless, as the learning rate was reduced, the model's mAP oscillations noticeably stabilized, indicating better robustness and generalization performance. Additionally, the transfer learning strategy validated the effectiveness of freezing certain layers during the early training stages. Freezing layers ensured that the pre-trained weights were not significantly altered in the initial training, thereby mitigating issues such as overfitting or convergence difficulties during the early stages (Figure 10).

# Analysis of brain tumor MRI image instance segmentation results

The improved Mask RCNN demonstrates outstanding performance in brain tumor MRI image instance segmentation. After optimization and training, the model not only accurately identifies and locates tumors but also provides in-depth analysis of tumor morphology and size. In practical use, the model first generates detection boxes for tumor regions. These detection boxes provide precise information about the tumor's location in the MRI image. Unlike traditional methods, the improved model offers higher localization accuracy, enabling medical professionals to quickly and accurately pinpoint the lesion area. For doctors and radiology technicians, this serves as an intuitive tool for rapid identification of the tumor's location. Moreover, the size of the detection box allows the calculation of the diameter of the tumor lesion area, providing essential reference information for subsequent treatment planning. However, the detection box can only provide approximate information. To obtain more detailed information about the tumor's structure and morphology, the model further generates a mask. This mask is a binary image in which the tumor area is labeled as 1, while other areas are labeled as 0. This means that through the mask, it is possible to not only obtain the accurate area of the tumor but also analyze its morphology more clearly, such as the presence of lobulation, the uniformity of the tumor, and more. This information is crucial for determining the malignancy level of the tumor, its growth pattern, and potential risks. Figure 11 illustrates the results of the improved Mask RCNN brain tumor MRI image instance segmentation analysis.

#### Ablation experiments

To comprehensively analyze the impact of various strategies on the segmentation performance of brain tumor MRI images, ablation experiments were devised for each strategy. All experiments were conducted with consistent hyperparameters and operating environments to ensure the validity and comparability of the results. The outcomes are presented in Table 2. It is evident from the table that employing either the BiFPN or CBAM attention mechanism independently significantly enhances the performance of Mask RCNN. BiFPN excels in capturing intricate image details, while CBAM automatically concentrates on crucial areas such as tumor edges and structural intricacies. Notably, the CBAM attention mechanism exhibits a slight advantage across all metrics. This is attributed to its ability to effectively focus on specific tumor areas during the processing of brain tumor MRI images, thereby enhancing model segmentation accuracy. When both BiFPN and CBAM were implemented in the segmentation of brain tumor MRI images, the model demonstrated optimal performance in Precision, Recall, F1 Score, and mAP. This indicates that these two strategies complement each other, collectively enhancing the model's segmentation capabilities when dealing with MRI images depicting complex tumor structures and diverse morphologies. Compared to the original Mask RCNN model, the enhanced model exhibited improvements of 0.67% in Precision, 0.79% in Recall, 1.14% in F1 Score, and 1.88% in mAP. These experimental findings unequivocally demonstrate the effectiveness of our proposed enhancement strategy in elevating the instance segmentation performance of brain tumor MRI images. This offers robust support for subsequent in-depth research and clinical applications (Table 2).







Table 2: The impact of different	t optimization	strategies on	model	performance.
----------------------------------	----------------	---------------	-------	--------------

Algorithms	Precision (%)	Recall (%)	F1 Score (%)	mAP (%)
Mask RCNN	90.12	90.65	91.23	93.24
Mask RCNN+BiFPN	90.37	90.96	91.71	94.13
Mask RCNN+CBAM	90.54	91.21	91.98	94.79
Mask RCNN+BiFPN+CBAM	90.79	91.44	92.37	95.12

#### Comparison of different model performances

To further validate the efficacy of the enhanced Mask RCNN model in instance segmentation of brain tumor MRI images, a comparative analysis was conducted with two prominent instance segmentation models, namely YOLACT and SOLO. To ensure fair evaluation, consistent hyperparameter configurations were maintained across all models, including batch size, learning rate, weight decay, and other factors. Each model underwent exhaustive training until convergence. The performance of these models is outlined in Table 3. The tabulated results reveal that the enhanced Mask RCNN model outperforms both YOLACT and SOLO across all four metrics. Notably, the improved Mask RCNN achieved a remarkable mAP of 95.12%, showcasing a significant superiority over the other two models. In the context of processing MRI brain tumor images, the ability to capture detail and guide attention proves pivotal for enhancing model performance. Despite the commendable performance of YOLACT and SOLO in various scenarios, they still fall behind our improved model in this study. This comparative analysis underscores the substantial superiority of the proposed enhancement strategy in the task of segmenting MRI brain tumor images (Table 3).

#### Factors affecting model performance

**Impact of tumor shape on model performance:** The configuration of brain tumor shapes, indicative of biological structural variations, significantly influences model performance. To elucidate the impact of different tumor shapes on the performance of the enhanced Mask RCNN brain tumor instance segmentation model, tests were conducted on two prototypical tumor morphologies: Circsular and irregular shapes. Figure 12 illustrates the experimental results. In scenarios involving circular tumor structures, characterized by regularity and continuity facilitating image segmentation, the enhanced Mask RCNN model showcases exceptional performance. It accurately identifies circular tumors and maintains high-precision segmentation, demonstrating the improved model's sensitivity and accuracy in handling regular structures. In instances of irregular-shaped tumors, the increased morphological variability

Table 3:	Comparison	of different	model	performances.
----------	------------	--------------	-------	---------------

presents greater challenges to image segmentation. Nevertheless, the enhanced Mask RCNN model exhibits robustness in this complex environment. While minor recognition errors may occur in specific areas, the model effectively performs tumor identification and segmentation tasks overall. Based on these experimental findings, it is concluded that the enhanced Mask RCNN brain tumor instance segmentation model maintains outstanding performance under both circular and irregular tumor morphologies. This robust performance establishes a reliable foundation for practical applications in medical image analysis, ensuring accurate segmentation results when diagnosing tumors of diverse shapes (Figure 12).

Effect of the number of brain tumors on model performance: The number of brain tumors is directly related to the complexity and challenges of medical image analysis. To systematically assess the impact of varying numbers of tumors on the enhanced Mask RCNN instance segmentation model, MRI images containing both single and multiple brain tumors were tested. The experimental results, as shown in Figure 13, demonstrate that the model exhibits excellent segmentation performance for images with a single brain tumor. This is mainly attributed to the relatively simple and clear structural features provided by a single tumor, allowing the model to easily capture and achieve highprecision segmentation. However, when multiple brain tumors are present in the image, especially accompanied by smaller tumors, the situation becomes more complex. In the context of multiple tumors, small tumors are often more challenging to detect due to their small size and variable shapes. Despite the challenges, the improved Mask RCNN model still demonstrates its superior performance. The attention mechanism can provide more focus on small tumor targets, and the ability to extract multi-scale features enables accurate detection and segmentation even in complex backgrounds. Based on our experimental results, it can be concluded that the improved Mask RCNN brain tumor segmentation model performs excellently in both single and multiple tumor scenarios. This further emphasizes the value of the improved Mask RCNN model in medical image analysis, providing more accurate and reliable diagnostic tools for medical

Algorithms	Precision (%)	Recall (%)	F1 Score (%)	mAP (%)
YOLACT	90.33	91.01	90.87	91.92
SOLO	89.7	90.76	90.23	91.37
Improved Mask RCNN	90.79	91.44	92.37	95.12



Figure 12: Effects of brain tumors with different shapes on model performance.



professionals (Figure 13).

# CONCLUSION

In this research, a precise method is proposed for detecting and segmenting brain tumors in MRI images. Recognizing the limitations of conventional Mask RCNN models in handling this task, the CBAM (Convolutional Block Attention Module) hybrid attention mechanism has been integrated to enhance the model's capability to identify and extract crucial features. This integration ensures a more accurate capture and enhancement of tumor features in MRI images, making the model more responsive to such features. Additionally, to optimize feature extraction, the BiFPN (Bidirectional Feature Pyramid Network) feature fusion technique has been incorporated, ensuring accurate segmentation of brain tumors across various scales and shapes. Comparative analysis with the original Mask RCNN model demonstrates significant improvements in multiple performance metrics in our enhanced approach. Specifically, precision has reached 90.79%, indicating a 0.67% increase over the original model. Recall has reached 91.44%, reflecting a 0.79% improvement, and mAP (mean Average Precision) has reached 95.12%, showcasing a 1.88% increase. To establish the superiority of our method, comparisons were made with instance segmentation models like YOLACT and SOLO, revealing that the enhanced Mask RCNN outperforms them in the instance segmentation of brain tumor MRI images. Finally, leveraging the instance segmentation results, a detailed analysis is conducted to extract key medical indicators of brain tumors, including area and diameter, providing crucial reference data for the medical field.

# REFERENCES

- Ghandour F, Squassina A, Karaky R, Diab-Assaf M, Fadda P, Pisanu C. Presenting psychiatric and neurological symptoms and signs of brain tumors before diagnosis: A systematic review. Brain Sci. 2021;11(3):301.
- Arabahmadi M, Farahbakhsh R, Rezazadeh J. Deep learning for smart Healthcare-A survey on brain tumor detection from medical imaging. Sensors. 2022;22(5):1960.
- Pandiselvi T, Maheswaran R. Efficient framework for identifying, locating, detecting and classifying MRI brain tumor in MRI images. J Med Syst. 2019;43(7):189.
- Hashemzehi R, Mahdavi SJ, Kheirabadi M, Kamel SR. Detection of brain tumors from MRI images base on deep learning using hybrid model CNN and NADE. Biocybern Biomed Eng. 2020;40(3):1225-1232.

- Çinar A, Yildirim M. Detection of tumors on brain MRI images using the hybrid convolutional neural network architecture. Med Hypotheses. 2020;139:109684.
- Anand L, Rane KP, Bewoor LA, Bangare JL, Surve J, Raghunath MP, et al. Development of machine learning and medical enabled multimodal for segmentation and classification of brain tumor using MRI images. Comput Intell Neurosci. 2022;2022(1):7797094.
- Zhan T, Shen F, Hong X, Wang X, Chen Y, Lu Z, et al. A glioma segmentation method using cotraining and superpixel-based spatial and clinical constraints. IEEE Access. 2018;6:57113-57122.
- Thayumanavan M, Ramasamy A. An efficient approach for brain tumor detection and segmentation in MR brain images using random forest classifier. Concurr Eng. 2021;29(3):266-274.
- Alam MS, Rahman MM, Hossain MA, Islam MK, Ahmed KM, Ahmed KT, et al. Automatic human brain tumor detection in MRI image using template-based K means and improved fuzzy C means clustering algorithm. Big data cogn. 2019;3(2):27.
- Raza A, Ayub H, Khan JA, Ahmad I, S. Salama A, Daradkeh YI, et al. A hybrid deep learning-based approach for brain tumor classification. Electron. 2022;11(7):1146.
- Emam MM, Samee NA, Jamjoom MM, Houssein EH. Optimized deep learning architecture for brain tumor classification using improved hunger games search algorithm. Comput Biol Med. 2023;160:106966.
- Deng W, Shi Q, Luo K, Yang Y, Ning N. Brain tumor segmentation based on improved convolutional neural network in combination with non-quantifiable local texture feature. J Med Syst. 2019;43:1-9.
- Zheng P, Zhu X, Guo W. Brain tumour segmentation based on an improved U-Net. BMC Med Imaging. 2022;22(1):199.
- 14. Ranjbarzadeh R, Bagherian Kasgari A, Jafarzadeh Ghoushchi S, Anari S, Naseri M, Bendechache M. Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images. Sci Rep. 2021;11(1):1-7.
- Zhao X, Wu Y, Song G, Li Z, Zhang Y, Fan Y. A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. Med Image Anal. 2018;43:98-111.
- Vijila Rani K, Joseph Jawhar S, Palanikumar S. Nanoscale imaging technique for accurate identification of brain tumor contour using NBDS method. J Ambient Intell Humaniz Comput. 2021;12(7):7667-7684.
- Xu M, Yoon S, Fuentes A, Park DS. A comprehensive survey of image augmentation techniques for deep learning. Pattern Recognit. 2023;137:109347.
- 18. Haque MA, Marwaha S, Deb CK, Nigam S, Arora A, Hooda KS,

et al. Deep learning-based approach for identification of diseases of maize crop. Sci Rep. 2022;12(1):6334.

- 19. Zhang C, Bao N, Sun H, Li H, Li J, Qian W, et al. A deep learning image data augmentation method for single tumor segmentation. Front Oncol. 2022;12:782988.
- 20. Wang S, Sun G, Zheng B, Du Y. A crop image segmentation and extraction algorithm based on mask RCNN. Entropy. 2021;23(9):1160.
- Azam H, Tariq H, Shehzad D, Akbar S, Shah H, Khan ZA. Fully automated skull stripping from brain magnetic resonance images using mask RCNN-based deep learning neural networks. Brain Sci. 2023;13(9):1255.
- 22. Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning. Neurocomputing. 2021;452:48-62.
- 23. Kim HI, Kim Y, Kim B, Shin DY, Lee SJ, Choi SI. Hyoid bone tracking in a videofluoroscopic swallowing study using a deep-

learning-based segmentation network. Diagn. 2021;11(7):1147.

- Chen N, Man Y, Sun Y. Abnormal cockpit pilot driving behavior detection using YOLOv4 fused attention mechanism. Electron. 2022;11(16):2538.
- Wang Y, Huang X, Xie G, Zhang N. A high-precision automatic recognition method based on target detection for nanometer scaled precipitates or carbides in different alloys. J Mater Res Technol. 2023;26:7767-7774.
- Li Y, Li H, Fan D, Li Z, Ji S. Improved sea ice image segmentation using U2-net and dataset augmentation. Appl Sci. 2023;13(16):9402.
- Li B, Gao Q. Defect detection for metal shaft surfaces based on an improved YOLOv5 algorithm and transfer learning. Sensors. 2023;23(7):3761.
- 28. Xie Q, Wu M, Bao J, Zheng P, Liu W, Liu X, et al. A deep learningbased detection method for pig body temperature using infrared thermography. Comput Electron Agric. 2023;213:108200.