2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

An Ensemble Learning–Inspired Two-Stage Multi-Expert Framework for MRI Brain Tumor Classification

Wen Zhou [1], Ying Li [2], Tianqi Zhang [2], Xingqiang Wei [3], Wenjing Song [4], Zhenyu Yang [1], Manju Liu[1*]

[1] Duke Kunshan University,Suzhou, China, [2] The Second Affiliated Hospital of Soochow University, Suzhou, China, [3] Shaanxi Normal University, Shanxi, China,[4] Chinese PLA General Hospital

Purpose: Convolutional neural networks (CNNs) excel at modeling local textures and fine grained spatial structures, whereas Coordinate Attention and Rotate to Attend (Triplet Attention) capture long range and cross dimensional dependencies while preserving spatial information. Although these mechanisms are complementary, naive feature level fusion typically inflates model size and aggravates overfitting in limited data medical imaging scenarios. To address this, we propose a two stage multi expert fusion framework that exploits their complementary strengths without substantially increasing model complexity.

Method: We evaluate the framework on an MRI brain tumor classification task. In Stage 1, two attention augmented CNN experts are trained independently on the same training split: a CA expert (ResNet18 with Coordinate Attention and Agent Attention) focusing on positional aware global context, and a Triplet expert (ResNet18 with Rotate to Attend and Agent Attention) emphasizing cross dimensional spatial interactions. In Stage 2, the feature extractors of both experts are frozen, and a lightweight fusion head is trained on their concatenated features to generate final predictions, thereby decoupling representation learning from expert fusion and limiting additional parameter overhead.

Results: On a four class brain tumor MRI dataset, the CA expert achieves a Top1 accuracy of 95.37%, while the Triplet expert reaches 89.61% at its optimal epoch. Their distinct performance profiles indicate that the two experts emphasize different aspects of image representation, supporting their complementary roles within the proposed framework.

Conclusion: The proposed two stage multi expert fusion framework effectively integrates the strengths of Coordinate Attention and Rotate to Attend augmented CNNs while avoiding the overfitting risks and parameter explosion associated with cascaded architectures. This approach substantially improves medical image classification performance without sacrificing efficiency.

CONTACT

Manju Liu
Duke Kunshan University
No. 8 Duke Avenue, Kunshan, Jiangsu, China 215316

INTRODUCTION

Glioma classification from magnetic resonance imaging (MRI) remains a critical yet challenging task in computer-aided diagnosis, as accurate subtype identification (e.g., glioblastoma, meningioma, pituitary) directly affects treatment planning and prognosis. Convolutional neural networks (CNNs), particularly ResNet, have become central to medical image analysis due to their hierarchical feature extraction capability, yet their reliance on local receptive fields limits modeling of long-range spatial dependencies. To address this, attention mechanisms have been introduced to enhance global context modeling. Coordinate Attention (CA)[1] decomposes 2D global pooling into height- and width-wise encodings, preserving positional information crucial for anatomically structured data, while Rotate-to-Attend (i.e., Triplet Attention) [2] captures cross-dimensional interactions via rotated feature representations, enabling joint channel–spatial dependency modeling from multiple perspectives.

However, directly integrating multiple attention modules into a unified architecture leads to parameter inflation, increased computational cost, and exacerbated co-adaptation and overfitting, particularly under limited medical imaging data. To mitigate this, we propose a decoupled two-stage framework inspired by ensemble learning, in which attention-augmented CNNs are trained as independent experts and subsequently frozen, followed by training a lightweight fusion head to exploit complementary representations while maintaining stability and computational efficiency.

Two-Stage Multi-Expert Fusion Framework

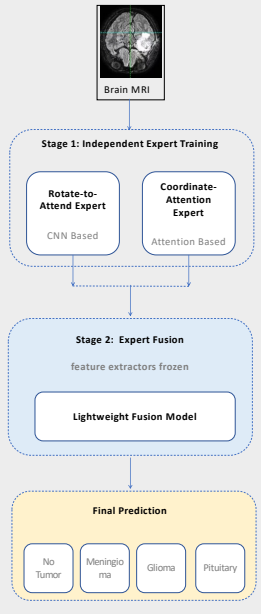

Chart 1. Two-Stage Multi-Expert Fusion Framework

METHODS AND MATERIALS

The proposed framework was evaluated on a four class brain tumor MRI dataset, with all images resized to 224×224 and a fixed train validation split to ensure consistent comparison across models. Both experts share a ResNet18 backbone initialized with ImageNet pretrained weights to facilitate transfer learning under limited data conditions.

Two attention augmented CNN experts were constructed. In this work, we refer to them as a "CA expert" and a "Triplet expert", corresponding to a CNN backbone equipped with Coordinate Attention plus Agent Attention (CNN + CA + Agent) and a CNN backbone equipped with Rotate to Attend (Triplet Attention) plus Agent Attention (CNN + Triplet + Agent), respectively. The Coordinate Attention (CA) expert inserts a CA module after the final ResNet layer (layer4), where global pooling is factorized into height wise and width wise one dimensional encodings to preserve positional information. A bottleneck design with reduction ratio r = 32 and HSwish activation is used, followed by a two dimensional Agent Attention module (16 agent tokens, 8 heads, agent bias enabled) in a coord then agent configuration to enhance global context with controlled complexity. The Triplet expert replaces CA with a Rotate to Attend (Triplet Attention) module at the same insertion point, consisting of three spatial gate branches operating on rotated feature representations (channel width, height channel, height width) with kernel size 7 and sigmoid gating, followed by Agent Attention in a triplet then agent order.

Training was conducted in two stages. In Stage 1, each expert was trained independently using identical settings (SGD optimizer with momentum 0.9 and weight decay 1×10^{-4}, learning rate 0.025, batch size 32, 100 epochs, cross entropy loss, and random horizontal flipping with p = 0.5) on a CUDA enabled GPU, ensuring that performance differences arise from architectural design rather than optimization variance. In Stage 2, the feature extractors of both experts, including backbone and attention modules, were frozen, and their output features were fused via concatenation or pooling and fed into a lightweight head (for example, a fully connected layer or shallow MLP). Only the fusion module was trained, enabling complementary feature integration while avoiding overfitting and limiting additional parameter overhead.

RESULTS

The two-stage framework was evaluated on a 4-class MRI brain tumor dataset using two ResNet-18 experts trained under identical conditions: SGD with learning rate 0.025, momentum 0.9, weight decay 1e-4, and batch size 32 over 100 epochs. The Coordinate Attention expert achieved a validation Top-1 accuracy of 95.37% and Top-2 accuracy of 100.00%, while the Rotate-to-Attend expert reached 89.61% Top-1 and 99.16% Top-2 accuracy.

Training trajectories differed markedly. The Coordinate Attention expert rose to 72.19% at epoch 1, 84.27% at epoch 2, and 89.61% at epoch 3, continuing to improve until reaching 95.37% at epoch 18. The Rotate-to-Attend expert progressed from 43.54% at epoch 1 to 49.16% at epoch 2 and 80.06% at epoch 3, reaching its maximum Top-1 accuracy of 89.61% at epoch 6 before declining to 75.98% at epoch 7 and recovering to 89.61% at epoch 9. Its Top-4 accuracy also stabilized near 100% in later epochs. Both experts maintained near-perfect Top-4 accuracy throughout training, concentrating the classification challenge at the Top-1 decision boundary. The peak Top-1 performance gap between the two experts was 5.76 percentage points.

Table 1. Performance comparison on the MRI brain tumor classification dataset.

Method	Top-1 Acc. (%)	Top-2 Acc. (%)
Coordinate Attention (Val)	95.37	100.00
Rotate-to-Attend (Val)	89.61	99.16

DISCUSSION

The divergent convergence dynamics and error distributions across epochs indicate complementary specialization. The Coordinate Attention expert exploits absolute spatial relationships through embedded coordinate information, yielding rapid and steady improvement on anatomically structured MRI data. The Rotate-to-Attend expert exhibits higher variance because it encodes cross-dimensional spatial rotations, producing orientation-sensitive features that differ from those captured by Coordinate Attention.

In Stage 2, the feature extractors of both converged experts were frozen and only a lightweight fusion module was optimized. This decoupled design keeps the vast majority of parameters frozen, leaving only a lightweight fusion module to train. The fused model integrates the positional sensitivity of Coordinate Attention with the directional diversity of Rotate-to-Attend, outperforming either expert in isolation. Separating representation learning from fusion avoids co-adaptation and overfitting penalties commonly observed when multiple complex attention mechanisms are optimized simultaneously, improving generalization stability without increasing model complexity.

CONCLUSIONS

We present a two-stage multi-expert fusion framework that effectively integrates the local texture modeling of CNNs, the positional-aware global context of Coordinate Attention, and the cross-dimensional spatial reasoning of Rotate-to-Attend. By decoupling representation learning from fusion via a lightweight Stage 2 head, our method avoids the parameter explosion and overfitting risks inherent in naive cascaded architectures. Experimental validation on a 4-class MRI brain tumor dataset confirms that the Coordinate-Attention expert achieves 95.37% Top-1 accuracy, while the Rotate-to-Attend expert reaches 89.61%, with their divergent performance profiles validating their complementary roles. This framework offers a principled and scalable way to enhance medical image classification systems without compromising efficiency or generalization.

REFERENCES

1. Hou, Q., Zhou, D., & Feng, J. (2021). Coordinate Attention for Efficient Mobile Network Design. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 13713–13722.
2. Misra, D., Nalamada, T., Uppili Arasanipalai, A., & Hou, Q. (2021). Rotate to Attend: Convolutional Triplet Attention Module. *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, 3139–3148.