Weekly Report 1: 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

Preprocessing and Model Preparation for Cross-Architectural Knowledge Distillation Using the BraTS23 Dataset

I. Introduction

This report provides an update on the initial phase of our research work involving cross-architectural knowledge distillation for medical image analysis using the BraTS23 dataset. The focus of this update is on dataset preprocessing before training the hybrid teacher model, which combines VGG19 and Vision Transformer (ViT). Additionally, we discuss the adjustments required for implementing multi-scale geometric feature fusion in later stages.

II. Problem Statement

Large, high-performing models in medical imaging achieve remarkable accuracy but come with significant computational costs. This makes them unsuitable for deployment in resource-limited environments such as edge devices. Knowledge distillation presents a promising solution to address this issue by transferring knowledge from complex models to smaller, efficient ones. However, there is a gap in techniques that integrate features from different architectures for optimal knowledge transfer. This project aims to fill that gap by implementing cross-architectural knowledge distillation with MSGFF to enhance both efficiency and accuracy in medical imaging tasks.

III. Literature Review

Yufan Liu et al. introduces a novel method [1] for distilling knowledge from Transformer models to Convolutional Neural Networks (CNNs). The authors propose two projectors, the Partially Cross Attention (PCA) projector and the Group-wise Linear (GL) projector, to align student features with teacher features in two projected feature spaces. Additionally, a multi-view robust training scheme is introduced to enhance the robustness and stability of the framework. Extensive experiments on datasets like ImageNet and CIFAR demonstrate that the proposed method outperforms 14 state-of-the-art methods, showcasing its effectiveness in cross-architecture knowledge distillation.

Semantic segmentation of remote sensing images aims [2] to classify each pixel, which is crucial for urban planning and dynamic monitoring. However, existing models struggle with small target pixels and tiny target sizes, leading to poor recognition and segmentation. Additionally, deeper feature extraction modules result in redundant parameters and increased computation time. To address these issues, the authors propose KDMSANet, a lightweight semantic segmentatio network that uses knowledge distillation, a multiscale pyramidal pooling module, and an attention mechanism to enhance feature fusion and focus. They trained teacher-student models to create lightweight network models through model pruning and distillation. Experiments on the Vaihingen and Potsdam datasets showed that the proposed network significantly reduces the number of parameters while maintaining accuracy, with the student model's size reduced by 43.6%, training efficiency improved by 22.3%, and accuracy reaching 99.30% of the teacher model.

The document [3] "Feature Extraction from Point Clouds" by Gumhold, Wang, and MacLeod presents a method for detecting feature lines in point clouds without surface reconstruction. It involves assigning penalty weights to points and edges in a neighbor graph to identify feature patterns, followed by recovering feature lines and junctions using wedge and corner fitting. Key steps include graph construction using Delaunay filtering, density estimation, and point classification into surface, crease, or border categories, with penalty functions evaluating the likelihood of feature points. A modified minimum spanning tree extracts feature patterns, while least squares fitting projects noisy data onto feature lines. The approach is robust against noise and has applications in surface meshing, point cloud enhancement, and non-photorealistic rendering, with future work focusing on integrating detection and recovery stages and addressing isolated peaks.

[4] The 3D point cloud (3DPC) has advanced with deep learning (DL), but DL faces challenges like data scarcity and high computational needs. Deep transfer learning (DTL) reduces costs by using knowledge from one task to train

another and is effective for aligning point clouds. Domain adaptation (DA), a subset of DTL, improves data quality by addressing noise and missing points. This paper reviews techniques for understanding 3DPC using DTL and DA, covering applications like object detection and segmentation, and discusses the pros and cons of these frameworks, open challenges, and future research directions.

IV. Solutions

- **Solution 1**: By distilling knowledge from a large, high-performing teacher model into a smaller, more efficient student model, you reduce computational costs while maintaining high detection accuracy.
- **Solution 2**: Our approach focuses on capturing both local and global features, enabling the model to better differentiate between tumor sub-regions and improve segmentation accuracy.
- **Solution 3**: Knowledge distillation helps improve generalization by transferring robust feature representations from the teacher to the student. Additionally, multi-scale feature fusion ensures the model is robust to variations in object size, shape, and appearance.

V. Dataset Overview: BraTS23

The BraTS23 (Brain Tumor Segmentation 2023) dataset consists of multi-modal MRI scans (T1, T1Gd, T2, and FLAIR) of brain tumor patients. This dataset is widely used for tumor segmentation and classification tasks. Since our work involves knowledge distillation and geometric feature fusion, proper preprocessing of this dataset is crucial to ensure optimal model performance.

Preprocessing the BraTS23 Dataset

To effectively train our hybrid teacher model, we need to preprocess the BraTS23 dataset. The preprocessing pipeline involves several key steps:

1) Data Acquisition and Loading
    a) Extract MRI scans.
    b) Organize images based on their modalities (T1, T1Gd, T2, and FLAIR) and corresponding ground truth segmentation masks.
    c) Convert images from DICOM or NIfTI format to a standardized format (PNG or NumPy arrays) for compatibility with PyTorch and torchvision models.
2) Image Normalization and Intensity Standardization
    a) Normalize pixel intensities to a range of [0,1] or use z-score normalization to standardize MRI intensities across patients.
    b) Apply histogram equalization or CLAHE (Contrast Limited Adaptive Histogram Equalization) to enhance image contrast.
3) Resizing and Resampling
    a) Convert images to a uniform size (224x224 pixels) to match the input requirements of VGG19 and ViT.
    b) If needed, resample the images to a consistent voxel spacing using interpolation techniques to preserve anatomical structures.
4) Data Augmentation To enhance model robustness, apply the following augmentations:
    a) Geometric transformations: Random rotations, flipping, and affine transformations.
    b) Intensity-based transformations: Gaussian noise addition, brightness and contrast adjustments.
    c) Elastic deformation: To simulate real-world variations in MRI scans.
5) Mask Preprocessing and Label Encoding
    a) Convert segmentation masks into categorical labels for classification.
    b) Ensure masks align properly with the preprocessed images.
    c) One-hot encode labels if required for multi-class segmentation.
6) Dataset Splitting
    a) Split the dataset into training (80%), validation (10%), and test (10%) sets.
    b) Ensure balanced distribution of tumor types across splits.

VI. Hybrid Teacher Model: VGG19 + Vision Transformer (ViT)

Model Architecture

Our hybrid teacher model integrates two different architectures:

1) VGG19 with Batch Normalization
    a) Extracts spatial features from MRI scans.
    b) Fully connected classifier adapted to the number of classes in BraTS23.
2) Vision Transformer (ViT-B/16)
    a) Captures global contextual features using self-attention mechanisms.
    b) Modified classification head to match the number of tumor classes.
3) Fusion Mechanism
    a) The outputs from VGG19 and ViT are concatenated.
    b) A final linear classifier is used to generate predictions.

VII. Future Adjustments for Multi-Scale Geometric Feature Fusion

To implement multi-scale geometric feature fusion, additional adjustments to preprocessing and model input representation are required:

1) Multi-Resolution Image Processing
    a) Generate multiple resolutions of MRI scans (e.g., 112x112, 224x224, 448x448).
    b) Enable different models to extract features at varying scales.

2) Feature Map Alignment
    a) Ensure extracted feature maps from VGG19 and ViT are geometrically aligned before fusion.
    b) Use deformable convolutions or spatial transformers if necessary.
3) Graph-Based Feature Fusion
    a) Explore graph neural networks (GNNs) to model spatial relationships across different resolutions.
    b) Implement geometric-aware attention mechanisms to enhance object localization in tumors.

VIII. Conclusion

In this phase, we have outlined the preprocessing steps necessary to prepare the BraTS23 dataset for training. These steps include data acquisition, normalization, augmentation, and dataset splitting. Additionally, we have provided an overview of our hybrid teacher model, which combines VGG19 and ViT for feature extraction and classification. In the next phase, we will explore multi-scale geometric feature fusion, requiring additional adjustments to preprocessing and feature alignment.

This foundational work ensures that the dataset is well-prepared for training, paving the way for effective knowledge distillation and improved medical image analysis.

IX. Next Steps

1) Implement the preprocessing pipeline by writing scripts for each step and ensuring that the processed images maintain anatomical integrity and consistency across different modalities.
2) Train the hybrid teacher model on the preprocessed BraTS23 dataset.
3) Develop and integrate multi-scale geometric feature fusion mechanisms.

References

[1] Y. Liu, J. Cao, B. Li, W. Hu, J. Ding, L. Li, *Cross-Architecture Knowledge Distillation* [Online]. Available: https://arxiv.org/abs/2207.05273

[2] Y. Yang, Y. Wang, J. Dong, B. Yu. ”Diabetic retinopathy: A growing cause of blindness in Bangladesh,” in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024, pp. 1-14.

[3] S. Gumhold, X. Wang, R. MacLeod, *Feature Extraction From Point Clouds* [Online]. Available: https://graphics.stanford.edu/courses/cs164-10-spring/Handouts/papers_gumhold.pdf

[4] S. S. Sohail , Y. Himeur, H. Kheddar , A. Amira, F. Fadli, S. Atalla, A. Copiaco, W. Mansoor, *Advancing 3D Point Cloud Understanding through Deep Transfer Learning: A Comprehensive Survey* [Online]. Available: https://arxiv.org/pdf/2407.17877v1

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