Brain Tumor Detection

Damilola Olaiya*
Carleton University
Ottawa, Canada
damilolaolaiya@cmail.carleton.ca

Hamza Cecen*
Carleton University
Ottawa, Canada
hamzacecen@cmail.carleton.ca

WooSeok Kim*
Carleton University
Ottawa, Canada
wooseokkim@cmail.carleton.ca

*These authors contributed equally to this work.

Abstract

Magnetic Resonance Imaging (MRI) has a central role in the detection and clinical assessment of brain tumors. Recent advancements in machine learning and object detection, especially YOLO-related frameworks, have significantly improved automated tumor localization in 2D multiplanar MRI slices. One of these, Pretrained Knowledge Guided YOLO (PK-YOLO) has demonstrated state-of-the-art performance by integrating SparK-pretrained RepViT backbone and specialized optimization techniques. However, the original PK-YOLO design exhibits notable limitations, primarily stemming from a plane-shift domain gap, with its backbone is pretrained exclusively on axial images, resulting in substantial performance disparities across planes.

To solve this limitation, we propose a 3D Multiplanar Fusion framework. In this method, three different backbones, each of which is pretrained on an anatomical plane (*i.e.*, axial, coronal, or sagittal), are used. Then the outputs of the backbones are combined using an adaptive router module that learns how much weight to give each plane.

1 Introduction

Over the past few years, advances in machine learning have opened up the possibilities to integrate deep learning methods into the medical domain[1]. Throughout active research, several papers employ multiple approaches (*e.g.*, image processing, convolutional networks, attention-based networks) to improve brain tumor detection performance [19].

Spotting brain tumors matters a lot when looking at medical scans, since finding them early and correctly helps doctors figure out what's wrong, decide on treatments, and keep track of how patients are doing [8]. MRIs are common for this job because they show fine details and distinguish soft tissues well [16]. Especially useful are multi-angle MRI views, like top, front, and side cuts, which give different body viewpoints, making it easier for physicians to pin down where the tumor sits and understand its shape.

Among these methods, YOLO-based models have drawn significant attention due to their efficiency, robust detection performance, and real-time capabilities. However, applying YOLO architectures in medical imaging remains challenging because MRI scans differ substantially from natural images in terms of intensity patterns, noise characteristics, and underlying anatomical structures.

To address these challenges, M. Kang *et al.* [13] proposed a new You Only Look Once (YOLO)-based [20] brain tumor detection model by leveraging 2D multiplanar Magnetic Resonance Imaging (MRI) slices for model training and fine-tuning. The experiments show that it achieves the state-of-the-art performance, scoring the highest metrics on all three scan types. Despite its performance, we have identified gaps in the model performance demonstrated by the plane-shift domain gap.

2 Limitations of the Previous Work

3 Related Work

Research on brain tumor detection using deep learning largely intersects with three key areas: DETR framework, brain tumor detection, and multiplanar MRI analysis. The PK-YOLO paper introduces contributions in all three areas, and the prior work they reference can be grouped into the following themes.

3.1 DETR Framework

Object detection approaches based on the DEtection TRansformer (DETR) architecture have evolved significantly in recent years, especially through improved training strategies and backbone pretraining. [2] Early variants such as UP-DETR introduced unsupervised pretraining for Transformers [7], showing that self-supervised representation learning could improve downstream detection accuracy. Subsequent works such as Group DETR v2 [5] and Co-DETR [25] further advanced DETR performance by exploring encoder-decoder pretraining strategies and hybrid assignments, achieving strong results on benchmarks such as MS COCO.

Lightweight or real-time DETR models have also emerged, including:

- RT-DETR, which allows flexible inference speed by varying decoder layers without retraining [24]
- LW-DETR, which improves efficiency using pretraining strategies [4]
- Salience DETR, which introduces salience-guided supervision to accelerate and stabilize training [9]

Despite these advances, the PK-YOLO paper points out that DETR-based models have not been sufficiently explored for brain tumor detection in MRI images, and prior DETR variants struggle to match YOLO-based detectors in medical settings due to computation cost and limited domain adaptation.

3.2 Brain Tumor Detection

While YOLO models dominate natural image detection tasks, their application to medical imaging, especially brain tumors, remains relatively limited.

Prior work includes:

Table 1: Summary of DETR-Based Object Detection Frameworks

Model	Key Idea	Pretraining Strategy	Strengths	Limitations
UP-DETR	Unsupervised pretraining for DETR using self-supervised tasks	Transformer encoder pretraining (unsupervised)	Improves object query learning; boosts detection accuracy	Not domain-specific; weak performance on MRI images
Group DETR v2	Strong performance via encoder–decoder pretraining	Pretrained ViT-H encoder	Very high accuracy on COCO benchmark	Extremely high computational cost; not suitable for MRI
Plain-DETR	Simplified DETR using masked image modeling	MIM-based backbone pretraining	Backbone pretraining significantly improves representation quality	Slow convergence; limited performance on small medical objects
Co-DETR	Hybrid assignments and strong pretraining improve DETR accuracy	ViT-L + DETR augmentation pretraining	First DETR surpassing 66 AP on COCO	Too heavy for practical medical imaging scenarios
RT-DETR	Real-time DETR with flexible inference speed	Backbone pretraining (CNN or transformer)	Fastest DETR; adjustable speed without retraining	Lower accuracy than YOLO-based models on MRI
LW-DETR	Lightweight DETR for real-time detection	Encoder–decoder pretraining	Balanced accuracy and efficiency	Instability in detecting small objects (*e.g.*, small tumors)
Salience DETR	Salience-guided hierarchical refinement for faster training	Transformer pretraining + salience supervision	More efficient training; better convergence	Still less accurate than PK-YOLO on MRI-based tumor detection

- RCS-YOLO, which integrates region concentration modules to improve localization of tumors [11]
- BGF-YOLO, which enhances YOLOv8 using multiscale attentional feature fusion [12]
- Traditional YOLO variants such as YOLOv5 [10], YOLOv8, and YOLOv9 [22] have been used but still face challenges detecting small tumors and generalizing across multiplanar MRI slices

These YOLO-based approaches generally perform well in speed and accuracy, but lack mechanisms to integrate domain-specific pretrained knowledge, especially for complex medical imagery.

3.3 Multiplanar MRI Analysis

Multiplanar analysis (axial, coronal, sagittal slices) is critical in medical imaging because tumors may appear differently across orientations.

Previous works include:

- Piantadosi et al. used ensembles of 2D CNNs for MRI tissue segmentation [18]
- MPS-FFA model introduced multiplane and multiscale feature fusion for Alzheimer's disease classification [15]
- Studies evaluating object detection on multiplanar MRI slices (*e.g.*, Barbato and Menga with YOLOv5m) reported low mean Average Precision, highlighting the difficulty of multiplane data

Challenges identified in previous studies: [17]

- Significant variation in lesion size and position across planes
- Increased number of small lesions due to slice orientation
- Difficulty training a single model to perform well on all three anatomical planes

3.4 Mixture-of-Experts

Integrating various knowledge into one for machine learning is especially important as it can enable utilization of pretrained knowledge. N. Shazeer *et al*. [21] proposed a neural network architecture called Mixture-of-Experts which considers each model as an expert and uses a gating layer to combine the models into one. Its potential lies in computational efficiency, scalability, and leveraging specialized expert modules. G. Chen *et al*. [3] used router module for multimodal large language model (MLLM) called LION, treating each AdaptFormer [6] as an expert. Each AdaptFormer is allocated to perform image-level and region-level task given the input image and prompt, and LION adopts a router module to control the ratio between image-level and region-level knowledge. In terms of medical machine learning area, some works [14, 23] employ mixture-of-experts architecture to handle brain tumor detection problem.

4 Method

here

5 Experiments

here

6 Conclusion

References

[1] S. Asif, Y. Wenhui, S. ur Rehman, Q. ul ain, K. Amjad, Y. Yueyang, S. Jinhai, and M. Awais. Advancements and prospects of machine learning in medical diagnostics: unveiling the future of diagnostic precision. *Archives of Computational Methods in Engineering*, 32(2):853–883, 2025.

[2] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer, 2020.

[3] G. Chen, L. Shen, R. Shao, X. Deng, and L. Nie. Lion: Empowering multimodal large language model with dual-level visual knowledge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 26540–26550, 2024.

[4] Q. Chen, X. Su, X. Zhang, J. Wang, J. Chen, Y. Shen, C. Han, Z. Chen, W. Xu, F. Li, et al. Lw-detr: A transformer replacement to yolo for real-time detection. *arXiv preprint arXiv:2406.03459*, 2024.

[5] Q. e. a. Chen. Group detr v2: Strong object detector with encoder-decoder pretraining. *arXiv:2211.03594*, 2022.

[6] S. Chen, C. Ge, Z. Tong, J. Wang, Y. Song, J. Wang, and P. Luo. Adaptformer: Adapting vision transformers for scalable visual recognition. *Advances in Neural Information Processing Systems*, 35:16664–16678, 2022.

[7] Z. Dai, B. Cai, Y. Lin, and J. Chen. Unsupervised pre-training for detection transformers. *IEEE transactions on pattern analysis and machine intelligence*, 45(11):12772–12782, 2022.

[8] S. Das and R. S. Goswami. Advancements in brain tumor analysis: a comprehensive review of machine learning, hybrid deep learning, and transfer learning approaches for mri-based classification and segmentation. *Multimedia Tools and Applications*, 84(23):26645–26682, 2025.

[9] X. Hou, M. Liu, S. Zhang, P. Wei, and B. Chen. Salience detr: Enhancing detection transformer with hierarchical salience filtering refinement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17574–17583, 2024.

[10] G. Jocher. Yolov5 by ultralytics, may 2020. *URL https://github. com/ultralytics/yolov5*, 2023.

[11] M. Kang, 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*, pages 600–610. Springer, 2023.

[12] M. Kang, C.-M. Ting, F. F. Ting, and R. C.-W. Phan. Bgf-yolo: Enhanced yolov8 with multiscale attentional feature fusion for brain tumor detection. In *MICCAI*, 2024.

[13] M. Kang, F. F. Ting, R. C.-W. Phan, and C.-M. Ting. Pk-yolo: Pretrained knowledge guided yolo for brain tumor detection in multiplanar mri slices. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 3732–3741. IEEE, 2025.

[14] S. Li, X. Sui, X. Luo, X. Xu, Y. Liu, and R. Goh. Medical image segmentation using squeeze-and-expansion transformers. *arXiv preprint arXiv:2105.09511*, 2021.

[15] F. Liu, H. Wang, S.-N. Liang, Z. Jin, S. Wei, and X. Li. Mps-ffa: A multiplane and multiscale feature fusion attention network for alzheimer's disease prediction with structural mri. *Computers in Biology and Medicine*, 157:106790, 2023.

[16] P. Mittal et al. From black box ai to xai in neuro-oncology: a survey on mri-based tumor detection. *Discover Artificial Intelligence*, 5(1):1–21, 2025.

[17] S. Otake, T. Taoka, M. Maeda, and W. T. Yuh. A guide to identification and selection of axial planes in magnetic resonance imaging of the brain. *The neuroradiology journal*, 31(4):336–344, 2018.

[18] G. Piantadosi, M. Sansone, R. Fusco, and C. Sansone. Multi-planar 3d breast segmentation in mri via deep convolutional neural networks. *Artificial Intelligence in Medicine*, 103:101781, 2020.

[19] N. Rasool and J. I. Bhat. Brain tumour detection using machine and deep learning: a systematic review. *Multimedia tools and applications*, 84(13):11551–11604, 2025.

[20] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection, 2016.

[21] N. Shazeer, A. Mirhoseini, K. Maziarz, A. Davis, Q. Le, G. Hinton, and J. Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *arXiv preprint arXiv:1701.06538*, 2017.

[22] C.-Y. Wang, I.-H. Yeh, and H.-Y. Mark Liao. Yolov9: Learning what you want to learn using programmable gradient information. In *European conference on computer vision*, pages 1–21. Springer, 2024.

[23] R. K. Yadav, M. Kumar, and A. Nandi. A scalable brain tumor diagnosis from large-scale mri datasets using cnn-vit and expert-attention fusions. *International Journal of Computational Intelligence Systems*, 18(1):1–25, 2025.

[24] Y. e. a. Zhao. Detrs beat yolos on real-time object detection. In *CVPR*, 2024.

[25] Z. Zong, G. Song, and Y. Liu. Detrs with collaborative hybrid assignments training. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6748–6758, 2023.