

HG-YOLO: Improving Tumor Detection with PP-HGNet and Global Attention Mechanism

Kien Trang[1,2,3], Bao Quoc Vuong[2,3(✉)], An Hoang Nguyen[1,2,3], Fung Fung Ting[1], and Chee-Ming Ting[1]

[1] School of Information Technology, Monash University, Malaysia Campus, 47500 Subang Jaya, Malaysia
[2] School of Electrical Engineering, International University, Ho Chi Minh City 700000, Vietnam
vqbao@hcmiu.edu.vn
[3] Vietnam National University, Ho Chi Minh City 700000, Vietnam

Abstract. Cancer continues to be a major global health issue, which is defined by uncontrolled cell proliferation and the potential invasion to other areas of the body. The early and precise detection of tumors through medical imaging is essential in improving cancer prognosis and treatment outcomes. This study presents HG-YOLO, a novel YOLO-based architecture enhanced by the integration of the PP-HGNet backbone and a Global Attention Mechanism (GAM). HG-YOLO capitalizes on the robust feature extraction capabilities of PP-HGNet and the attention-enhancing properties of GAM to improve the potential features. This combination aims to improve the sensitivity and precision of tumor localization, especially in complex cases where tumors are small or poorly delineated. Overall, the model is assessed using the Brain Tumor Detection 2020 (Br35H) dataset with the Magnetic Resonance Imaging (MRI) images. Comparative studies show that our HG-YOLO outperforms the prior versions in terms of Precision, Recall, mAP50 and mAP50-95 - giving 0.934, 0.915, 0.953, and 0.728, respectively.

Keywords: YOLO · brain tumor · MRI · deep learning · PP-HGNet

1 Introduction

The global healthcare system still faces significant challenges, leading to delays in cancer screenings, diagnoses, and treatments in some low-income nations. The International Agency for Research on Cancer reports that there were approximately 19.3 million new cases of cancer and nearly 10.0 million cancer-related deaths in 2020 alone [1]. The human brain is a biological marvel that coordinates innumerable physiological and cognitive functions in the human body. It is also vulnerable to certain diseases, including tumors. In fact, brain tumors can be either primary or metastatic, which are serious concerns due to their potential to disrupt neurological functions and threaten life [2]. Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) remain

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
O. Gervasi et al. (Eds.): ICCSA 2025, LNCS 15648, pp. 332–344, 2025.
https://doi.org/10.1007/978-3-031-97000-9_21

the primary non-invasive imaging modalities for diagnosing brain tumors, while MRI is particularly vital for defining tumor boundaries and assessing invasion into adjacent structures [3].

One of the main challenges is that the size of lesions or tumors is still small in the early stages of many diseases, making them hard to discover and diagnose promptly. The ability to accurately detect such lesions would directly benefit the prognosis of patients and greatly raise their percentage of successful treatment with proper medical intervention. Deep Learning (DL) has demonstrated significant capability in the analysis of complicated medical datasets. Convolutional neural network (CNN) is considered a particular category of deep learning models, which have been widely utilized in medical imaging techniques such as X-rays, CT scans, and MRIs.

On the other hand, You Only Look Once (YOLO) model is proposed to overcome the end-to-end training compared with previous models. In general, YOLO-based models consist of three parts: backbone, neck, and head. Thanks to the single network architecture, these models can achieve high accuracy in real-time and is improved over the series of YOLO [4–6]. Several enhanced models have been developed to work with different types of medical diseases, such as tumors [7], lesions [8], and anatomical structures [9]. Besides, Real-time Detection Transformer (RT-DETR) [10] has recently been introduced as a competitor in the field of object detection. Integrating the use of a high-end feature extraction backbone and transformer-based architecture, RT-DETR shows robust performance in many tasks. Inspired by the previous studies, the contribution of this study is mentioned as follows:

- We propose using the feature extraction backbone in RT-DETR, which integrates PP-HGNet backbone into the YOLO-based model. This enhances the ability of the feature extraction process, which benefits dealing with the more complex scenarios of images.
- We improve the neck part by using Repeated Normalized Cross Stage Partial with Efficient Large Kernel Attention Network (RepNCSPELAN) and Global Attention Mechanism (GAM) to focus more on potential features in different scales. This also enhances the ability to detect small and low-contrast objects.
- We evaluate the proposed HG-YOLO on the Brain Tumor Detection 2020 (Br35H) dataset, which demonstrates superior performance in almost all metrics compared to the previous models.

2 Related Works

Conventional approaches to medical disease classification and lesion detection involve manual annotations of the region of interest, which are time-consuming and error-prone [11]. Recently, the development of Artificial Intelligence-based methods has contributed greatly to many healthcare applications, especially in the medical image analysis fields [12]. Several deep learning frameworks in recent years have been proposed with extreme levels of accuracy in different medical object detection tasks with real-time processing, such as CNN and RCNN-based models, U-Net series, Vision Transformers, etc. Among them, brain tumor is one of the most difficult targets for the latest medical object detection approaches using various DL architectures and algorithms.

In [13], the proposed NeuroNet19 model achieved highly competitive results of brain tumor classification with 99.3% accuracy and 99.2% in precision and recall on a combined dataset from 3 separate data sources Figshare, SARTAJ, and BR35H. The model utilized the Inverted Pyramid Pooling Module (iPPM) for enhancing the feature extraction capability by capturing multi-scale feature maps from an adjusted VGG19 architecture. Another study also implemented the combined dataset is Islam et al. [14], in which they achieved significant brain tumor classification results using transfer learning of multiple modified architectures, such as DenseNet121, VGG19, InceptionV3, and MobileNet. However, even though the above studies obtained great results in the combined MRI dataset, they are subjective to broaden their target imaging techniques such as CT-scans to strengthen their model's practical implementation ability.

In comparison to CNN-based methods, the approaches using YOLO-based models also obtain notable outcomes in brain tumor detection research. In [15], the authors validate the capability of YOLOv5 and YOLOv7 models in the same combined dataset and achieved more than 90% detection results in all Precision, Recall, F1-score, mAP50, and mAP50-95, respectively. These results demonstrate high potential in applying YOLO-based models in brain tumor detection problems. Mridul et al. [16] integrates the Segment Anything Model with the YOLOv9-based architecture named SIYO, composing of many innovative modules such as PGI and GELAN. The SIYO model surpasses the previous versions and obtained 94.7% in mAP50. Later, Ming Kang et al. [17] designed an enhanced model named BGF-YOLO by combining multiple modules: the Bi-level Routing Attention, the Generalized Feature Pyramid Networks and Fourth heading Head upon a YOLOv8-based architecture. The designed BGF-YOLO achieved state-of-the-art results in the brain tumor Br35H dataset of 91.9% in precision and 92.6% in recall, while obtaining a firm increase of 4.7% in mAP50 being 97.4% compared to the base YOLOv8x.

From the above studies, it is certain that YOLO-based architectures possess the potential to improve the brain tumor detection problem, especially in later versions with a refined architecture and the latest feature extraction modules. In this study, we aim to propose a deep-learning YOLOv11-based model for medical tumor detection. The results illustrate that our proposed model has further advanced the detection outcome of the previous versions of the YOLO-based model.

3 Proposed Method

3.1 Overview

In this study, a novel YOLO-based model is proposed to apply in classifying the medical disease. Similar to other YOLO versions, our model comprises three parts: backbone, neck, and head. Inspired from the Real-time Detection Transformer (RT-DETR), PP-HGNet is used to replace the original backbone module of YOLOv11, which aims to improve the feature extraction process, especially for small and unclear objects. In addition, the Repeated Normalized Cross Stage Partial with Efficient Large Kernel Attention Network (RepNCSPELAN) is implemented to deal with feature processing in the neck part, which might obtain more effective features. At the end of the neck part, we integrate the Global Attention Mechanism (GAM) before the head part, which may enhance the

attention ability of each detection scale. This is designed to improve the detection ability to focus on different scales, potentially increasing its performance while maintaining a balance in sensitivity. The architecture of our proposed model is shown in Fig. 1.

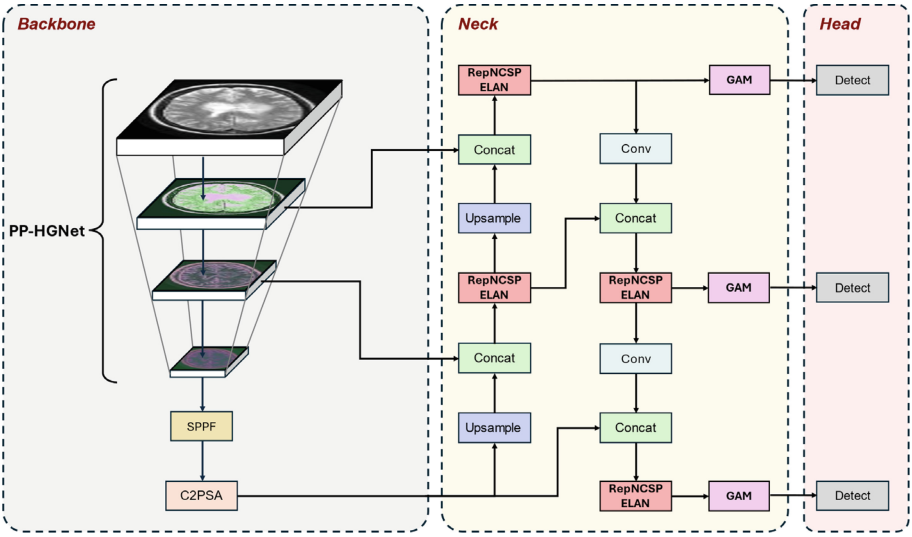

Fig. 1. Our proposed HG-YOLO model.

3.2 HG-YOLO Architecture

The main architecture of the proposed work is mainly based on YOLOv11. Therefore, some main blocks are still preserved to guarantee the advantages of the YOLO-based model. Initially, the Spatial Pyramid Pooling Fast (SPFF) module is integrated after the PP-HGNet, specifically engineered to extract more features from various regions of an image at multiple scales. This enhances the capability of the network to detect objects of varying sizes, particularly in identifying small objects. Additionally, Cross-stage Partial Spatial Attention (C2PSA) is applied to provide the attention mechanism at the end of the backbone part. This plays the role of emphasizing spatial relationships in feature maps to focus on critical areas of an image. The C2PSA block contains dual Partial Spatial Attention (PSA) modules, which operate on distinct branches. This configuration guarantees that the model prioritizes spatial information while simultaneously achieving a balance between speed and accuracy.

On the other hand, Repeated Normalized Cross Stage Partial with Efficient Large Kernel Attention Network (RepNCSPELAN) is used to replace the C3k2 block as default in the neck part. The RepNCSPELAN block is introduced in the YOLOv9, which is inspired by the architecture combination of Cross Stage Partial Network (CSPNet) [18] and Efficient Layer Aggregation Network (ELAN) [19]. This configuration ensures high performance and preservation of lightweight architecture at the same time. The illustration of RepNCSPELAN block is depicted in Fig. 2.

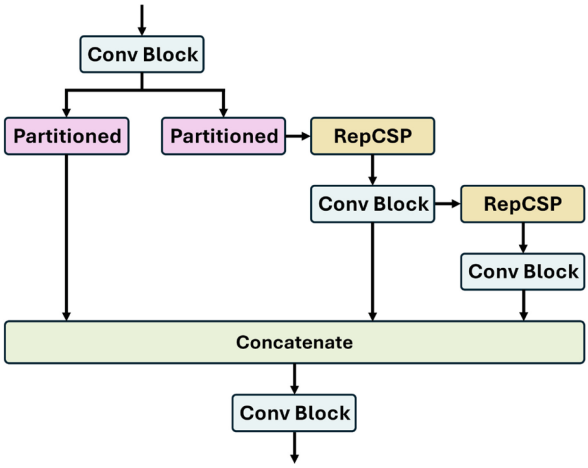

Fig. 2. The RepNCSPELAN architecture.

3.3 PP-HGNet Backbone

PP-HGNet is a high-performance backbone network introduced by the Baidu team [20], which aims to be more suitable for high-end GPU platforms. The architecture of this network employs a learnable downsampling layer (LDS Layer) from VOVNet and also amalgamates the strengths of pre-trained models such as ResNet [21], and PP-LCNet [22]. This method employs a hierarchical structure to extract features, which facilitates the effective handling of complicated image data. The illustration of PP-HGNet backbone is shown in the Fig. 3.

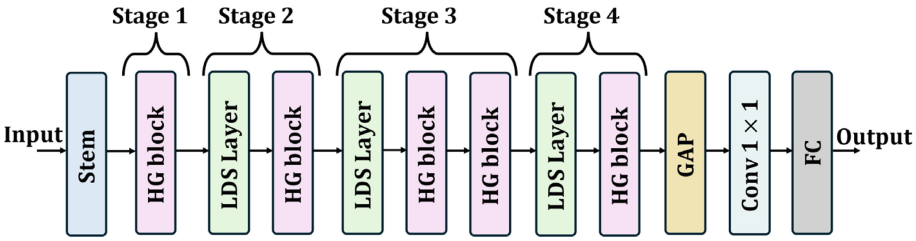

Fig. 3. The PP-HGNet architecture.

The main core of PP-HGNet is the HG Blocks positioned from the second to the fifth layers. These blocks play a crucial part in extracting the features and are meticulously designed for hierarchical data processing, which is illustrated in Fig. 4. This multilayered architecture enables the network to efficiently integrate elements at multiple levels of abstraction, aggregating both deep and general patterns the data. After the HG Block, Learnable Down-Sampling Layers are integrated, which contain a group of depth-wise convolutional and batch normalization layers to adaptively diminish the dimensions of the feature maps.

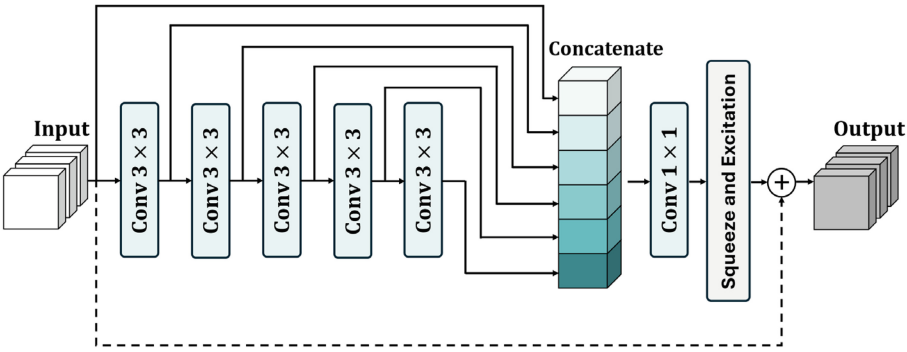

Fig. 4. The HGBlock interconnection.

Thus, in these blocks, the network takes an input and applies multiple convolutional layers, each denoted as Conv 3 × 3. These layers apply a convolution operation, which can be mathematically represented as:

$$F_l(X) = \sigma(W_l \cdot X + b_l) \tag{1}$$

where X is the input of the l-th layer, W_l and b_l are the weight and bias. The activation function is represented by σ. After the series of convolutions, features extracted at multiple scales are created. The feature maps from each stage get through the normalization process and are afterward concatenated, which is denoted as F_{concat} via (2).

$$F_{concat} = \text{CONCATENATE}(F_1, F_2, .., F_5) \tag{2}$$

To convert the feature map F_{concat} to a given size, the 1 × 1 convolution operation is applied to achieve a fixed-size feature map F'_{concat}. The concatenated feature map F_{concat} is then processed through a Squeeze-and-Excitation (SE) block, which adaptively recalibrates channel-wise feature responses for enhancing the representational power of the network. This process is formulated in (3).

$$S = \sigma\left(W_2 \cdot \delta\left(W_1 \cdot GAP\left(F'_{concat}\right) + b_1\right) + b_2\right) \tag{3}$$

where GAP represents global average pooling, W_1, b_1 and W_2, b_2 are the weights and biases of two fully connected layers, respectively. δ denotes the ReLU activation function, while σ is the sigmoid function. Finally, the optional step for adding the input features to the output is proposed to supplement the strength of extracted features.

3.4 Global Attention Mechanism

To achieve the enhancement in the context understanding, the Global Attention Mechanism (GAM) is applied to the end of each scale detection. The purpose of this implementation is to emphasize the essential features, which contribute to the final result. The organization of GAM is depicted in Fig. 5.

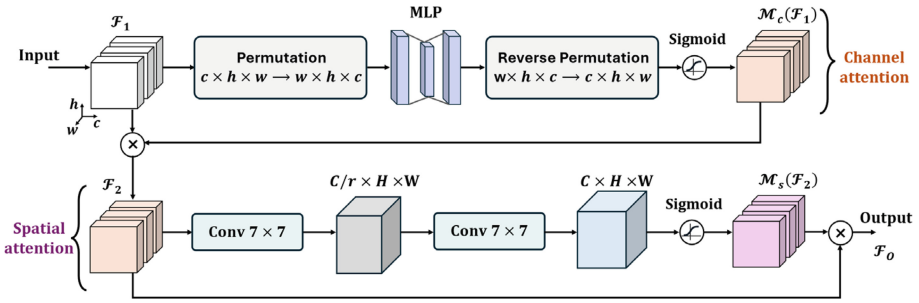

Fig. 5. The GAM interconnection.

In the channel attention part, the rearrangement of the position of input dimensions (channel, spatial, and height) from the feature map is conducted, and the feature information of each channel is vectorized for processing. The spatial features are then fed into a two-layer of Multilayer Perceptron (MLP). The training of the MLP effectively yielded the weight values for all spatial feature information across each dimension. Similar to other networks, the Sigmoid activation function is used to prevent the linearity of features. After these operations, the enhanced result $\mathcal{M}_c(\mathcal{F}_1)$ is reshaped to the same size of the initial feature map $\mathcal{F}_1$. Next, the initial feature map $\mathcal{F}_1$ is multiplied element-wise by channel attention-weighted feature map $\mathcal{M}_c(\mathcal{F}_1)$ to produce the enhanced result $\mathcal{F}_2$ for the next stage, as formulated in (4).

$$\mathcal{F}_2 = \mathcal{M}_c(\mathcal{F}_1) \otimes \mathcal{F}_1 \tag{4}$$

To continue the attention in spatial perspective, the enhanced result $\mathcal{F}_2$ from the previous stage with C channels is initially reduced to C/r by employing 7×7 convolution kernel, thereby improving the concentration of spatial feature information. The resulting features are subsequently rearranged to match the dimensions of the previous enhanced feature map $\mathcal{F}_2$ via a 7×7 convolution kernel, and these weights are compressed into weight values through the Sigmoid activation function. At the end, an element-wise multiplication between the spatial attention-weighted feature map and the feature map $\mathcal{F}_2$ is conducted to generate the final enhancement in both channel and spatial attention, as described in (5).

$$\mathcal{F}_O = \mathcal{M}_s(\mathcal{F}_2) \otimes \mathcal{F}_2 \tag{5}$$

4 Experiments and Results

4.1 Dataset

The dataset applied in this work is the Brain Tumor Detection 2020 (Br35H) dataset [23], which comprises approximately 700 MRI images in total collected from diverse patient demographics. This dataset encompasses a wide variety of tumor forms and stages, which are annotated by medical professionals to provide accurate ground-truth labels for the training of machine learning models.

4.2 Implementation Setting

In terms of conducting experiments, we implemented on the computer with Ubuntu 20.04 LTS, 12th generation of Intel core i7, and RTX 3090 containing 24Gb VRAM. The development framework is Pytorch 2.0, CUDA 11.7. The size of images in the dataset is rescaled to 640 × 640, and the training epoch and batch size are set to 300 and 32, respectively. The optimizer for training is AdamW, while the initial learning rate is 0.01.

4.3 Assessment Metrics

To evaluate the advantages and drawbacks of the proposed framework, the selected assessment metrics are mAP50, mAP50-95, precision, and recall, while the Mean Average Precision (mAP)50 and mAP50-95 are considered the main metrics to compare with other models and ablation studies.

The precision is defined as the ratio of true positive predictions to the total number of positive predictions, as shown in (6).

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6}$$

where *TP* are the correct objects, *FP* are the objects that were incorrectly identified. The Eq. (7) describes the recall metric, which measures the ability of model to detect all relevant instances in the dataset.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{7}$$

where *FN* are the objects that the model cannot detect. Additionally, mAP50 refers to the mean average precision calculated at an intersection-over-union (IoU) threshold of 50%. This means that detections are considered correct if the IOU is 50% or higher. On the other hand, mAP50-95 averages the AP calculated at multiple IOU thresholds ranging from 50% to 95%. This metric provides a measure of model performance through varying levels in the overlap criterion.

4.4 Quantitative Results

To give a comprehensive look of the learning ability, Fig. 6 shows the training progress with four evaluation metrics of our model. Overall, almost all the values reached saturation point and fluctuated less after 200 epochs. Only the mAP50-95 reaches around 0.7, while other metrics increase over 0.9. To emphasize the robustness of the proposed framework for medical disease detection, Table 1 displays the comparative results between other methods (YOLOv9 [4], YOLOv10 [5], YOLOv11 [6]) and our study. It is clear that our proposed architecture, which integrates PP-HGNet and GAM, achieves good performance. HG-YOLO achieves the highest values compared to other models. The precision of our model also reaches the best result, which matches the case of YOLOv9 at 0.934. However, in the case of mAP50-95, our model scores slightly lower but remains competitive.

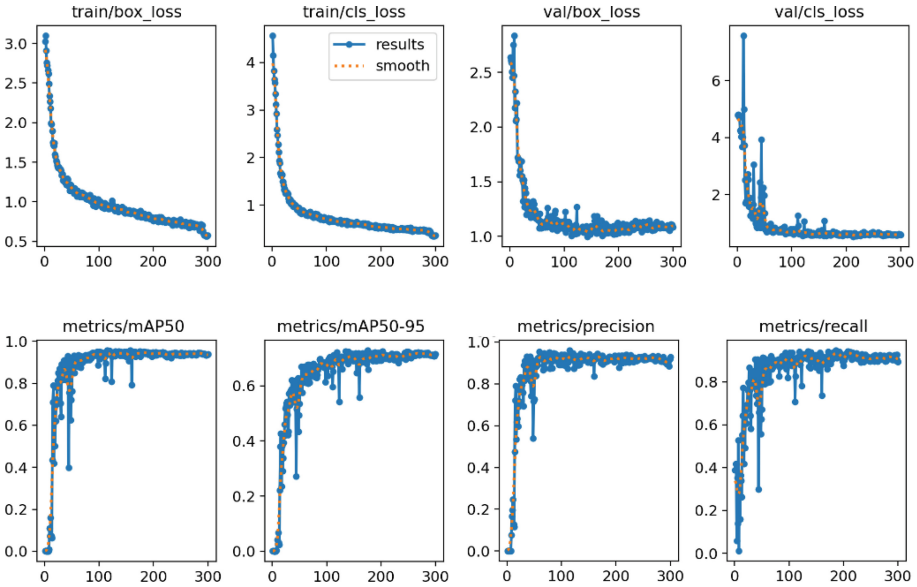

Fig. 6. The training progress of proposed model.

Table 1. Performance comparison of other models.

Model	Precision	Recall	mAP50	mAP50-95
YOLOv9	0.934	0.91	0.942	0.733
YOLOv10	0.908	0.831	0.927	0.717
YOLOv11	0.945	0.886	0.936	0.725
HG-YOLO (our)	**0.934**	**0.915**	**0.953**	**0.728**

4.5 Qualitative Results

Besides analyzing the quantitative results, we also conduct the detection of other models and our proposed framework perceptibly. The visualization of this comparison is shown in Fig. 7. In the first instance, while YOLOv10 and YOLOv11 have duplicated error detections, only YOLOv9 and our proposed model detect correctly the mass. Similarly, YOLOv10 and YOLOv11 models fail to detect the mass object in the second sample scan, while the remaining models successfully identify the tumor. In the third sample scan, YOLOv9 fails to detect the tumor entirely, whereas YOLOv10, YOLOv11, and the proposed model successfully identify the tumor. For the final cases, YOLOv9, YOLOv10, and YOLOv11 models fail to accurately detect the tumor, while our model shows advanced detection capability by accurately placing a correct bounding box. Overall, these results suggest that our proposed model achieves more reliable and precise mass localization compared to existing YOLO-based approaches.

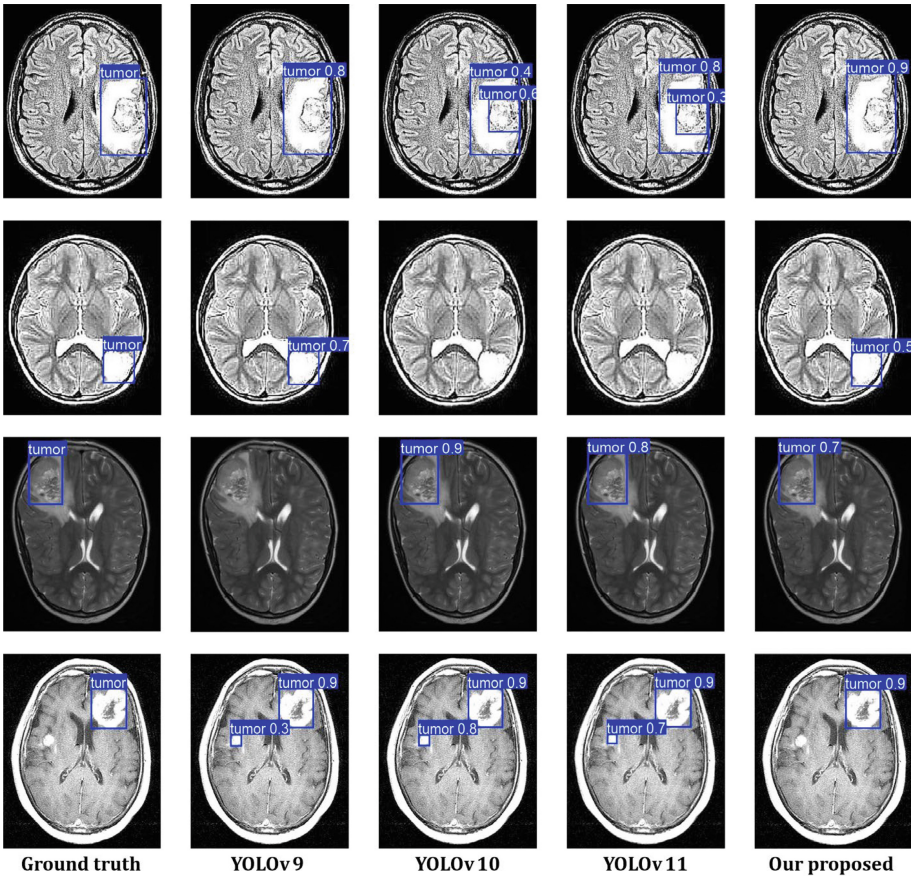

Fig. 7. The comparison of visualization between different models

4.6 Ablation Studies

In order to provide the comparative effects of implemented settings, we conduct the three sets of ablation studies.

Effect of Attention Module

The GAM module has been chosen to be applied to our proposed framework. To carry out the selection of this, Table 2 shows the performance comparison of different attention approaches such as Coordinate Attention (CA) [24] and Triplet Attention (TA) [25]. Among these methods, GAM consistently outperforms CA and TA in almost all metrics. Specifically, GAM achieves the highest Precision at 0.934, which surpasses CA at 0.922 and TA at 0.925. This indicates that GAM provides a more accurate distinction between tumors and non-tumor regions. In terms of mean Average Precision, GAM reaches an mAP50 of 0.953, slightly exceeding CA at 0.947 and TA at 0.945. For the stricter mAP50-95 metric, GAM still scores the highest value of 0.728. As a result, this emphasizes the selection of our implementation.

Table 2. Performance of different attention modules.

Method	Precision	Recall	mAP50	mAP50-95
CA	0.922	0.905	0.947	0.725
TA	0.925	0.925	0.945	0.726
GAM	**0.934**	**0.915**	**0.953**	**0.728**

Effect of Bounding Box Optimization Method

This optimization is primarily aimed to improve the bounding box regression for object detection tasks. To enhance the traditional IoU, advancements for IoU are proposed, such as CIoU [26], SIoU [27] and DIoU [28]. The comparison with different types of IoU is shown in Table 3. Overall, CIoU achieves almost the highest values among metrics compared to other methods. Although GIoU records the highest mAP50-95 of 0.732, which is higher than CIoU at 0.728 and SIoU at 0.726, CIoU still reaches the overall balance among the metrics. It achieves the highest Precision of 0.934, Recall of 0.915, and mAP50 of 0.953, compared to the other methods. This can be explained that CIoU considers not only the overlapping area but also the distance between the predicted and ground-truth box centers and the aspect ratio consistency, which leads to more precise and stable bounding box regression for optimization. Thus, CIoU is selected as the bounding box optimization method for our framework.

Table 3. Performance of different bounding box optimization methods.

Method	Precision	Recall	mAP50	mAP50-95
SIoU	0.921	0.9	0.94	0.726
GIoU	0.909	0.905	0.947	0.732
CIoU	**0.934**	**0.915**	**0.953**	**0.728**

5 Conclusion

This research presents the HG-YOLO model, which is a deep learning model designed to improve tumor detection in MRI images. By integrating the PP-HGNet backbone and Global Attention Mechanism (GAM), our model enhances feature extraction and improves the attention across different scales, leading to more accurate detections. Through experiments, HG-YOLO significantly outperformed previous models (YOLOv9, YOLOv10, and YOLOv11), demonstrating high precision, recall, and mAP scores on the Br35H dataset. Specifically, our proposed model achieved precision and recall rates of 0.934 and 0.915, respectively, and a mAP50 of 0.953, indicating superior detection performance. The combination of advanced feature extraction and focused attention on relevant features allows HG-YOLO to accurately identify even small or

low-contrast tumors. This advancement suggests that the model could contribute to more reliable and early tumor detection, which offers promising applications in clinical environments. Future work can explore refining the attention mechanisms and feature extraction strategies to enhance detection performance while maintaining a lightweight architecture for practical settings.

Acknowledgments. This research is funded by International University, VNU-HCM under grant number T2023-05-EE.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this paper.

References

1. Sung, H., et al.: Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. **71**(3), 209–249 (2021). https://doi.org/10.3322/CAAC.21660
2. Chera, B.S., Kirwan, J., Mendenhall, W.M.: Chapter 79 - Management of central nervous system metastases in breast cancer. In: Bland, K.I., Copeland, E.M. (eds.) The Breast, 4th edn., pp. 1297–1319. W.B. Saunders, Philadelphia (2009). https://doi.org/10.1016/B978-1-4160-5221-0.00079-6
3. Talukder, Md.A., et al.: An efficient deep learning model to categorize brain tumor using reconstruction and fine-tuning. Expert Syst. Appl. **230**, 120534 (2023). https://doi.org/10.48550/arXiv.2305.12844
4. Wang, C.-Y., Yeh, I.-H., Mark Liao, H.-Y.: YOLOv9: learning what you want to learn using programmable gradient information. In: Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T., Varol, G. (eds.) ECCV 2024. LNCS, vol. 15089, pp. 1–21. Springer, Cham (2025). https://doi.org/10.1007/978-3-031-72751-1_1
5. Wang, A., et al.: YOLOv10: real-time end-to-end object detection. https://arxiv.org/abs/2405.14458
6. Glenn, J.: YOLOv11. GitHub (2024). https://github.com/ultralytics/ultralytics/releases/tag/v8.3.50
7. Sobek, J., et al.: MedYOLO: a medical image object detection framework. J. Digit. Imaging. Inform. Med. **37**(6), 3208–3216 (2024). https://doi.org/10.1007/s10278-024-01138-2
8. AlSadhan, N.A., Alamri, S.A., Ben Ismail, M.M., Bchir, O.: Skin cancer recognition using unified deep convolutional neural networks. Cancers **16**(7), 1246 (2024). https://doi.org/10.3390/cancers16071246
9. Zeng, P., et al.: TUSPM-NET: a multi-task model for thyroid ultrasound standard plane recognition and detection of key anatomical structures of the thyroid. Comput. Biol. Med. **163**, 107069 (2023). https://doi.org/10.1016/j.compbiomed.2023.107069
10. Zhao, Y., et al.: DETRs beat YOLOs on real-time object detection. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 16965–16974 (2024). https://doi.org/10.1109/CVPR52733.2024.01605
11. Zhao, Y., et al.: Deep learning solution for medical image localization and orientation detection. Med. Image Anal. **81**, 102529 (2022). https://doi.org/10.1016/j.media.2022.102529
12. Alowais, S.A., et al.: Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med. Educ. **23**, 689 (2023). https://doi.org/10.1186/s12909-023-04698-z

13. Haque, R., Hassan, M.M., Bairagi, A.K., Shariful Islam, S.M.: NeuroNet19: an explainable deep neural network model for the classification of brain tumors using magnetic resonance imaging data. Sci. Rep. **14**, 1524 (2024). https://doi.org/10.1038/s41598-024-51867-1
14. Islam, M.M., et al.: Transfer learning architectures with fine-tuning for brain tumor classification using magnetic resonance imaging. Healthc. Anal. **4**, 100270 (2023). https://doi.org/10.1016/j.health.2023.100270
15. Almufareh, M.F., Imran, M., Khan, A., Humayun, M., Asim, M.: Automated brain tumor segmentation and classification in MRI using YOLO-based deep learning. IEEE Access **12**, 16189–16207 (2024). https://doi.org/10.1109/ACCESS.2024.3359418
16. Mayankeyshwar, M., Kumar, L., Wagh, M.P., Behuria, S., Yadav, D.: Brain tumor detection and segmentation using SAM integrated YOLOv9 scheme. In: 2024 IEEE 1st International Conference on Advances in Signal Processing, Power, Communication, and Computing (ASPCC), pp. 67–72 (2024). https://doi.org/10.1109/ASPCC62191.2024.10881978
17. Kang, M., Ting, C.-M., Ting, F.F., Phan, R.C.-W.: BGF-YOLO: enhanced YOLOv8 with multiscale attentional feature fusion for brain tumor detection. In: Linguraru, M.G., et al. (eds.) MICCAI 2024. LNCS, vol. 15008. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-72111-3_4
18. Wang, C.-Y., Liao, H.-Y.M., Wu, Y.-H., Chen, P.-Y., Hsieh, J.-W., Yeh, I.-H.: CSPNet: a new backbone that can enhance learning capability of CNN. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, pp. 1571–1580. IEEE (2020). https://doi.org/10.1109/CVPRW50498.2020.00203
19. Wang, C.-Y., Liao, H.-Y.M., Yeh, I.-H.: Designing network design strategies through gradient path analysis. J. Inf. Sci. Eng. (JISE) **39**(4), 975–995 (2023). https://doi.org/10.6688/JISE.202307_39(4).0016
20. Paddle, P.: PP-HGNet series - 2022. GitHub. https://github.com/PaddlePaddle/PaddleClas/blob/release/2.5.2/docs/zh_CN/models/ImageNet1k/PP-HGNet.md. Accessed 09 Jan 2025
21. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90
22. Cui, C., et al.: PP-LCNet: a lightweight CPU convolutional neural network. arXiv preprint, arXiv:2109.15099 (2021). https://doi.org/10.48550/arXiv.2109.15099
23. Br35H: Brain tumor detection 2020. Kaggle. https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection. Accessed 09 Jan 2025
24. Hou, Q., Zhou, D., Feng, J.: 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. IEEE (2021). https://doi.org/10.1109/CVPR46437.2021.01350
25. Misra, D., Nalamada, T., Arasanipalai, A.U., Hou, Q.: Rotate to attend: convolutional triplet attention module. In: 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 3138–3147. IEEE, January 2021. https://doi.org/10.1109/WACV48630.2021.00318
26. Zheng, Z., et al.: Enhancing geometric factors in model learning and inference for object detection and instance segmentation. IEEE Trans. Cybern. **52**(8), 8574–8586 (2022). https://doi.org/10.1109/TCYB.2021.3095305
27. Gevorgyan, Z.: SIoU loss: more powerful learning for bounding box regression. arXiv preprint, arXiv:2205.12740 (2022). https://doi.org/10.48550/arXiv.2205.12740
28. Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., Ren, D.: Distance-IoU loss: faster and better learning for bounding box regression. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 07, April 2020. https://doi.org/10.1609/aaai.v34i07.6999