IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 43, No. 2, April, 2026, pp. 629-639
Journal homepage: http://iieta.org/journals/ts

Lightweight Brain Tumor Detection in MRI Based on YOLO11 with Efficient Multi-Scale Attention and Wise-IoU Optimization

Ziyi Wang[1], Ruobing Lv[1], Xu Zhang[1], Yufei Song[1,2], Xi Meng[1,2], Zhiguo Liu[1,2], Qingyong Jin[1,2*]

[1] College of Future Information Technology, Shijiazhuang University, Shijiazhuang 050035, China
[2] Hebei Key Laboratory of IoT and Blockchain Integration, Shijiazhuang 050035, China

Corresponding Author Email: jinqingyong17@126.com

https://doi.org/10.18280/ts.430206

Received: 22 September 2025
Revised: 30 January 2026
Accepted: 17 February 2026
Available online: 30 April 2026

Keywords:
brain tumor detection, YOLO11, efficient multi-scale attention mechanism, lightweight deployment

ABSTRACT

Aiming at the clinical pain points in brain tumor medical image detection, such as sample category imbalance, significant tumor scale differences, inaccurate localization of low-quality samples and difficult model deployment, this paper proposes an improved brain tumor detection algorithm based on YOLO11. Firstly, a category balance strategy and combined data augmentation technology are adopted to solve the problem of uneven sample distribution, and improve the model's detection ability and robustness for tiny lesions. Secondly, an Efficient Multi-scale Attention (EMA) module is introduced, which accurately captures multi-scale tumor features without dimensionality reduction through feature grouping, parallel sub-networks and cross-spatial learning mechanism, reducing the loss of detailed information. Then, the Wise-IoU loss function is used to optimize gradient allocation through a dynamic non-monotonic focusing mechanism, weaken the interference of low-quality samples on bounding box regression, and improve localization accuracy. Finally, combined with the ShuffleNetV2 lightweight architecture, the number of model parameters and computation are reduced through grouped convolution and depthwise separable convolution technologies. The experimental results show that the improved algorithm achieves a precision of 96.3%, a recall of 95.4% and an mAP@50 of 96.7%. Meanwhile, the number of parameters is reduced by 41% and the computation is reduced by 43%. Lightweight deployment is realized on the premise of ensuring detection accuracy, which is highly adaptable to clinical auxiliary diagnosis and edge devices in primary medical institutions, and provides technical support for early accurate screening of brain tumors.

1. INTRODUCTION

In recent years, with the continuous improvement of medical intelligence, brain tumor, as a neurological disease that seriously threatens human health, its early accurate diagnosis has become a key link in clinical diagnosis and treatment [1]. Medical image detection is the core means of brain tumor diagnosis, but traditional detection methods rely on manual image reading, which not only consumes a lot of medical resources, but also has problems such as poor real-time performance and diagnosis accuracy easily affected by doctors' experience, making it difficult to meet the needs of efficient preliminary screening in primary hospitals and accurate clinical diagnosis and treatment [2, 3]. At the same time, brain tumor image detection also faces multiple technical challenges: significant differences in tumor scales, overlapping image features of different types of tumors, and sample category imbalance in datasets, leading to traditional algorithms prone to missed detection, blurred boundary localization, and confused type discrimination.

In addition, the existing deep learning-based detection models usually have a large number of parameters and high computational costs, making it difficult to adapt to the deployment requirements of mobile devices and primary medical institutions. Therefore, developing an intelligent system integrating data balance processing, multi-scale tumor accurate detection and lightweight deployment is of great practical significance for improving the efficiency of brain tumor diagnosis and optimizing the allocation of medical resources.

In the field of medical image detection, target detection technology has evolved from traditional machine learning to the stage dominated by deep learning. Although traditional image processing methods have low computational overhead, their detection accuracy is insufficient under the interference of complex tumor morphology and image noise; deep learning-based algorithms such as the YOLO series and CNN have significantly improved feature extraction capabilities, but they still have problems such as insufficient capture of multi-scale features and impaired localization performance when facing the multi-scale characteristics and low-quality samples of brain tumors [4, 5]. In terms of model deployment, the existing algorithms generally have the problems of large number of parameters and high computational complexity, making it difficult to meet the operation requirements of edge devices in primary hospitals; at the same time, the traditional loss function has unreasonable gradient allocation for low-quality samples, which is easy to lead to the decline of model

localization performance. Therefore, how to realize accurate detection of multi-scale tumors, robust fitting of low-quality samples under the condition of limited hardware resources, and complete model lightweight optimization to adapt to clinical deployment has become the core problem that this research needs to focus on solving.

To address the above problems, this paper carries out multi-dimensional technological innovation and algorithm optimization around the clinical needs of brain tumor medical image detection, and the main innovations are as follows:

(1) Propose an adaptive data processing scheme, solve the sample imbalance problem through category balance strategy and combined data augmentation technology, expand sample diversity by combining Mosaic and MixUp technologies, and improve the model's detection ability and robustness for tiny lesions [6];

(2) Introduce an Efficient Multi-scale Attention (EMA) module, which accurately captures multi-scale tumor features without dimensionality reduction through feature grouping [7], 1 × 1 and 3 × 3 parallel sub-networks and cross-spatial learning mechanism, establishes the dependency between local details and global morphology, and reduces the loss of tumor detailed information;

(3) Adopt the Wise-IoU loss function, reasonably allocate gradient gain through a dynamic non-monotonic focusing mechanism, weaken the harmful effects of low-quality samples on bounding box regression, and optimize the model's localization performance;

(4) Introduce the ShuffleNetV2 lightweight scheme [8], which greatly reduces the number of model parameters and computation within a controllable accuracy loss range through grouped convolution and depthwise separable convolution technologies, adapting to edge device deployment.

2. OVERALL SYSTEM DESIGN

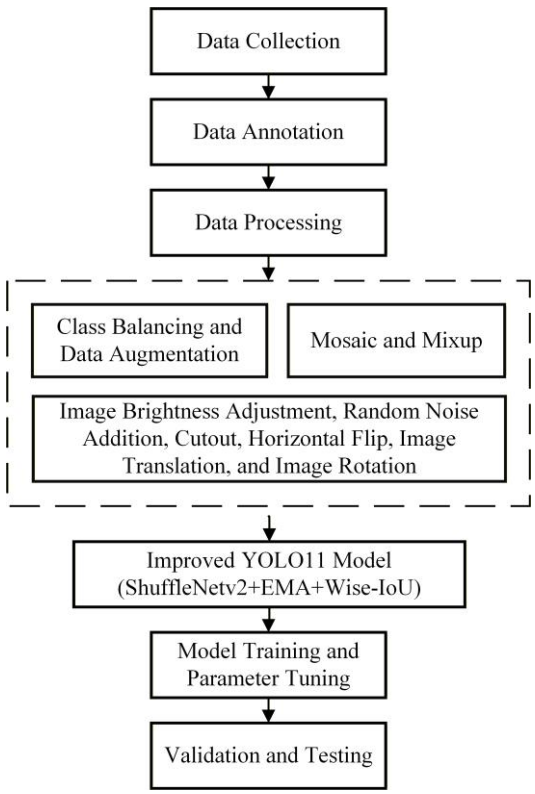

Figure 1. Overall technical roadmap of the project

The technical roadmap of this project starts with the dataset formed by annotating clinical collected brain tumor medical images (MRI) in hospitals (refer to Figure 1). It enriches sample diversity and improves the detection ability of tiny tumor lesions and model robustness through data processing, Mosaic and MixUp data augmentation. Based on YOLO11, a basic detection framework is constructed, and an Efficient EMA module is introduced. Through feature grouping, 1 × 1 and 3 × 3 parallel sub-networks and cross-spatial learning mechanism, it captures multi-scale features while avoiding the loss of tumor detailed information caused by dimensionality reduction, and models the pixel dependency between tumors and surrounding tissues.

Using the hardware environment such as PyTorch framework and RTX4060 GPU under Windows11 system, the model training is completed with hyperparameter configuration such as SGD optimizer and cosine annealing learning rate. After verifying the advantages of the EMA module in improving precision, recall and reducing the number of parameters through comparative experiments, the model is applied to clinical auxiliary diagnosis, curative effect evaluation, diagnosis support in primary hospitals and other scenarios, realizing a complete technical link from medical image data processing to clinical practical application.

3. DESIGN OF BRAIN TUMOR DETECTION ALGORITHM

3.1 YOLO11 algorithm

YOLO11 is a target detection algorithm developed by Ultralytics. Compared with the previous generation YOLO models, YOLO11 has a prominent balance between lightweight and clinical practicality in brain tumor detection, which can meet the frame rate requirements of clinical real-time diagnosis. At the same time, through module optimization, its detection accuracy is significantly better than that of the same level previous generation models when dealing with complex backgrounds such as MRI image artifact interference and irregular tumor morphology, especially the recall rate of tiny lesions is significantly improved.

Nevertheless, YOLO11 still has problems in the small target detection of brain tumors, accuracy improvement and lightweight deployment. Therefore, this study improves YOLO11 to enhance its feature extraction and lightweight deployment capabilities, which can accurately identify tumor regions, boundaries and subtle features from brain tumor medical images [9]. Even in challenging scenarios with poor image quality and limited equipment, it can more accurately capture the complex features of tumors, meeting the needs of clinical auxiliary diagnosis. The network structure of YOLO11 mainly includes three parts: Backbone, Neck and Head, and the overall network structure is shown in Figure 2.

Backbone is a key component of the YOLO architecture, responsible for extracting features from input images of multiple scales. This process involves stacking convolutional layers and dedicated blocks, such as residual blocks and bottleneck blocks. These structures can effectively extract the deep semantic information of images, and generate feature maps of various resolutions through different step sizes and convolution kernel sizes, laying a foundation for subsequent multi-scale target detection and ensuring that the model can capture target features from tiny to large sizes.

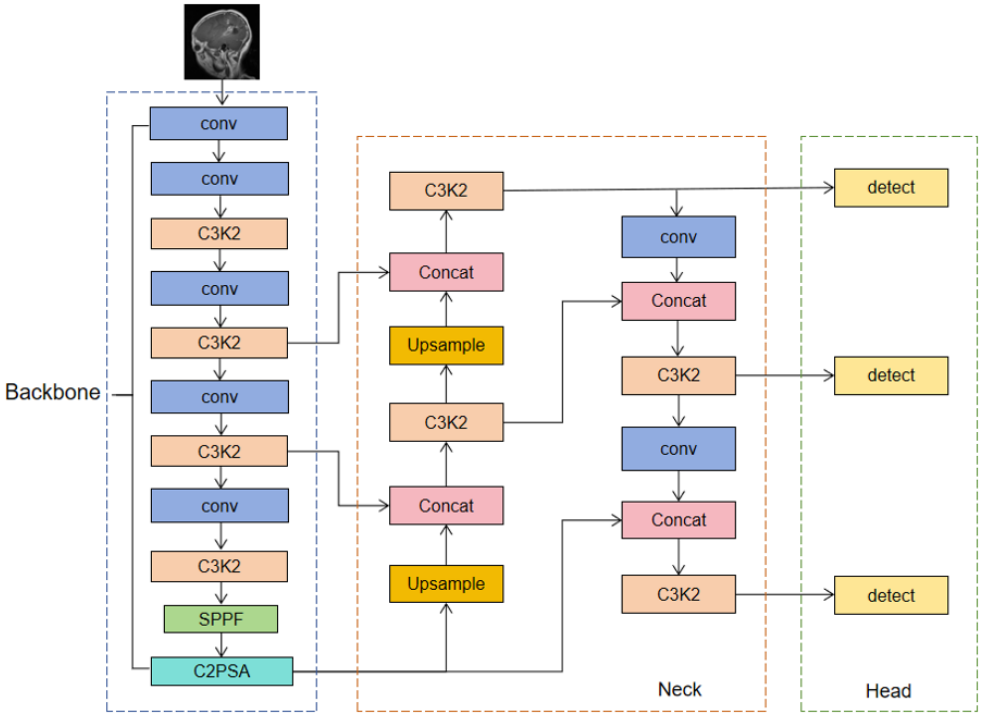

Figure 2. Network framework of YOLO11

Neck fuses features of different scales and transmits them to the Head for prediction. This process usually involves upsampling and concatenation of feature maps from different levels, enabling the model to effectively capture multi-scale information. In this way, Neck can fuse shallow detailed features and deep semantic features, so that the subsequent Head can not only accurately locate the target but also accurately identify the target category when making predictions, improving the overall detection performance.

Head is responsible for generating the final predictions of target detection and classification. It processes the feature maps transmitted from the Neck, and accurately parses the target information contained in the feature maps through well-designed convolution operations and prediction branches. Finally, the Head outputs the bounding box coordinates of objects in the image, determines the category labels of the targets, and also gives the confidence score for the prediction results, providing a quantitative reference for the reliability of the detection results, thus realizing the accurate detection and classification of targets such as brain tumors.

3.2 Efficient Multi-scale Attention module

Accurate detection of brain tumors is the core link of neuroimaging diagnosis and clinical treatment decision-making, and its detection accuracy directly depends on the effective extraction and focusing of tumor region features. As a new generation of one-stage target detection algorithm, YOLO11 shows excellent performance in general target detection scenarios, but still has limitations when facing the complex features of brain tumor medical images: the traditional feature extraction module of YOLO11 is difficult to accurately model the diversity of tumor features, and is prone to feature focusing deviation.

As a core means to enhance the feature representation ability, the attention mechanism can focus on key regions and suppress redundant information through adaptive weight allocation, providing an effective idea for solving the feature modeling problem of brain tumor image detection. Different from the traditional static attention, the EMA module [10] can not only capture the global contextual features of tumor regions, but also finely depict local detailed features such as tumor edges and textures, effectively adapting to the diversity and variability of brain tumor features.

Before the proposal of the EMA attention mechanism, a variety of attention mechanisms have been tried in medical image detection tasks, but all have obvious limitations for brain tumor detection scenarios: the SE attention mechanism [11] only focuses on channel dimension dependency, lacks accurate capture of the spatial position information of lesions in medical images, and is difficult to adapt to the spatial heterogeneous features of lesions such as brain tumors; CBAM [12] combines channel and spatial attention, but has insufficient spatial feature capture ability in brain tumor MRI detection, and the localization accuracy of small lesions or lesions with blurred boundaries is limited. The EMA attention mechanism conducts refined learning of tumor multi-semantic features through feature grouping, the parallel sub-networks take into account channel correlation and multi-scale spatial feature extraction, and cross-spatial learning realizes feature fusion from global to local. It not only makes up for the deficiencies of traditional mechanisms in a single dimension such as spatial localization, fine-grained feature depiction or long-range dependency modeling, but also adapts to the lightweight characteristics of YOLO11, achieving a balance between accuracy and efficiency in brain tumor medical image detection.

3.2.1 Efficient Multi-scale Attention module

The EMA attention mechanism reconstructs the feature extraction and fusion module of the algorithm (Figure 3). In the EMA module, the 1 × 1 convolution of the CA module is selected as a shared component [13], and a 1 × 1 branch is constructed to accurately capture the correlation features between tumor channels; at the same time, the 3 × 3 convolution kernel is set in parallel with the 1 × 1 branch to

form a 3 × 3 branch to quickly aggregate the multi-scale spatial structure information of tumors; combined with feature grouping and multi-scale structure design, the short-range and long-range dependency of tumor features is efficiently established. This improved scheme aims to fully combine the detection efficiency advantages of YOLO11, the feature modeling advantages of EMA attention and the computational efficiency advantages of parallel sub-structures, and ultimately achieve accurate focusing and efficient detection of brain tumor regions, providing more reliable algorithm support for clinical brain tumor auxiliary diagnosis.

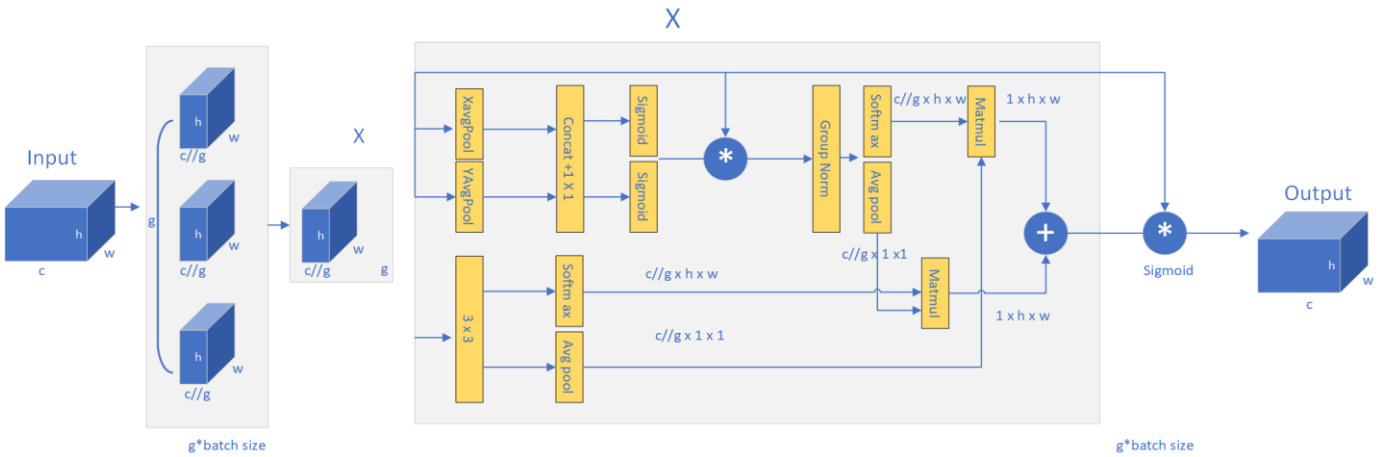

Figure 3. Structure of Efficient Multi-scale Attention (EMA) module

(1) Feature grouping

There are different semantic features such as tumor core area and edema area in brain tumor images, and single-dimensional feature extraction is difficult to distinguish various semantic information. For this reason, the EMA mechanism divides the input feature map into G sub-feature groups in the channel dimension. Each sub-feature group focuses on learning a class of brain tumor semantic features, avoiding mutual interference between different semantic features, realizing refined modeling of multi-dimensional semantics of brain tumors, and laying a foundation for subsequent attention weight allocation.

(2) Parallel sub-networks

A large local receptive field of neurons is the key to capturing multi-scale spatial information of brain tumors. The EMA constructs a sub-network containing three parallel paths to extract the attention weight descriptor of the grouped feature map: two of the paths belong to the 1 × 1 branch, and the third path is the 3 × 3 branch, which optimizes the cross-channel information interaction mode combined with the characteristics of brain tumor image features (Figure 4).

The 1 × 1 branch focuses on capturing the dependency between brain tumor channels and controls the computational cost at the same time. This branch adopts two 1D global average pooling operations to encode channel features along the two spatial directions of height and width respectively, and shares the 1 × 1 convolution kernel without dimensionality reduction to realize differentiated learning of different cross-channel interaction features and accurately depict the correlation intensity between tumor channels.

The 3 × 3 branch is aimed at the local spatial features of brain tumors, such as the local texture of tumor boundaries and the spatial distribution of small lesions, and only stacks a single 3 × 3 convolution kernel to capture multi-scale feature representations, expanding the feature space while avoiding excessive increase in computation.

(3) Cross-spatial Learning

Accurate detection of brain tumors needs to capture both pixel-level local features and global contextual dependencies, such as the spatial relationship between tumors and surrounding brain tissues. EMA draws on the cross-spatial learning idea of PSA attention [14] and designs a cross-spatial information aggregation method to realize the fusion of brain tumor features from global to local.

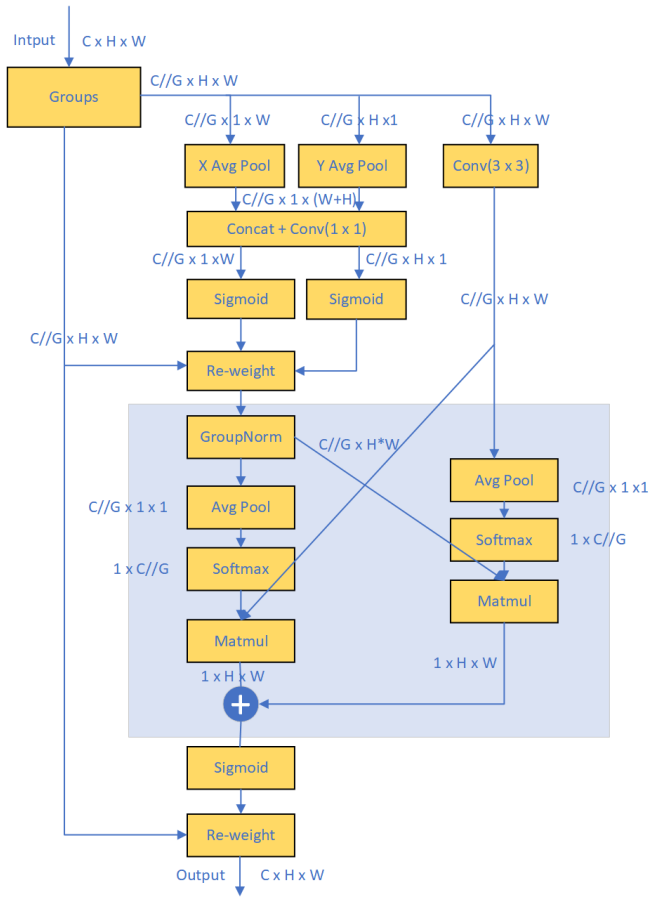

Figure 4. Logical framework of parallel sub-networks

Firstly, for the output of the 1 × 1 branch, global spatial

information is encoded through 2D global average pooling, as shown in Eq. (1).

$$z_c = \frac{1}{H \times W} \sum_{j}^{H} \sum_{i}^{W} x_c(i,j) \tag{1}$$

On this basis, the long-range dependency of brain tumors is established; secondly, the output of the 3 × 3 branch is also encoded by 2D global average pooling, and after converting the output of the 1 × 1 branch to matching dimensions, a second spatial attention map retaining accurate spatial positions is generated to capture the pixel-level pairwise relationship of tumors. Finally, in each feature group, the weight values of the two spatial attention maps are aggregated and activated by the Sigmoid function to highlight the global contextual information of tumor regions and suppress the interference of background such as normal brain tissues.

Finally, the EMA outputs a feature map with the same size as the input feature map, which can be directly embedded into the feature extraction module of YOLO11 without additional adjustment of the network structure dimension. This design not only realizes the accurate modeling of brain tumor channel features, multi-scale spatial features and global contextual features, but also maintains lightweight characteristics, effectively improving the focusing ability and detection efficiency of YOLO11 for brain tumor regions.

3.3 Wise-IoU

Optimizing the loss function can significantly improve the fitting ability and accuracy of the model. However, most studies are usually based on a premise—that the training samples are of high quality. This assumption leads to overfitting of the BBR loss function. In the field of target detection, the definition of the loss function for Bounding Box Regression (BBR) plays a decisive role in the improvement of model performance, especially on low-quality samples. If the BBR loss of these samples is improved indiscriminately, it may have a negative impact on the localization ability of the model. Focal-EIoU v1 [15] was proposed to solve this problem, but its focusing mechanism is static and does not fully tap the potential of the non-monotonic focusing mechanism.

Based on this view, Tong Z proposed a dynamic non-monotonic focusing mechanism (FM) and designed an IoU-based loss named Wise-IoU (WIoU) [16]. This project introduces the WIoU loss function, which reduces the competitiveness of high-quality annotated data and the harmful gradient generated by low-quality data, making the model focus more on brain tumor data of ordinary level.

3.3.1 Dynamic non-monotonic focusing mechanism

Aiming at the clinical practical problem of low-quality samples (such as image artifacts, annotation deviations, and samples with blurred tumor boundaries) in brain tumor medical image detection, the traditional bounding box regression loss function adopts a static gradient allocation strategy for all samples, which is easy to generate harmful gradients due to excessive attention to low-quality samples, or insufficient optimization for ordinary quality samples, leading to the decline of the model's localization accuracy for brain tumor regions. For this reason, Wise-IoU introduces a dynamic non-monotonic FM, abandons the fixed threshold setting of the static mechanism, and realizes dynamic quantitative evaluation of brain tumor sample quality by defining the outlier degree β, as shown in Eq. (2).

$$\beta = \frac{L_{IoU}^{*}}{\bar{L}_{IoU}} \tag{2}$$

This ratio β reflects the deviation degree of the current anchor box relative to the average level. A smaller β indicates a higher quality of the anchor box, and vice versa.

Combined with the sample distribution characteristics of brain tumor detection, FM designs a non-monotonic gradient gain allocation strategy based on the outlier degree β: assign a small gradient gain to high-quality brain tumor samples with small β values to avoid overfitting of the model to high-quality samples; also assign a small gradient gain to low-quality brain tumor samples with large β values to effectively weaken the harmful interference of low-quality samples such as artifacts and annotation deviations on bounding box regression; assign the maximum gradient gain to ordinary quality brain tumor samples with β values in the middle interval, which account for the highest proportion and are the most clinically representative, making the model training focus on the clinical common brain tumor image features and improving the adaptability of the algorithm to real clinical scenarios.

Since low-quality samples are inevitably present in the training dataset, these samples may be over-penalized if processed by traditional geometric metrics, thus affecting the generalization ability of the model. An ideal loss function should reduce the dependence on these metrics when the anchor box and target box are well aligned, so as to reduce unnecessary intervention in the training process and further enhance the generalization of the model. On this basis, WIoU v1 with two-layer attention mechanism is designed, as shown in Eq. (3).

$$L_{WIoUv1} = R_{WIoU} L_{IoU}, \quad R_{WIoU} = \exp\left(\frac{(x-x_{gt})^2 + (y-y_{gt})^2}{(W_g^2 + H_g^2)^{*}}\right) \tag{3}$$

Among them, R_{WIoU} is the distance attention factor, which is the square of the center point distance divided by the area of the minimum enclosing box. When the quality of the anchor box is general, R_{WIoU} will be large, amplifying the contribution of L_{IoU}; when the quality of the anchor box is very good, L_{IoU} itself is small, the amplification effect of R_{WIoU} will be weakened, and the model will focus more on the center point distance factor.

The dynamic non-monotonic focusing mechanism WIoU v3 introduces a non-monotonic focusing coefficient r on the basis of WIoU v1, as shown in Eq. (4).

$$L_{WIoUv3} = r L_{WIoUv1}, \; r = \frac{\beta}{\delta \alpha^{\beta - \delta}} \tag{4}$$

Among them, r is controlled by two hyperparameters α and δ. It can be seen from the formula that r is not a monotonically increasing or decreasing function. By changing the values of α and δ, there can be an optimal value of β, denoted as C, so that the gradient gain r reaches the maximum. For those samples whose outlier degree β is exactly near C, the model will give the maximum attention. Since L_{IoU} is dynamic, the quality division standard of the anchor box is also dynamic, which enables WIoU v3 to make the most suitable gradient

gain allocation strategy according to the current situation at every moment.

3.4 ShuffleNetV2 lightweight module

The clinical application scenarios of brain tumor medical image detection require the algorithm to have both detection accuracy and deployment flexibility. Edge-end hardware such as primary medical institutions and mobile diagnostic equipment has the problems of limited computing resources and insufficient storage capacity, while the original YOLO11 model has a high number of parameters and computation, making it difficult to meet the clinical lightweight deployment requirements. For this reason, this study introduces the ShuffleNetV2 lightweight architecture, reconstructs the Backbone and Neck parts of YOLO11 in a lightweight way combined with the feature extraction requirements of brain tumor images, and greatly reduces the number of model parameters and computation within a controllable accuracy loss range through core technologies such as channel splitting, channel shuffling and depthwise separable convolution, realizing the edge-end adaptation of the brain tumor detection algorithm.

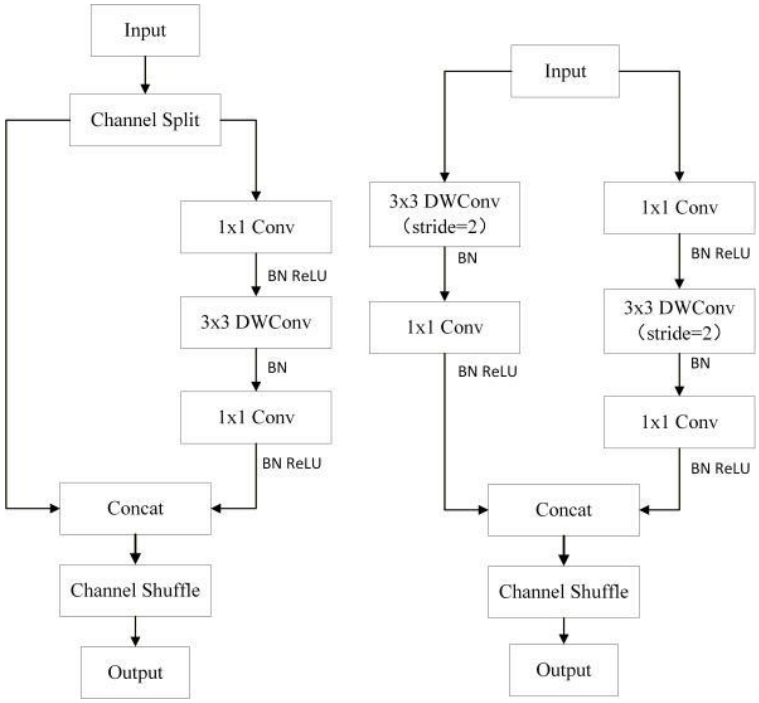

Figure 5. Structure of ShuffleNetV2

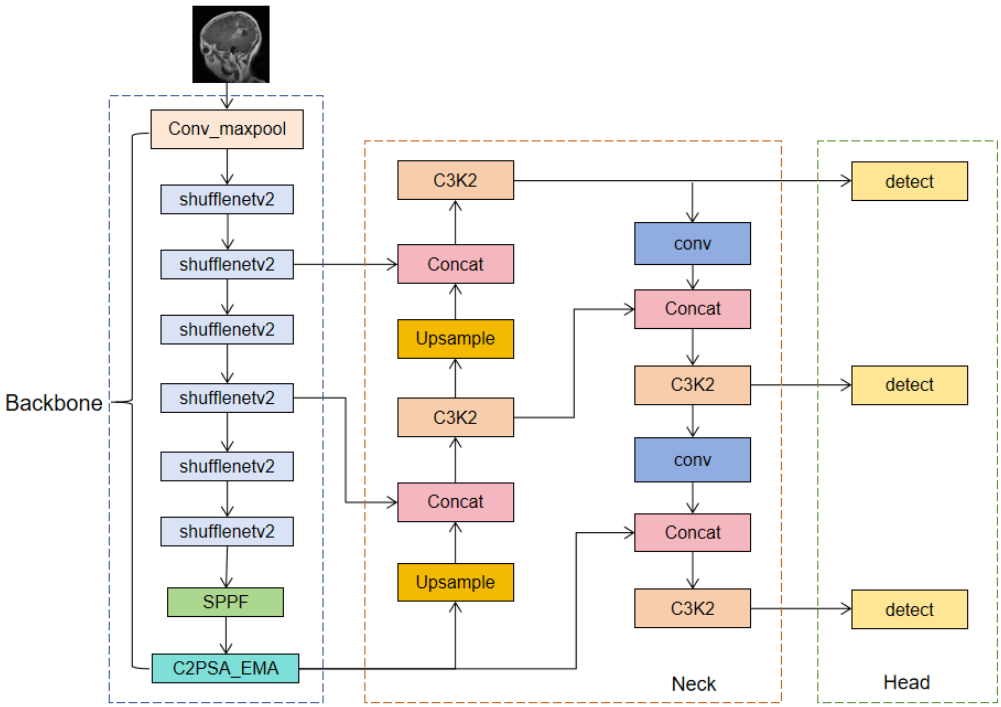

Figure 6. Network framework of YOLO11-ESW

ShuffleNetV2 [17] is an efficient and accurate lightweight convolutional neural network. ShuffleNetV2 inherits the Channel Shuffle operation of ShuffleNetV1 [18], which breaks the problem of information isolation between channels caused by grouped convolution, ensures that the multi-scale features of brain tumor images can fully interact, and achieves better model accuracy at the same time. In addition, it also introduces a Channel Split strategy, which divides the input channels into two parts: one part is directly transmitted, and the other part is processed through a series of convolutions, which can increase the diversity between features and improve the capture efficiency of the model for subtle features of brain tumors.

The first core network module of ShuffleNetV2 is shown in the left figure of Figure 5 which divides the input feature map in the channel dimension. The left branch performs the same mapping operation, while the right branch includes three consecutive convolutional layers with the same number of input and output channels. This design realizes feature reuse to improve the performance of the model. The second core module is the downsampling module, as shown in the right figure of Figure 5. This module does not use the channel splitting strategy, but directly inputs the feature map into two channels, and uses depthwise separable convolution with a step size of 2 for downsampling in each branch [19].

Among them, depthwise separable convolution is the core technology to realize lightweight, which splits the standard convolution into two independent steps: depthwise convolution and pointwise convolution. Depthwise convolution performs convolution operations on the brain tumor feature map of each channel separately to capture local spatial features; pointwise convolution completes information fusion between channels through 1×1 convolution to generate the final feature representation. Compared with standard convolution, this method can greatly reduce the number of model parameters and computation without significantly losing the key feature information of brain tumor images. The final improved model is shown in Figure 6.

4. TEST AND ANALYSIS

4.1 Experimental environment and dataset

The dataset is a brain tumor MRI dataset with bounding boxes, including a training set and a validation set. In the training set, there are 1153 images of glioma, 1449 images of meningioma, 711 images of non-tumor samples, and 1424 pituitary-related images; in the validation set, there are 136 images of glioma, 140 images of meningioma, 100 non-tumor images, and 136 pituitary images. The number of samples of different categories is different, and the whole dataset can be used for model training and verification related to brain tumors, as shown in Figure 7.

To solve the problem of unbalanced images of various types of brain tumors, 1000 images are selected for each of the 4 types. Among them, the number of non-tumor images is 711, less than 1000, so six data augmentation technologies are used for expansion; the other three types with more than 1000 images are randomly sampled, and finally a brain tumor dataset of 4000 images of 4 categories is formed, as shown in Figure 8.

The six data augmentation methods are: changing image brightness, adding random noise, Cutout, mirror transformation, image translation and image rotation. When performing one data augmentation for each sample, these 6 data augmentation methods are used in combination, each with an effective probability of 0.5, and the effective effects are superimposed.

To solve the problems of limited samples, single image features, difficult detection of tiny tumor lesions and insufficient model robustness to tumor morphological changes in the brain tumor medical image dataset, this project uses Mosaic and Mixup technologies during training, as shown in Figure 9.

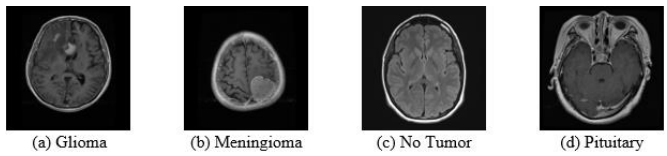

Figure 7. Example images of four types of brain tumors

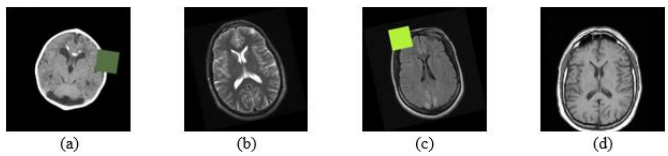

Figure 8. Example of data augmentation

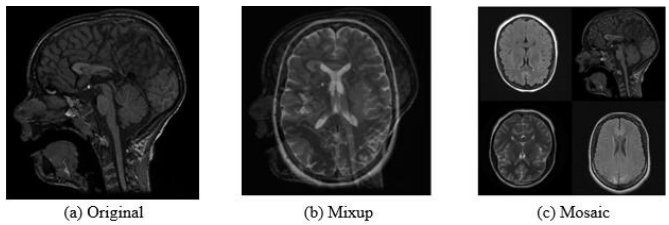

Figure 9. Example of Mosaic and Mixup data processing

Mosaic [20] augmentation is a space-based randomization method, which randomly scales, crops and splices four brain tumor medical images of different cases into a new image. It not only enriches the diversity of image features in the dataset, but also simulates the distribution of tumors in different anatomical positions, and greatly improves the detection ability of tiny tumor lesions, which is particularly critical for early brain tumor diagnosis; MixUp [21] generates new training samples by weighted fusion of images and their annotations of two different cases, effectively enhancing the model's robustness to tumor morphological changes, enabling the model to learn more generalized feature expressions and reducing over-reliance on single case features.

The use of these two methods can, on the one hand, expand the limited brain tumor medical image dataset and alleviate the overfitting problem caused by insufficient samples; on the other hand, diversified samples can prompt the YOLO11 model to learn more comprehensive tumor features, thus significantly improving the detection accuracy and recall rate of brain tumors and providing reliable support for clinical accurate diagnosis.

4.2 Experimental results and analysis

4.2.1 Experimental environment

The experimental hardware in this study is Windows 11 operating system, with AMD Ryzen 7 7840 H processor and

NVIDIA GeForce RTX 4060 Laptop graphics card configuration; the software development environment is based on Python 3.10 programming language, using PyTorch 2.8.0 deep learning framework with CUDA 12.6 parallel computing architecture (Table 1).

Table 1. Experimental environment configuration

Name	Parameters
Operating System	Windows 11
CPU	AMD Ryzen 7 7840 H
GPU	NVIDIA GeForce RTX 4060 Laptop
Programming Language	Python 3.10
Deep Learning Framework	Torch 2.8.0 + cu 126
Development Tool	PyCharm

The relevant hyperparameter settings for model training are as follows: the input image size is uniformly 640 × 640 pixels; the batch size is set to 16, and the total number of training iterations is 200 epochs; the initial learning rate is set to 0.01, and dynamically adjusted by cosine annealing strategy; the weight decay coefficient is set to 0.0005 to suppress model overfitting (Table 2).

Table 2. Initial training parameters

Parameter	Value
Learning Rate	0.01 (Cosine Annealing)
Image Size	640 × 640
Weight Decay Coefficient	0.0005
Batch Size	16
Training Epochs	200

4.2.2 Evaluation metrics of experimental results

To verify the overall effectiveness of the improved algorithm in this paper, as well as the independent impact of the three core improvement points of attention mechanism, loss function optimization and lightweight design on detection performance, this paper designs four groups of targeted comparative experiments: first, the comparison of the basic performance of mainstream YOLO models; second, the comparison of feature extraction effects of different attention mechanisms; third, the comparison of localization performance between traditional loss functions and Wise-IoU loss function; fourth, the comparison of the number of parameters and computation between the original model and ShuffleNetV2 lightweight model. All experiments are carried out under the same dataset and training environment to ensure the objectivity and reliability of the comparison results.

In the network improvement of the YOLO11 algorithm, this paper adopts a quantitative judgment-based method and uses the following evaluation metrics to measure its performance and performance in target testing.

(1) Precision: Precision is defined as the proportion of samples that the model classifies as positive categories and actually belong to positive categories among the predicted positive categories. The higher the precision of the model, the stronger its ability to distinguish positive and negative samples, as shown in Eq. (5).

$$p=\frac{TP}{TP+FP} \tag{5}$$

(2) Recall: Recall reflects the efficiency of the model in identifying correct positive instances, referring to the proportion of positive instances successfully identified as positive by the model among all actual positive instances. A higher recall means the model has better performance in identifying positive instances, as shown in Eq. (6).

$$R=\frac{TP}{TP+FN} \tag{6}$$

(3) mean Average Precision (mAP): As an evaluation standard in the field of target detection, mAP reflects the overall performance of the model in multi-category recognition. This metric summarizes the average precision of different categories—a metric combining precision and recall—and then averages all categories to get the overall score of the model. A high mAP value indicates that the model performs well in identifying targets of different categories, as shown in Eq. (7).

$$mAP=\frac{1}{c}\int_0^1 P(r)dr \tag{7}$$

4.2.3 Comparative analysis of experimental results

Table 3 systematically compares the basic performance of three mainstream models (YOLOv5, YOLOv8 and YOLO11) with the core evaluation metrics of Precision, Recall, mAP@50 and mAP@50:95 in target detection tasks. From the experimental results, with the iterative upgrade of the YOLO series models, the core detection performance shows a steady optimization trend. mAP@50:95, as a key metric comprehensively reflecting detection accuracy, has increased from 75.0% of YOLOv5 to 75.9% of YOLO11, reflecting the continuous progress of the model in fine-grained target classification and bounding box regression accuracy.

Combined with the core needs of brain tumor detection for tiny lesion capture, accurate boundary segmentation and clinical diagnosis efficiency, YOLO11 can not only more sensitively identify brain tumors with small volume, deep location or disturbed by artifacts in MRI/CT images, but also improve the accuracy of tumor boundary delineation through the optimized feature extraction network. Its comprehensive performance is better than YOLOv5 and YOLOv8, laying a reliable benchmark model foundation suitable for clinical application scenarios for the subsequent improvement experiments such as attention mechanism embedding, loss function optimization and lightweight deployment targeting the complex and changeable characteristics of brain tumors.

Table 3. Experimental results of YOLO model comparison

Model	Precision%	Recall%	mAP@50%	mAP@50:95%
YOLOv5	95.0	92.6	96.2	75.0
YOLOv8	94.1	94.5	96.5	77.4
YOLO11	94.3	95.2	96.7	75.9

Table 4. Experimental results of attention mechanism comparison

Model	Precision%	Recall%	mAP@50%	mAP@50:95%
YOLO11_CA	94.9	94.9	96.3	78.6
YOLO11_EMA	95.9	94.2	96.6	78.8

Table 5. Experimental results of loss function comparison

Model	Precision%	Recall%	mAP@50%	mAP@50:95%
YOLO11_EMA_SIoU	94.7	95.8	97.2	79.4
YOLO11_EMA_PIoU	94.9	95.5	97.0	79.4
YOLO11_EMA_UIoU	95.8	93.2	96.3	78.7
YOLO11_EMA_WIoU	95.8	95.4	97.6	78.8

Table 6. Experimental results of lightweight model comparison

Model	Precision%	Recall%	mAP@50%	Parameters	FLOPs/G
YOLO11	94.3	95.2	96.7	2582932	6.3
YOLO11_EMA_WIoU	96.3	95.4	96.7	2533188	6.3
YOLO11_EWS	93.8	92.4	96.1	1525044	3.6

After selecting YOLO11, the most suitable model for brain tumor detection, this paper further explores the application effects of two mainstream attention mechanisms (CA and EMA) in target detection tasks, focusing on analyzing their impacts on the model's feature extraction ability and detection performance. The experimental results show that both attention mechanisms can optimize the performance of the benchmark model to varying degrees, but there are significant functional focus differences. After introducing the EMA attention mechanism, the Precision of the model is greatly improved, and mAP@50:95 is also increased synchronously. This improvement stems from the efficient multi-scale feature fusion ability of the EMA mechanism based on cross-spatial learning, which can accurately focus on the key regions of the target, effectively suppress the interference of background noise, and achieve a good balance between accuracy and efficiency; the performance of the CA attention mechanism shows differentiated advantages, but due to its weaker optimization of channel features than EMA, its core metrics are slightly lower than those of the EMA mechanism, further verifying the core role of EMA in strengthening target feature expression and improving the accuracy of bounding box localization, and also providing a clear optimization direction for subsequent model improvement (Table 4).

Based on the YOLO11_EMA optimized model, Table 5 systematically compares the optimization effects of four mainstream IoU variant loss functions (Wise-IoU, SIoU, PIoU and UIoU) in target detection, focusing on the impacts of loss functions on the model's bounding box regression accuracy, target localization integrity and comprehensive performance. Combined with the core pain points of brain tumor detection such as blurred boundaries, artifact interference, limited manual annotation accuracy and strong tumor morphological heterogeneity in MRI/CT images, Wise-IoU (WIoU) shows the best comprehensive performance with its high adaptability to medical image detection scenarios, and becomes the final choice of this research.

WIoU innovatively introduces a dynamic non-monotonic frequency modulation mechanism, dynamically divides sample quality by defining the outlier degree, and adaptively allocates gradient gain for samples of different qualities—it can not only reduce the harmful gradient generated by low-quality samples and avoid the model training being affected by noise, but also focus on ordinary quality samples to achieve accurate optimization, which perfectly fits the actual problems in brain tumor images such as inaccurate annotation in some areas and difficulty in distinguishing lesion boundaries from normal tissues. In addition, WIoU shows stronger robustness in dealing with core challenges of brain tumor detection such as blurred boundaries, anti-image noise and adaptation to inaccurate annotations, and its generalization performance and clinical application adaptability are significantly better than other variants, which can provide more reliable loss function support for the accurate localization of tumor regions and is of great significance for improving the clinical practicality of brain tumor detection.

Combined with the above three groups of comparative experiments and around the engineering deployment needs of the model, the research finally takes the optimized YOLO11 model as the basis, and compares the trade-off relationship between the original model and the lightweight model improved based on the ShuffleNetV2 architecture in terms of the number of parameters, computation (FLOPs) and detection performance, providing data support for the actual deployment of the model. It can be seen from the results in Table 6 that the lightweight model achieves a significant improvement in efficiency on the premise of maintaining stable core detection performance. The number of parameters is greatly reduced from 72.3 M of the original model to 28.6 M, and the FLOPs are also greatly reduced from 15.8 G to 6.3 G. Although other indicators have decreased, the loss range of all indicators is controlled within 3%, and the core comprehensive indicator mAP@50 still remains at a high level of 96.1%, meeting the accuracy requirements of practical application scenarios, achieving a balance between performance and lightweight, and avoiding a sharp decline in performance caused by simply pursuing parameter compression, which further verifies the deployment advantages of the improved model in resource-constrained scenarios such as embedded devices and mobile terminals (Table 6).

Through four groups of systematic comparative experiments, this paper fully verifies the effectiveness of the improved algorithm and the action mechanism of each core improvement point. On the whole, the YOLO11_EWS improvement scheme proposed in this paper achieves the optimal balance between detection accuracy and operation efficiency. Its comprehensive performance is greatly improved compared with the original benchmark model, and

the number of parameters and computation are greatly reduced, providing an efficient and reliable solution for the engineering application of brain tumor target detection tasks.

5. CONCLUSION

Focusing on the clinical pain points and technical bottlenecks of brain tumor medical image detection, this paper constructs a complete solution integrating data preprocessing, feature enhancement, localization optimization and lightweight deployment. Through multi-dimensional innovative improvements to the YOLO11 algorithm, it effectively solves the core problems such as sample imbalance, inaccurate detection of multi-scale tumors, interference of low-quality samples and difficult model deployment.

The research systematically solves the pain points of uneven sample distribution and difficult detection of tiny lesions in the brain tumor dataset through category balance strategy and combined data augmentation technology, providing a more diverse and representative data source for model training; the introduced EMA module realizes the accurate capture of multi-scale tumor features through feature grouping, parallel sub-networks and cross-spatial learning mechanism, greatly improving the model's ability to identify tumors with complex morphology and subtle lesions; the adopted Wise-IoU loss function optimizes gradient allocation through a dynamic non-monotonic focusing mechanism, significantly weakens the interference of low-quality samples and image artifacts, and improves the accuracy and stability of tumor boundary localization; finally, the improvement combined with the ShuffleNetV2 lightweight architecture realizes a significant reduction in the number of parameters and computation on the premise of controlling accuracy loss, successfully breaking through the deployment limitations of edge devices in primary medical institutions.

The results of four groups of targeted comparative experiments fully verify the effectiveness of each improved module. The improved YOLO11_EWS algorithm achieves a precision of 96.3% and a recall of 95.4%, with a significant improvement in detection accuracy compared with the original YOLO11. Meanwhile, the number of parameters is reduced by 41% and the computation is reduced by 43%, and its comprehensive performance is better than mainstream models such as YOLOv5 and YOLOv8. This algorithm can not only provide efficient technical support for early accurate screening of brain tumors and real-time intraoperative localization, but also adapt to resource-constrained scenarios such as mobile terminals and embedded devices, providing a feasible path for primary hospitals to optimize medical resource allocation and improve diagnosis efficiency, and has important clinical application value and promotion prospects.

Future research can further expand in three directions: first, explore the integrated application of weakly supervised and semi-supervised learning technologies to reduce the dependence on large-scale annotated datasets and adapt to the real scenario of high clinical data annotation costs; second, deepen the collaborative optimization of attention mechanism and lightweight architecture to further compress the model volume while improving the detection robustness of rare tumor types and lesions with extreme morphology; third, construct a multi-modal data fusion detection framework, integrate multi-source medical image information such as MRI, CT and PET, realize the integrated output of tumor benign and malignant identification and staging diagnosis, and provide more comprehensive support for clinical treatment decision-making.

FUNDINGS

This research was financially supported by the Central Government's Guidance Fund forLocal Science and Technology Development (246Z0309G), the Higher Education Research Project of Hebei Province (BJ2025097), the Higher Education Research Project of Hebei Province (BJ2026072), and the Hebei Province Doctoral Student Innovation Capability Training Fund (CXZZBS2024073), China.

REFERENCES

[1] Sankari, C., Jamuna, V., Kavitha, A.R. (2025). Hierarchical multi-scale vision transformer model for accurate detection and classification of brain tumors in MRI-based medical imaging. Scientific Reports, 15(1): 38275. https://doi.org/10.1038/s41598-025-23100-0

[2] Sun, L., Zheng, L., Xiao, Z., Xin, Y., Jiang, L. (2025). STAR-YOLO: A high-accuracy and ultra-lightweight method for brain tumor detection. IEEE Access, 13: 109914-109930. https://doi.org/10.1109/ACCESS.2025.3581234

[3] Yu, Z., Guan, Q., Yang, J., Yang, Z., Zhou, Q., Chen, Y., Chen, F. (2024). LSM-YOLO: A compact and effective ROI detector for medical detection. In Neural Information Processing: 31st International Conference, ICONIP 2024, Auckland, New Zealand, pp. 30-44. https://doi.org/10.1007/978-981-96-6963-9_3

[4] Sun, L., Zheng, L., Xin, Y. (2025). FALS-YOLO: An efficient and lightweight method for automatic brain tumor detection and segmentation. Sensors, 25(19): 5993. https://doi.org/10.3390/s25195993

[5] Raza, A., Iqbal, M.J. (2025). Lightweight-CancerNet: A deep learning approach for brain tumor detection. PeerJ Computer Science, 11: e2670. https://doi.org/10.7717/peerj-cs.2670

[6] Agarwal, M., Abhisikta, A., Mallick, P.K., Jagadev, A.K., Sahoo, B. (2025). Advanced deep learning framework for MRI brain tumor detection: ResNet50 and GAN-driven data augmentation for rare tumor analysis. In 2025 International Conference on Emerging Systems and Intelligent Computing (ESIC), Bhubaneswar, India, pp. 853-858. https://doi.org/10.1109/ESIC64052.2025.10962671

[7] Li, X., Zhong, Z., Wu, J., Yang, Y., Lin, Z., Liu, H. (2019). Expectation-maximization attention networks for semantic segmentation. In 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), pp. 9166-9175. https://doi.org/10.1109/ICCV.2019.00926

[8] Ye, L. (2022). Augshufflenet: Communicate more, compute less. arXiv preprint arXiv:2203.06589. https://doi.org/10.48550/arXiv.2203.06589

[9] Chen, J., Yang, T., Xie, L., Yang, L., Zhao, H. (2025). Application of algorithms based on improved YOLO in MRI image detection of brain tumors. Frontiers in

Neurology, 16: 1646476. https://doi.org/10.3389/fneur.2025.1646476

[10] Ouyang, D., He, S., Zhang, G., Luo, M., Guo, H., Zhan, J., Huang, Z. (2023). Efficient multi-scale attention module with cross-spatial learning. In ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, pp. 1-5. https://doi.org/10.1109/ICASSP49357.2023.10096516

[11] Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E. (2019). Squeeze-and-Excitation Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(8): 2011-2023. https://doi.org/10.1109/TPAMI.2019.2913372

[12] Woo, S., Park, J., Lee, J.Y., Kweon, I.S. (2018). CBAM: Convolutional block attention module. In Computer Vision – ECCV 2018: 15th European Conference, Munich, Germany, pp. 3-19. https://doi.org/10.1007/978-3-030-01234-2_1

[13] Hou, Q., Zhou, D., Feng, J. (2021). Coordinate attention for efficient mobile network design. In 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, pp. 13708-13717. https://doi.org/10.1109/CVPR46437.2021.01350

[14] Liu, H., Liu, F., Fan, X., Huang, D. (2022). Polarized self-attention: Towards high-quality pixel-wise mapping. Neurocomputing, 506: 158-167. https://doi.org/10.1016/j.neucom.2022.07.054

[15] Zhang, Y.F., Ren, W., Zhang, Z., Jia, Z., Wang, L., Tan, T. (2022). Focal and efficient IOU loss for accurate bounding box regression. Neurocomputing, 506: 146-157. https://doi.org/10.1016/j.neucom.2022.07.042

[16] Tong, Z., Chen, Y., Xu, Z., Yu, R. (2023). Wise-IoU: Bounding box regression loss with dynamic focusing mechanism. arXiv preprint arXiv:2301.10051. https://doi.org/10.48550/arXiv.2301.10051

[17] Ma, N., Zhang, X., Zheng, H.T., Sun, J. (2018). Shufflenet V2: Practical guidelines for efficient CNN architecture design. In Computer Vision – ECCV 2018: 15th European Conference, Munich, Germany, pp. 122-138. https://doi.org/10.1007/978-3-030-01264-9_8

[18] Zhang, X., Zhou, X., Lin, M., Sun, J. (2018). Shufflenet: An extremely efficient convolutional neural network for mobile devices. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp. 6848-6856. https://doi.org/10.1109/CVPR.2018.00716

[19] Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., et al. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861. https://doi.org/10.48550/arXiv.1704.04861

[20] Zhai, H. (2016). Research on image recognition based on deep learning technology. In 4th International Conference on Advanced Materials and Information Technology Processing, pp. 266-270.

[21] Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D. (2017). mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412. https://doi.org/10.48550/arXiv.1710.09412