Problem Statement & Solution: Cross-Architectural Knowledge Distillation in Medical Imaging: Multi-Scale Geometric Feature Fusion for MRI Scan Classification

Mahiyat Nawar Mantaqa
2122455042

MD. Shafat Islam Khan
2121517642

***Abstract*—Brain tumor segmentation is a challenging task due to the intricate and variable nature of MRI scans. This report presents a novel approach using a hybrid teacher model that combines a convolutional neural network (VGG19) and a Vision Transformer (ViT) to effectively capture both local and global features. We then employ cross-architectural knowledge distillation to transfer the rich representations learned by the teacher to a more efficient student model. Additionally, a multi-scale geometric feature fusion technique is incorporated to enhance the detection and segmentation of complex tumor regions. The following sections describe the methods and rationale behind each solution in detail.**

I. Introduction

The BraTS23 dataset includes brain MRI scans that are characterized by complex anatomical structures and subtle variations in tumor regions. Accurate segmentation is essential for diagnosis and treatment planning, but challenges such as overlapping intensities and irregular boundaries make this task difficult. High-performing deep learning architectures, like transformers, excel in modeling long-range dependencies but come with high computational costs. Our approach addresses these challenges by integrating the strengths of CNNs and transformers in a hybrid teacher model and then distilling its knowledge into a smaller, more resource-efficient student model. This process is further enhanced by using multi-scale feature fusion to capture both detailed textures and broader spatial context.

II. Problem Statement & Solutions

A. Problem 1: High Computational Expense of Transformer Models

Challenge: Transformer models, while effective in capturing long-range dependencies, require significant computational resources due to their complex self-attention mechanisms. This makes them less practical for real-time medical applications in settings with limited resources. **Solution**: To reduce computational overhead, we use a knowledge distillation strategy. The heavy, high-performing hybrid teacher model is first trained on the BraTS23 dataset. Its learned representations, which are rich in both local details (from VGG19) and global context (from ViT), serve as guidance for training a much smaller student model. Instead of relying solely on hard labels (i.e., the true class labels), the student model is also trained to mimic the teacher's output distributions. This "soft-label" training helps the student model to learn finer distinctions in the data, allowing it to maintain high accuracy while operating with significantly reduced computational demands.

B. Problem 2: Complexity of Medical Images

Challange: Brain MRI scans are challenging due to their intricate details and subtle differences between healthy tissue and tumor regions. The images often contain overlapping structures and low contrast areas, making accurate segmentation a difficult task. **Solution**: Our approach uses multi-scale geometric feature fusion to address this complexity. The hybrid teacher model extracts features at various scales using both VGG19 and ViT. The CNN component (VGG19) captures fine, local details, while the transformer component (ViT) is better suited for understanding the overall structure and global context of the image. By fusing these features from multiple scales, the model can integrate information about small-scale textures and large-scale spatial relationships. This comprehensive feature representation enables more precise localization and segmentation of tumor regions, even when the images contain subtle and overlapping details.

C. Problem 3: Differentiating Between Tumor Sub-Regions

Challange: For effective diagnosis and treatment planning, it is crucial to accurately differentiate between various tumor sub-regions, such as the enhancing tumor, non-enhancing tumor, and edema. However, overlapping intensities and ambiguous boundaries make this differentiation particularly challenging. Solution: The hybrid teacher model is designed to extract complementary features that address this challenge. The local features derived from the VGG19 branch provide detailed information about the texture and boundaries of the tumor, while the global features obtained from the ViT branch offer insights into the spatial relationships across the entire brain. When these features are combined, they create a robust representation that enables the model to distinguish between closely overlapping tumor sub-regions. The subsequent knowledge distillation process transfers this nuanced understanding to the student model, ensuring that it, too, can effectively differentiate between tumor sub-regions despite the inherent difficulties in the image data.

D. Problem 4: Limited Annotated Data in Medical Imaging

Challenge: A significant limitation in medical imaging is the scarcity of annotated data, which can lead to overfitting and hinder the model's ability to generalize to new cases. **Solution**: Knowledge distillation plays a critical role in overcoming the data limitation issue. By transferring the rich, generalized features from the teacher model to the student model, we effectively regularize the training process. The teacher's guidance helps the student model focus on the most informative aspects of the data, thereby reducing the risk of overfitting. Additionally, the use of multi-scale feature fusion ensures that the model can robustly handle variations in tumor size, shape, and appearance. This combined approach enhances the model's ability to generalize from a limited number of annotated examples, making it more effective when applied to new patient data.

III. Methodology Overview

1) Hybrid Teacher Model
    a) VGG19 Branch: This branch of the teacher model is responsible for extracting detailed, local features from the MRI scans. Early layers are frozen to retain the pre-trained weights, while later layers are fine-tuned to adapt to the specific task of tumor segmentation
    b) ViT Branch: The transformer branch captures global dependencies and overall context in the images. The ViT model is adjusted to ensure its output aligns with the segmentation requirements.
    c) Feature Fusion: The outputs from both branches are combined. This concatenation of local and global features is then processed through an additional classifier that integrates these diverse representations into a single, robust feature vector for each image.
2) Knowledge Distillation Process
    a) Features are extracted at multiple scales from both the CNN and transformer components. This multi-scale extraction enables the model to capture a broad spectrum of details, from minute textures to global spatial arrangements.
    b) The features obtained from different scales are combined through a weighted aggregation process, allowing the model to emphasize the most informative features at each scale. This comprehensive feature integration is crucial for accurately detecting and segmenting the various tumor regions within the MRI scans.

IV. Conclusion

This report has outlined an innovative approach to brain tumor segmentation using the BraTS23 dataset. By combining the detailed local feature extraction capabilities of VGG19 with the global contextual understanding of ViT, our hybrid teacher model is able to address the inherent challenges of medical image complexity and limited annotated data. The process of knowledge distillation transfers these strengths to a smaller, more efficient student model, ensuring that high accuracy is maintained even in resource-constrained settings. Additionally, the integration of multi-scale geometric feature fusion enhances the model's ability to accurately delineate tumor sub-regions, addressing the challenges posed by overlapping and irregular tumor boundaries.

Through the use of these techniques—cross-architectural knowledge distillation and multi-scale feature fusion—we achieve a robust and efficient segmentation model. This approach not only minimizes computational overhead but also improves generalization, making it a promising solution for practical medical imaging applications.

Future work will focus on fine-tuning the multi-scale fusion parameters and further refining the distillation process to ensure even greater performance and adaptability in clinical environments.

Date of Submission: 16/2/2025 CSE499B Section 20