(Received 12 June 2024, Accepted 14 September 2024, Scheduled 26 September 2024)

YOLOv8-DEC: Enhancing Brain Tumor Object Detection Accuracy in Magnetic Resonance Imaging

Zekun Lin, Weiming Lin*, and Fuchun Jiang

School of Opto-Electronic and Communication Engineering, Xiamen University of Technology, Xiamen 361024, China

ABSTRACT: Brain tumors are characterized by the fast growth of aberrant brain cells, which poses a considerable risk to an adult's health since it can result in severe organ malfunction or even death. Magnetic resonance imaging (MRI) provides vital information for comprehending the nature of brain tumors, directing treatment approaches, and enhancing diagnostic precision. It displays the diversity and heterogeneity of brain tumors in terms of size, texture, and location. However, manually identifying brain tumors is a difficult and time-consuming process that could result in errors. It is proposed that an enhanced You Only Look Once version 8 (YOLOv8) model aids in mitigating the drawbacks associated with manual tumor detection, with the objective of enhancing the accuracy of brain tumor detection. The model employs the C2f_DySnakeConv module to improve the perception and discrimination of tumors. Additionally, it integrates Content-Aware ReAssembly of FEatures (CARAFE) to efficiently expand the network's receptive area to integrate more global contextual information, and Efficient Multi-Scale Attention (EMA) to improve the network's sensitivity and resolution for lesion features. According to the experimental results, the improved model performs better for brain tumor detection than both the open-source model and the original YOLOv8 model. It also achieves higher detection accuracy on the brain tumor image dataset than the original YOLOv8 model in terms of precision, recall, mAP@0.5, and mAP@0.5~0.95 above, respectively, of 2.71%, 2.34%, 2.24%, and 3.73%.

1. INTRODUCTION

Brain and spinal cord make up Central Nervous System (CNS), which controls several essential processes like decision-making and tissue organization [1]. Because of the complex architecture of the brain, diagnosing and treating CNS diseases, particularly brain tumors, present substantial obstacles. These tumors come in over 130 different varieties, ranging from benign to malignant [2]. They can be classified as primary tumors, meaning that they originate within or surrounding brain cells, or secondary tumors, meaning that they occur from cancer spreading to the brain. Accurate tumor diagnosis is crucial but challenging due to invasiveness, subjectivity, and flaws in traditional biopsy and histological grading systems. For effective therapeutic methods to be developed for the management of brain tumors, a timely and correct diagnosis is crucial [3].

Without question, Magnetic Resonance Imaging (MRI) has established itself as a major tool for the diagnosis of brain tumors. As a diagnostic modality, it is thorough and noninvasive, providing a complete view of cerebral anatomy without the dangers of ionizing radiation. Rich tissue contrast and high-resolution visualization are presented by this sophisticated imaging method, which makes use of axial, coronal, and sagittal multiplanar imaging capabilities [4]. However, despite the numerous advantages of MRI, reliance on subjective interpretation of MRI scans by radiologists poses a number of significant challenges. The inherent intricacies of neuroimaging, coupled with the necessity to meticulously examine numerous imaging slices and planes, make the process time-consuming and resource-intensive [5]. In addition, the potential for discrepancies in diagnostic conclusions between different observers highlights an important limitation, as it introduces an element of uncertainty that may affect clinical decision-making and, in turn, patient care.

Furthermore, the present techniques [6, 7] frequently fall short of expectations in terms of diagnostic accuracy, particularly when it comes to MRI-detected brain tumors. The complex and variable morphology of some brain tumors may lead to delayed detection or misdiagnosis. Therefore, in order to improve the accuracy and reliability of MRI diagnosis, better detection techniques are desperately needed.

The research in brain tumor categorization, detection, and segmentation has seen significant advancements. Various techniques have been explored to enhance classification accuracy, including residual networks [8] with skip connections to mitigate overfitting and reduce model complexity. U-Net architecture, featuring convolutional and deconvolutional layers [9], is well suited for biomedical image segmentation tasks. Innovations such as integrated batch normalization in convolutional neural networks (CNNs) have been proven effective for identifying gliomas and stroke lesions [10]. Multi-layer models [11] and machine learning [12] approaches have improved MRI tumor detection and classification across different modalities. Texture extraction methods [13], adaptive independent subspace analysis [14], and loss function optimization [15] have also contributed to these advancements. Additionally, systems combining CNNs with multimodal data [16] and transfer learning techniques [17] have enhanced tumor detection precision.

* Corresponding author: Weiming Lin (linwm@xmut.edu.cn).

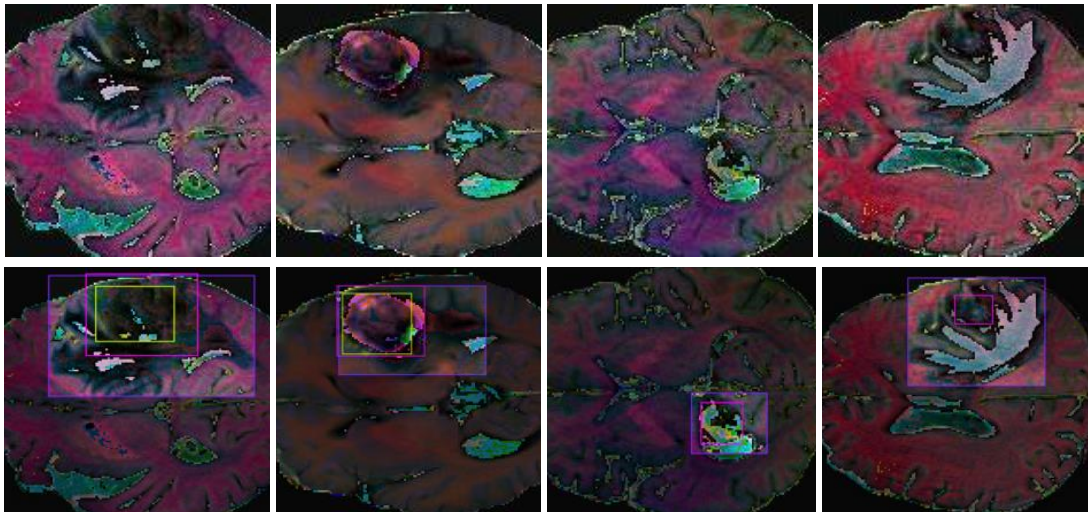


FIGURE 1. MRI images. First column shows a sample of the partial dataset and Second column shows the results of the target detection, where the boxes highlight the different tumor types, yellow for tumor core, blue for total tumor, and purple for enhanced tumor core.

Adaptive neuro-fuzzy systems [18] and genetic algorithms [19] have further expanded the range of tools available for brain tumor classification and grading.

A new era in medical image analysis has been ushered in by the development of deep learning models in recent years, particularly in the area of object detection. The You Only Look Once (YOLO) series, with YOLOv8 being the most developed, has a lot of promise in these developments for improved real-time object detection accuracy. YOLOv8 is a single-stage detection algorithm that lowers latency while maintaining very competitive detection performance in order to concurrently localize and recognize objects. When being applied to the difficult task of brain tumor object detection, it needs to be fine-tuned and developed in order to reach its maximum potential in the particular complexity of neuroimaging.

Thus, the purpose of this study is to enhance and modify the YOLOv8 framework in order to overcome the limitations of the existing brain tumor detection techniques. Our objective is to precisely enhance YOLOv8's brain tumor detection accuracy by making deliberate architectural modifications. Through thorough assessment and verification of these enhancements, our goal is to construct a more resilient and precise diagnostic framework that diminishes the ambiguity linked to manual interpretation, streamlines clinical judgment, and ultimately leads to better patient treatment and results in the demanding domain of brain tumor treatment.

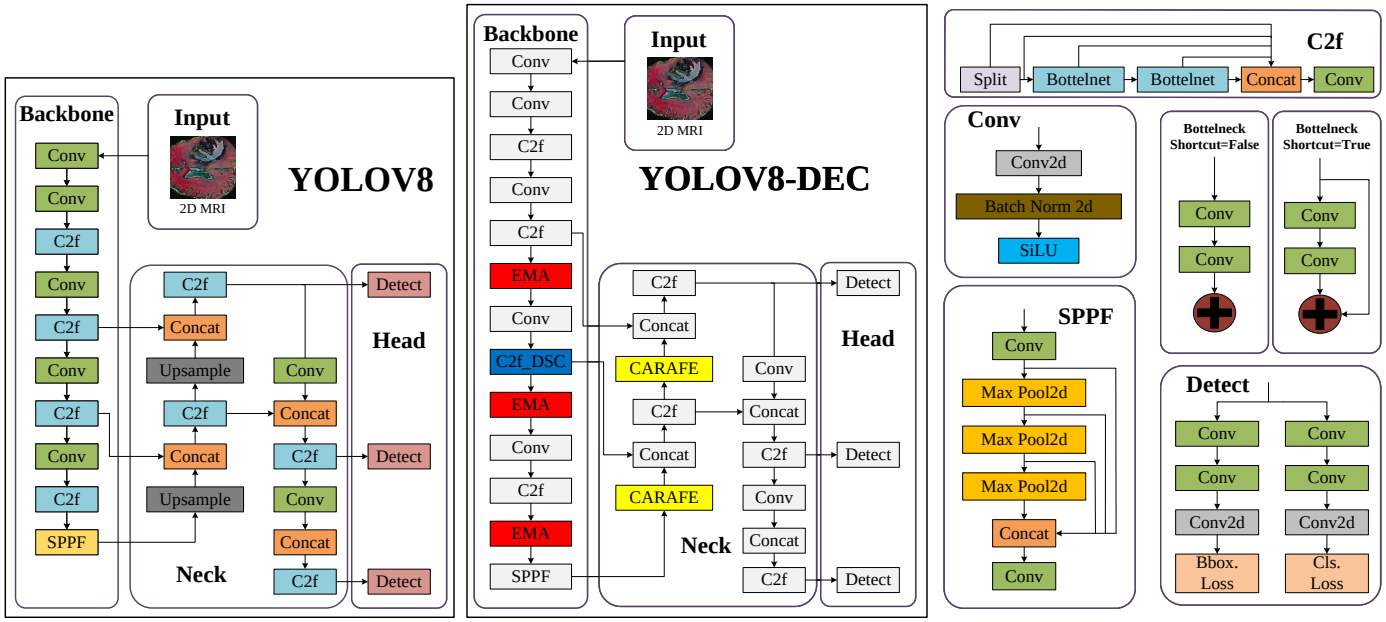
2. MATERIALS AND METHODS

2.1. Datasets and Preprocessing

This work uses a publicly available dataset from universe.roboflow.com [20] to guarantee the validity and efficacy of model training. There are 9,900 sample images in the dataset, and every image has at least one tumor annotation. They are T1-weighted images whose size is 139×132. With 6,072 samples for tumor core, 9,651 samples for total tumor, and 5,802 samples for enhanced tumor core, the dataset specifically covers a range of tumor labels, offering a large number of samples for model testing and training. We split the dataset into a training set, a test set, and a validation set in a $7 : 2 : 1$ ratio in order to make the most use of the data and guarantee the scientific quality of model training. The model was trained using the training set, which comprised 6,930 images. The validation set, which consists of 990 images, was used to tune hyperparameters and provide an initial assessment of the model. Finally, the model's detection accuracy and generalization capacity were evaluated using the 1,980 image test set.

Figure 1 shows the before and after comparisons of preprocessed brain tumor MRI dataset images after target detection. It has been color-mapped to improve the visual effect and identify various tissues or features. The grayscale values of the original MRI image were mapped into different colors, which converted the grayscale MRI images into RGB pseudo-color images. It is essentially a visual depiction of the original grayscale MRI picture after processing, where important characteristics are made more noticeable and easily identifiable by applying different colors to distinct ranges of grayscale values.

To reduce the risk of overfitting and improve the model's capacity for generalization, we used a static cropping picture preprocessing technique in this work. In order to focus on the crucial information that contains the tumor and significant surrounding anatomical structures, we systematically retained 24% to 82% of the horizontal width and 22% to 77% of the vertical height of the central region of each image during the preprocessing process using a four-week crop. By removing unnecessary information from the image's edge, it helps to concentrate the model's training on the key diagnostic features. This indirectly aids the model's ability to fully comprehend and learn about the tumor region and how it interacts with the surrounding tissues. By doing this, it guarantees that the model



Note: Conv is a convolution operation, Concat is the feature connection module, Upsample is upsampling module, Detect is the detection head, SPPF is the spatial pyramid pooling module, Max pool2d is the maximum pooling, Bbox.Loss is bounding box loss, Cls.Loss is classification loss. C2f_DySnakeConv is C2f with dynamic snake convolution module, Conv is a convolution operation, EMA is Efficient Multi-Scale Attention, CARAFE is upsampling network structure, Concat is the feature connection module, Detect is the detection head.

FIGURE 2. Structure comparison diagram.

may focus more on examining and recognizing the minute alterations and anatomical characteristics that are essential for diagnosis when working with actual cases, enhancing the localization accuracy and resilience to intricate situations in brain tumor detection jobs.

2.2. YOLOv8 Model Structure

In the evolution of YOLO framework, YOLOv8 stands out for striking a remarkable balance between speed and accuracy, along with noticeably enhanced detection capabilities. The ease with which it operates in practical application scenarios demonstrates its efficacy. Please refer to Figure 2 for the model's specific architecture arrangement. Three main components make up the architecture of YOLOv8: Head, Neck, and Backbone. The Backbone is based on CSPDarkNet53, which recovers features and optimizes them with residual connections for deep learning by using Cross Stage Partial (CSP) Networks. The Cross Convolutional Fusion (C2f) module of the neck efficiently integrates multi-scale features, and structural simplification refines the feature hierarchy, which is akin to the Advanced Feature Pyramid Network (FPN). Lastly, by managing categorization and bounding box prediction separately, the Head uses a decoupled design to increase detection accuracy and processing efficiency. YOLOv8's tripartite structure increases processing efficiency and detection accuracy in a synergistic way.

2.3. Improvement Methods

With improved YOLOv8 as its core, this study suggests a new brain tumor detection approach that aims to address the issues of low detection accuracy, high misdiagnosis rate, and difficulty in early identification in existing brain tumor detection methods. To achieve better detection performance, the method's primary objective is to optimize and improve three crucial factors.

The enhanced model utilized in this study is known as YOLOv8-DEC network, and Figure 2 depicts its architectural layout. In this work, we offer a sophisticated YOLOv8-based brain tumor detection model that has been improved by integrating three critical technologies: Content-Aware ReAssembly of FEatures (CARAFE) on the Neck structure, Efficient Multi-Scale Attention Module (EMA) interspersed in the Backbone structure, and Dynamic Snake Convolution (DSConv) fused in the Backbone.

By combining DSConv with the C2f module of YOLOv8 to generate C2f_DySnakeConv, the model achieves a substantial improvement in detection accuracy. EMA improves the model's capacity to identify tiny lesions at different sizes, and CARAFE enhances feature upsampling, enhancing global context knowledge, and increasing computational efficiency and detection precision. When taken as a whole, these improvements enable the model to diagnose complex brain tumor situations with increased accuracy and resilience.

2.4. C2f_DySnakeConv

The original YOLOv8 model may have difficulties in effectively detecting minor changes in complicated tumor morphology and surrounding structures, especially when the tumor is close to vascular structures, in the context of brain tumor detection. We address these issues by adding dynamic snake convolution (DSConv) [21] to the YOLOv8 C2f module, creating a new module called C2f_DySnakeConv.

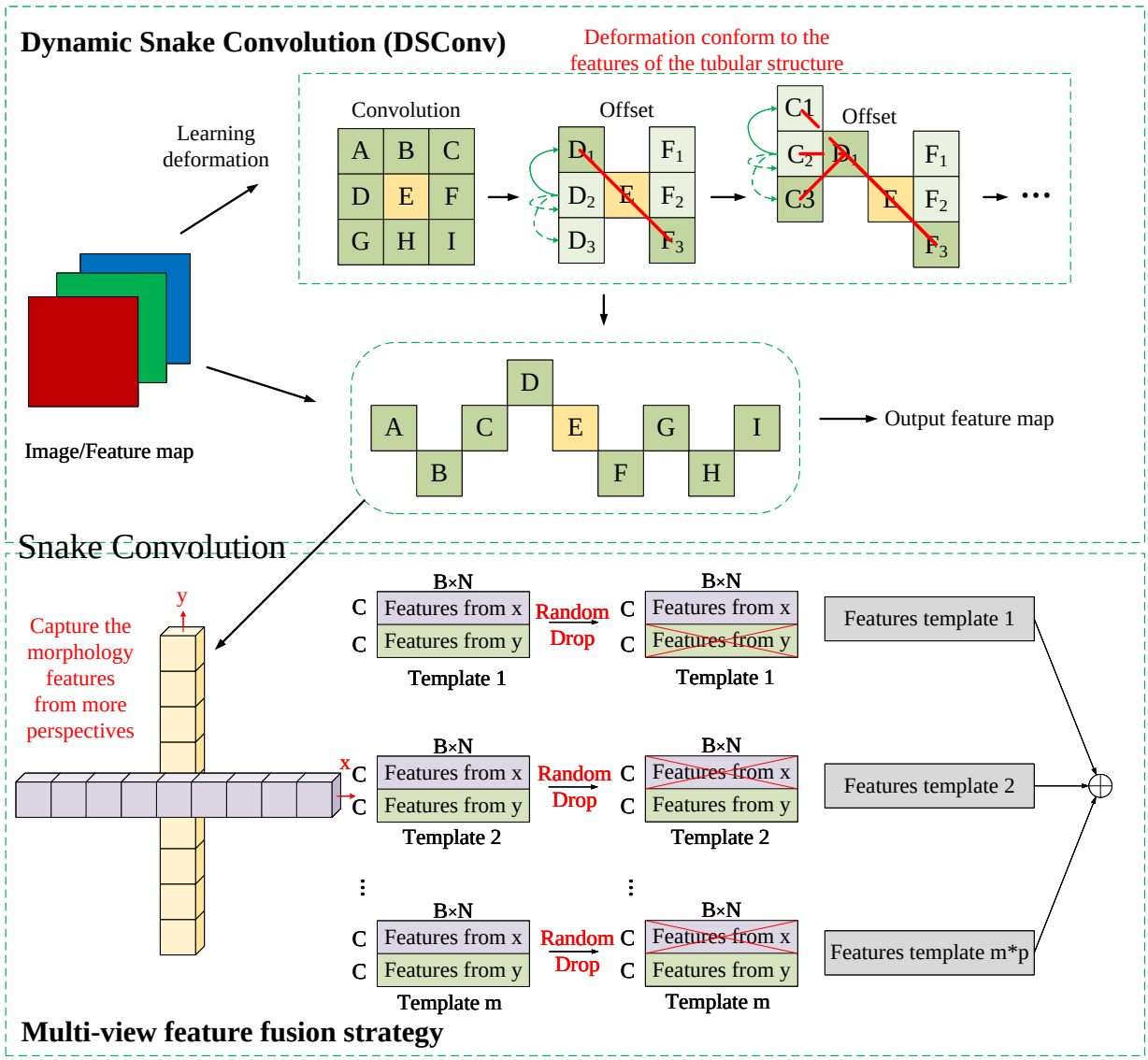


FIGURE 3. DySnakeConv structure.

The DSConv can detect tiny structural changes in brain tumors and the vasculature around them with great sensitivity by adaptively focusing on local elongated and tortuous features. In Figure 3, the intricate construction is displayed. This improves the model's ability to detect small lesions and complex tumor boundaries. Standard convolution kernels in DSConv are designed to freely deform along the x and y axes in order to account for the twisted morphology of tubular structures. Using a 3×3 convolution kernel as an example, all of the elements' initial positions are fixed. But in DSConv, these places are dynamically adjusted according to the offset of the position that comes before it. In order to ensure that the convolution kernel bends with the structure while paying attention to the object structure, this adjustment is made by adding up previous offsets.

Multi-view feature fusion strategy is the cornerstone of the C2f_DySnakeConv module. Figure 3 shows the architecture in detail. The objective is to proficiently handle the intricate and varied worldwide morphology of tubular formations and enhance the precision of segmentation by thoroughly examining their inherent structural data. The technique discussed is cleverly built on DSConv, which uses a thorough grasp of tubular structure morphology to build a large collection of morphological kernel templates. The key feature of DSConv is its ability to learn deformations based on input feature maps, allowing it to adaptively zero in on slender and winding local features while ensuring that the receptive field does not veer off-target due to significant deformation offsets, which is especially important when dealing with thin-walled tube structures.

Learning deformation is applied to the Image/Feature map, producing various Offest degrees dependent on convolution. At each layer of the Offset, features are extracted in both the X- and Y-axis directions, and DSConv is used to adaptively focus on the tubular structure to create numerous sets of multi-

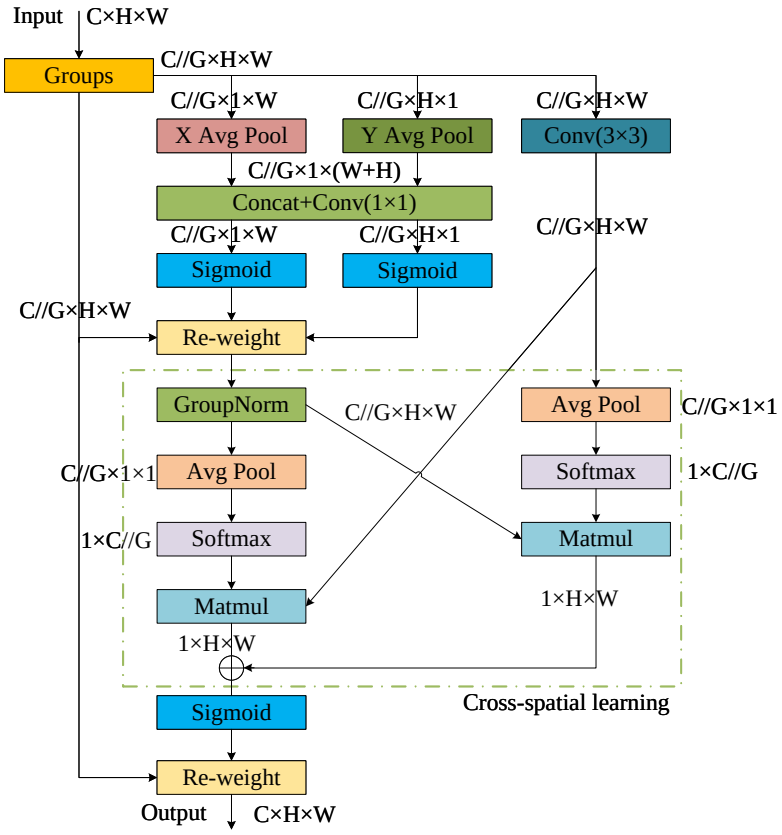


FIGURE 4. EMA network structure.

view Templates. During training, a Random Drop technique is applied to reduce fusion redundancy and prevent overfitting. By doing this, the model is able to better identify the target structure, increasing accuracy and decreasing needless computational effort.

2.5. Attention Mechanism Module

By supporting multi-scale feature learning and attaining, a more effective allocation of contextual attention, the Efficient Multi-Scale Attention (EMA) [22] mechanism, addresses these issues. This improvement effectively addresses the previously noted issues and boosts YOLOv8's performance in brain tumor identification tasks by significantly increasing sensitivity to tumor details and overall detection accuracy.

Figure 4 shows the EMA module's architecture in detail. Before exploring the EMA module's architecture and operation, let us explain its fundamental design principle: an inventive feature grouping technique expertly strikes a balance between computing efficiency and the variety of feature representations. To be more precise, the input feature map's C channels are separated into subsets of G equal or nearly equal numbers, and each subset, that is, a feature group, contains C/G channels. Moreover, it sets up a parallel processing environment, makes full use of the multi-core and parallel computing power of current computing platforms, and speeds up the process of training and inferring models. The job requirements, computing resources, and model complexity should all be considered when determining the number of groupings G. The choice of G is closely tied to the feature space's subdivision and the model's capacity for parallel processing.

The dual-path parallel processing technique of the EMA module is its essential component. In order to optimize the feature map from a global perspective, the 1×1 convolutional kernel's channel attention branching concentrates on two things: first, fine-tuning the global contextual information and realigning each channel's contribution; conversely, the 3×3 convolutional kernel is introduced with the intention of capturing the local features, specifically the fine-grained representation of the spatial structure. This allows the model to recognize and make use of the multiscale properties that span across the microscopic to the macroscopic scale. This two-pronged approach gives the model a thorough and in-depth comprehension of visual features.

The fusing mechanism of the EMA module is notably its essence. Cross-dimensionality is used to interactively fuse the outputs of two parallel branches. This strengthens information exchange between channels and preserves spatial position information at the pixel level, greatly enhancing the model's theoretical and practical understanding and representation of complex scenes. Equation (1) illustrates how the 2D global average pooling process works.

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

In this case, the value at location (i, j) in the input feature map is represented by x_{ij}, where H and W stand for the feature map's height and breadth, respectively. A preliminary attention map is then obtained by nonlinearly transforming the pooled results using the Softmax function. Then, to create a richer feature representation, features from two concurrent branches are combined by matrix multiplication.

The EMA module can improve the breadth and depth of feature learning, particularly when working on very complicated tasks that need to meet precise accuracy requirements, such as brain tumor image recognition. This highlights the model's enormous potential and importance in the field of medical image analysis, as it significantly enhances the model's performance while preserving high efficiency in areas like tumor detection.

2.6. CARAFE

While YOLOv8 has received a lot of attention for its effective speed and overall detection performance in brain tumor object identification, it has several drawbacks when it comes to small, complicated, and ambiguously bordered brain tumor objects. The main reason for this is that the convolutional neural network's upsampling processes may cause the loss of subtle local characteristics, especially in the network's neck region. Wang et al. [23] presented an inventive architecture named Content-Aware ReAssembly of FEatures (CARAFE) to solve this problem. By increasing the receptive field of the model and employing input feature maps to direct the upsampling process, they hoped to improve the integration of contextual information and feature utilization.

Figure 5 shows the CARAFE upsampling mechanism's complete architecture. A content-aware reconstruction module plus a kernel prediction module make up CARAFE's core division. To reduce the complexity of subsequent operations, the first input $H \times W \times C$ dimensional feature map is channel reduced in the kernel prediction module and transformed into a $C_m \times H \times W$ format. Through efficient dimension compression, computational optimization is accomplished in this step. Subsequently, convolutional kernels of size $k_{up} \times k_{up}$ are employed for the content encoding process, with the goal of producing the reassembly kernel, which is one of the essential parts for feature reassembly. After the formula explanation, this procedure finally produces a feature map size transformation that reflects an effective input information reorganization technique. Equation (2) is as follows:

$$C_m = \sigma^2 \times k_{up}^2 \tag{2}$$

Among these, σ, the upsampling ratio, is usually set to 2 with the intention of enhancing the feature representation, while C_m indicates the number of channels in the feature layer following dimensionality reduction. The dimension size that is utilized to predict the upsampling kernel is represented by the variable k_{up}. This is followed by a spatial expansion and reconfiguration of the channel data to create an upsampling kernel structure of size $\sigma H \times \sigma W \times {k_{up}}^2$. The Softmax function is then applied right after this to make sure that all weights add up to one, and normalization is achieved.

An efficient up-sampling kernel is used to recover and extract feature information before moving on to the content-aware rearrangement stage. A small area with a side length of k_{up} is chosen around each specified coordinate in the output feature map, and an element-wise multiplication with the upsampling kernel for that coordinate is then carried out. This operation's result is then merged back into the initial feature map, rebuilding a $\sigma H \times \sigma W \times C$ feature representation.

3. EXPERIMENT AND ANALYSIS

3.1. Experimental Environment and Parameter Configuration

The experiment was conducted on a 64-bit Ubuntu operating system with kernel version 4.15.0-213-generic. The GPU used was NVIDIA GeForce RTX 3090 with 24 GB of memory, and the host had 32 GB of RAM. The programming language utilized was Python 3.8.10, with CUDA v11.1 employed for GPU acceleration. The training was performed based on the deep learning framework PyTorch 1.9.1. Table 1 displays the training parameter configurations.

TABLE 1. Training parameter setting.

Parameter	Value	Parameter	Value
epochs	300	optimizer	SGD
patience	50	weight_decay	0.0005
batch	16	momentum	0.937
image_size	640	mask_ratio	4
workers	8	close_mosaic	10
final_learning_rate	0.01	patience	50

3.2. Evaluation Metrics

In this experiment, mean Average Precision (mAP), recall, and precision are used to evaluate the model. The mean is denoted by 'm' in mean Average Precision (mAP), and the average precision for this class of samples when the confusion matrix's Intersection over Union (IoU) threshold is set to 0.5 is denoted by AP@0.5. The mean accuracy of every sample class, or mAP@0.5, illustrates how the accuracy of the model varies with recall. A greater number signifies that the model sustains greater accuracy even at elevated recall percentages. With a step size of 0.05, the average mAP values at various IoU thresholds, spanning from 0.5 to 0.95, are represented by the variable mAP@0.5~0.95.

3.3. Experimental Result

3.3.1. Comparison Experiment of Mainstream Algorithms

Focused on brain tumor detection, a thorough comparative analysis was systematically conducted. This study's main goal is to thoroughly test and assess our newly proposed model's technical benefits and practical usefulness for the task of brain tumor object identification. In addition to comparison experiments with the original model YOLOv8. We have carefully chosen a number of representative models for comparison

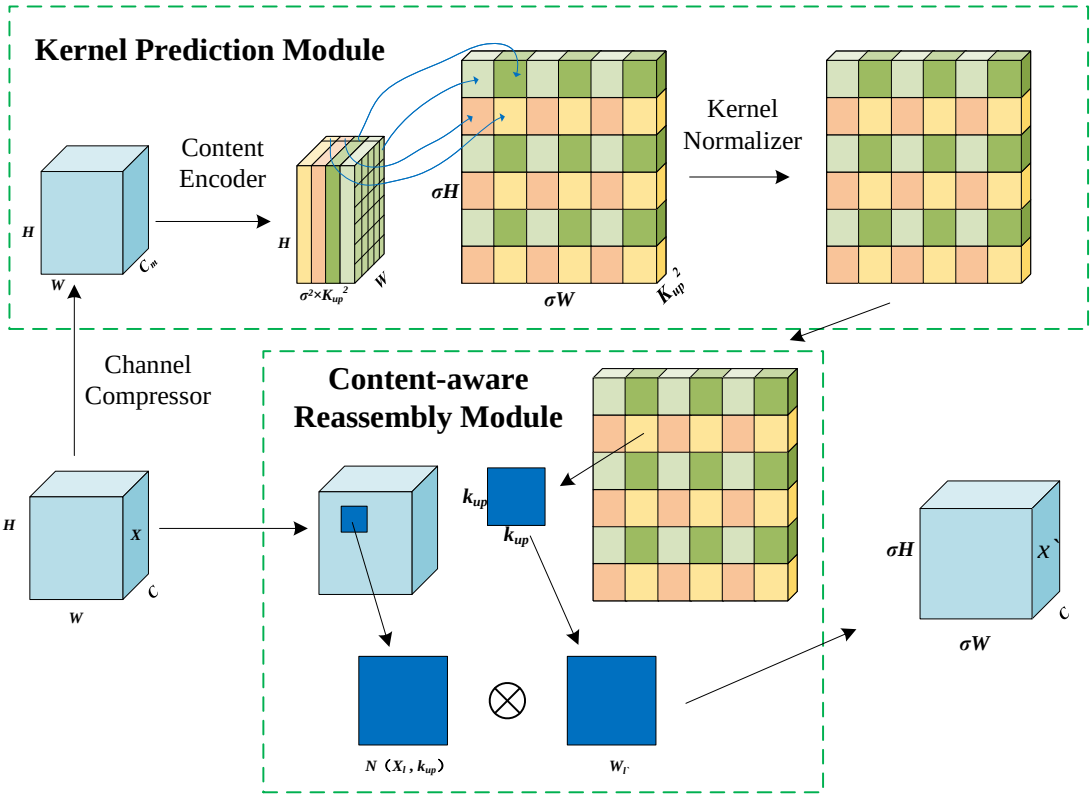


FIGURE 5. CARAFE upsampling network structure.

TABLE 2. Comparison experiments.

Models	Precision	Recall	mAP@0.5	mAP@0.5~0.95
Faster R-CNN	0.712	0.621	0.623	0.305
SSD	0.735	0.712	0.674	0.348
YOLOv5n	0.768	0.655	0.690	0.360
YOLOv7	0.889	0.748	0.809	0.514
YOLOv7-tiny	0.875	0.759	0.808	0.512
YOLOv7x	0.895	0.726	0.792	0.489
YOLOv8n	0.885	0.726	0.802	0.536
YOLOv9	0.863	0.682	0.757	0.480
RCS-YOLO	0.862	0.728	0.722	0.415
BGF-YOLO	0.854	0.721	0.807	0.401
YOLOv8-DEC	0.909	0.743	0.820	0.556

in order to guarantee the evaluation's comprehensiveness and authority. These models include RCS-YOLO [6] and BGF-YOLO [7], which are enhanced, and open-source versions of our proposed models, which are specifically intended to increase the accuracy of brain tumor detection. We have added a number of well-known YOLO family variations, such as the incredibly effective Faster R-CNN, SSD, YOLOv5n, YOLOv7, YOLOv7-tiny, YOLOv7x [24] and YOLOv9 [25], which have been proven to perform exceptionally well under a variety of performance conditions. This extensive coverage of the comparative framework was intended to guarantee a thorough and representative assessment and to offer a strong basis for the performance of the models to be compared scientifically.

Particularly crucially, all models' training processes are repeated under the same experimental conditions to guarantee the comparability and reliability of the evaluation results. Moreover, consistency and transparency are prioritized at every stage of the process, from data preprocessing to model configuration to final performance metrics. Table 2 shows the outcomes of the experiment.

It is important to note that variations in data preparation, training and validation protocols, and computational resources make direct comparisons in research challenging. Every model

TABLE 3. Ablation experiment.

EMA	CARAFE	C2f_DySnakeConv	Precision	Recall	mAP@0.5	mAP@0.5~0.95
			0.885	0.726	0.802	0.536
√			0.903	0.732	0.807	0.543
	√		0.899	0.735	0.809	0.545
		√	0.895	0.736	0.808	0.548
	√	√	0.914	0.742	0.817	0.552
√	√		0.906	0.746	0.817	0.551
√		√	0.887	0.745	0.815	0.551
√	√	√	0.909	0.743	0.820	0.556

utilized in the comparison studies was obtained from the open-source website GitHub, and every experiment's outcome was repeated in the same way.

The revised YOLOv8-DEC model has a slightly lower recall than YOLOv7-tiny, according to Table 2's testing results. In contrast, Precision, mAP@0.5, and mAP@0.5~0.95 are higher than those of the previous models, coming in at 90.9%, 82.0%, and 55.6%, respectively. These findings highlight the promise of the trustworthy brain cancer diagnostic approach that has been suggested. Comparisons between the outcomes of various experiments should be done carefully. However, the accuracy attained indicates the potential for enhancing the identification of brain tumors and emphasizes the superiority of the suggested optimized YOLOv8 model. It is clear from the study's research above that the YOLOv8-DEC algorithm performs better across a range of metrics.

3.3.2. Ablation Experiment

The baseline model YOLOv8n is used as the reference to determine the effectiveness of each enhancement module in the proposed research algorithm. Evaluation metrics such as precision, recall, mAP@0.5, and mAP@0.5~0.95 are employed. Several combinations of several enhancement modules are used in ablation tests, and the outcomes are outlined in Table 3.

This study has used a series of carefully planned trials using ablation analysis to systematically evaluate the contributions of the EMA, CARAFE, and C2f_DySnakeConv modules to the performance of the YOLOv8 brain tumor identification model. To enable thorough and in-depth examination, eight iterative experiments were conducted, each combining various combinations of these components into the YOLOv8 architecture. The experimental sequence was carefully designed: first, the EMA module was integrated alone to see its effect; next, the CARAFE and C2f_DySnakeConv modules were independently tested for efficacy; and last, all three modules were integrated for thorough optimization training while keeping the core structure of YOLOv8.

It is clear from the analysis of the experimental results in Table 3 that the accuracy, recall rate, mAP@0.5, and mAP@0.5~0.95 of the model have all improved by 2.0, 0.74, 0.62, and 1.26 percentage points, respectively, as a result of the independent application of the EMA model to improve the Backbone of the original YOLOv8n model. Comparably, accuracy, recall rate, mAP@0.5, and mAP@0.5~0.95 have all increased, by 1.58, 1.24, 0.87, and 1.68 percentage points, respectively, as a result of the CARAFE module's exclusive incorporation into the neck portion. Similarly, the accuracy, recall rate, mAP@0.5, and mAP@0.5~0.95 have all increased, by 1.13, 1.38, 0.75, and 2.24 percentage points, respectively, when the C2f-DSConv module was included. Moreover, the model outperforms both the individually improved models and the original YOLOv8n model when the modules are merged in pairs. For instance, when the EMA and CARAFE modules are combined, the accuracy, recall rate, mAP@0.5, and mAP@0.5~0.95 are increased relative to the original model by 2.37, 2.76, 1.87, and 2.8 percentage points, respectively. The EMA module and C2f-DSConv module together also lead to increases of 0.23, 2.62, 1.62, and 2.8 percentage points, in that order. Furthermore, there are increases of 3.28, 2.2, 1.87, and 2.99 percentage points for accuracy, recall rate, mAP@0.5, and mAP@0.5~0.95 when the CARAFE module and C2f-DSConv module are combined. When all three modules are integrated into the improved YOLOv8n original model, the accuracy of the comprehensive model is 0.005 percentage points less than when the CARAFE module and C2f-DSConv module are combined. All things considered, though, the comprehensive model outperforms the original YOLOv8n model, separately improved models, and pairwise merged models in terms of accuracy, recall rate, mAP@0.5, and mAP@0.5~0.95. The complete model has enhanced accuracy, recall rate, mAP@0.5, mAP@0.5~0.95, and mAP@0.5 by 2.71, 2.34, 2.24, and 3.73 percentage points, respectively, compared with the previous model.

4. DISCUSSION

In this work, we present an improved version of the YOLOv8 model for precise target identification of various brain tumor types. The C2f_DySnakeConv module, EMA attention mechanism, and CARAFE sampling module are all carefully incorporated into our suggested model through optimization.

The higher predictive performance of our suggested architecture is highlighted by comparison evaluations with current state-of-the-art models (see Table 2 for details). The same dataset and configuration environment were used for these comparisons, and the experimental results for the ablation experiments that examine the function of the

EMA attention mechanism, CARAFE sampling module, and C2f_DySnakeConv module are displayed in Table 3 and provide insight into the individual and combined contributions of these improved enhancements to the efficacy of the Yolov8-based brain tumor detection model. Understanding the role of each component through a methodical approach to ablation guides the improvement and efficiency optimization of this state-of-the-art procedure for the detection of brain cancers.

Our proposed enhanced model achieves significant outperformance in several key performance metrics, such as precision, recall, and average precision improvement in the range of mAP@0.5 and mAP@0.5~0.95 of 2.71%, 2.34%, 2.24%, and 3.73%, respectively, compared with the existing publicly available state-of-the-art models and the original YOLOv8. This accomplishment not only validates the efficacy of the suggested enhancement, but also represents a technological advance in the area of automated brain tumor identification.

The results of ablation experiments affirm the higher tumor detection accuracy achieved by C2f_DySnakeConv, validating its superior edge-awareness and detail-resolution capabilities. As a result, the utility of C2f_DySnakeConv extends to general and small target detection, including tumors.

In the clinical context, early and precise brain tumor diagnosis is essential for creating treatment regimens that work, increasing patient survival, and enhancing quality of life. The study's findings show that the improved YOLOv8 model can be a useful tool for clinicians, particularly in situations where resources are scarce, or the workforce is underqualified. It can also efficiently lighten doctors' workloads and expedite the diagnostic process, which speeds up the course of treatment.

Our study still has a lot of obstacles to overcome despite the amazing outcomes, one of which is how to better optimize the model for uncommon or unique forms of tumor patients. Furthermore, even with the increased precision of the model, the judgment of the model still requires professional verification in cases that are extremely complicated or ambiguous in order to guarantee the accuracy of the diagnosis.

Deep learning architectures have a strong dependence on training data, limiting recognition to similar input patterns. With this in mind, different MRI modalities have their own advantages and disadvantages for tumor detection. If an integrated multimodal deep learning framework can simultaneously analyze and fuse data from different MRI sequences, it is expected to optimize detection accuracy and robustness through synergy. However, the key obstacle is still the lack of rich multisequence MRI datasets, which hinders research progress and model refinement.

Future directions for research include investigating cross-modality fusion, which combines this model with data from other medical imaging modalities (such as PET or CT) to benefit from the complementary qualities of multimodal information and provide a more thorough and precise tumor characterization. Our focus on 2D MRI reflects its practical superiority in computational efficiency and accessibility, making it ideal for resource-constrained settings. Despite 3D MRI's higher fidelity, 2D MRI's pragmatic benefits underpin our YOLOv8 framework development, ensuring broad applicability. Future work will explore integrating 3D MRI for enhanced precision, while preserving our commitment to accessible diagnostics. By creating an interpretation mechanism that is more visible, potential biases or inaccuracies can be identified and corrected while also helping patients and doctors to better understand the model's decision-making process and maintain trust.

5. CONCLUSION

In this study, a novel YOLOv8-DEC model based on YOLOv8 has been developed for accurate detection of brain tumors from MRl. It was shown that by optimizing the YOLOV8 structure, introducing CARAFE on the Neck structure, EMA interspersed in the Backbone structure, and DSConv fused in the Backbone. The contribution of each module to the model was also verified under ablation experiments. These improvements enhance the target detection capability of the model corresponding to the brain tumor detection task. In addition, it is also superior to other excellent models under comparison experiments with other models.

ACKNOWLEDGEMENT

This work was supported by the Science and Technology Program of Fujian Province of China [grant number 2022J011271, 2023I0044].

REFERENCES

[1] Lee, D. Y., "Roles of mTOR signaling in brain development," *Experimental Neurobiology*, Vol. 24, No. 3, 177, 2015.

[2] Arabahmadi, M., R. Farahbakhsh, and J. Rezazadeh, "Deep learning for smart healthcare — A survey on brain tumor detection from medical imaging," *Sensors*, Vol. 22, No. 5, 1960, 2022.

[3] Tandel, G. S., A. Tiwari, and O. G. Kakde, "Performance optimisation of deep learning models using majority voting algorithm for brain tumour classification," *Computers in Biology and Medicine*, Vol. 135, 104564, 2021.

[4] Tiwari, A., S. Srivastava, and M. Pant, "Brain tumor segmentation and classification from magnetic resonance images: Review of selected methods from 2014 to 2019," *Pattern Recognition Letters*, Vol. 131, 244–260, 2020.

[5] Raut, G., A. Raut, J. Bhagade, J. Bhagade, and S. Gavhane, "Deep learning approach for brain tumor detection and segmentation," in *2020 International Conference on Convergence to Digital World — Quo Vadis (ICCDW)*, 1–5, Mumbai, India, Feb. 2020.

[6] Kang, M., C.-M. Ting, F. F. Ting, and R. C.-W. Phan, "RCS-YOLO: A fast and high-accuracy object detector for brain tumor detection," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 600–610, 2023.

[7] Kang, M., C.-M. Ting, F. F. Ting, and R. C.-W. Phan, "Bgf-yolo: Enhanced yolov8 with multiscale attentional feature fusion for brain tumor detection," *ArXiv Preprint ArXiv:2309.12585*, 2023.

[8] Ismael, S. A. A., A. Mohammed, and H. Hefny, "An enhanced deep learning approach for brain cancer MRI images classification using residual networks," *Artificial Intelligence in Medicine*, Vol. 102, 101779, 2020.

[9] Ronneberger, O., P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Im-*

age Computing and Computer-Assisted Intervention — MICCAI 2015, 234–241, 2015.

[10] Amin, J., M. Sharif, M. A. Anjum, M. Raza, and S. A. C. Bukhari, "Convolutional neural network with batch normalization for glioma and stroke lesion detection using MRI," *Cognitive Systems Research*, Vol. 59, 304–311, 2020.

[11] Çinar, A. and M. Yildirim, "Detection of tumors on brain MRI images using the hybrid convolutional neural network architecture," *Medical Hypotheses*, Vol. 139, 109684, 2020.

[12] Khan, M. A., I. Ashraf, M. Alhaisoni, R. Damaševičius, R. Scherer, A. Rehman, and S. A. C. Bukhari, "Multimodal brain tumor classification using deep learning and robust feature selection: A machine learning application for radiologists," *Diagnostics*, Vol. 10, No. 8, 565, 2020.

[13] Yang, A., X. Yang, W. Wu, H. Liu, and Y. Zhuansun, "Research on feature extraction of tumor image based on convolutional neural network," *IEEE Access*, Vol. 7, 24 204–24 213, 2019.

[14] Ke, Q., J. Zhang, W. Wei, R. Damaševičius, and M. Woźniak, "Adaptive independent subspace analysis of brain magnetic resonance imaging data," *IEEE Acces*, Vol. 7, 12 252–12 261, 2019.

[15] Thaha, M. M., K. P. M. Kumar, B. S. Murugan, S. Dhanasekeran, P. Vijayakarthick, and A. S. Selvi, "Brain tumor segmentation using convolutional neural networks in MRI images," *Journal of Medical Systems*, Vol. 43, 1–10, 2019.

[16] Li, M., L. Kuang, S. Xu, and Z. Sha, "Brain tumor detection based on multimodal information fusion and convolutional neural network," *IEEE Access*, Vol. 7, 180 134–180 146, 2019.

[17] Yang, Y., L.-F. Yan, X. Zhang, Y. Han, H.-Y. Nan, Y.-C. Hu, B. Hu, S.-L. Yan, J. Zhang, D.-L. Cheng, *et al.*, "Glioma grading on conventional MR images: A deep learning study with transfer learning," *Frontiers in Neuroscience*, Vol. 12, 804, 2018.

[18] Selvapandian, A. and K. Manivannan, "Fusion based glioma brain tumor detection and segmentation using ANFIS classification," *Computer Methods and Programs in Biomedicine*, Vol. 166, 33–38, 2018.

[19] Anaraki, A. K., M. Ayati, and F. Kazemi, "Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms," *Biocybernetics and Biomedical Engineering*, Vol. 39, No. 1, 63–74, 2019.

[20] Ghanem, Y., "Brain tumor detection dataset," [Online] Available: https://universe.roboflow.com/yousef-ghanem-jzj4y/brain-tumor-detection-fpf1f, Jul. 2022.

[21] Qi, Y., Y. He, X. Qi, Y. Zhang, and G. Yang, "Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 6070–6079, 2023.

[22] Ouyang, D., S. He, G. Zhang, M. Luo, H. Guo, J. Zhan, and Z. Huang, "Efficient multi-scale attention module with cross-spatial learning," in *ICASSP 2023 — 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5, Rhodes Island, Greece, Jun. 2023.

[23] Wang, J., K. Chen, R. Xu, Z. Liu, C. C. Loy, and D. Lin, "CARAFE: Content-aware reassembly of features," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 3007–3016, Seoul, Korea (South), Oct. 2019.

[24] Wang, C.-Y., A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 7464–7475, Vancouver, BC, Canada, Jun. 2023.

[25] Wang, C.-Y., I.-H. Yeh, and H.-Y. M. Liao, "YOLOv9: Learning what you want to learn using programmable gradient information," *ArXiv Preprint ArXiv:2402.13616*, 2024.