Project Proposal: Enhancing U-Net with Lightweight Attention for Efficient Medical Image Segmentation

Jason Zou (jz3481)

I. Problem Overview

Medical image segmentation is a critical task in healthcare, enabling precise delineation of anatomical structures and abnormalities for diagnosis and treatment planning. Traditional U-Net architectures have been widely adopted for this purpose due to their effectiveness in capturing context and producing accurate segmentations. However, the standard U-Net can be computationally intensive, potentially posing challenges for deployment in resource-constrained environments such as portable medical devices or more real-time applications. This project aims to enhance the U-Net architecture by integrating lightweight attention mechanisms in order to improve segmentation accuracy while reducing computational complexity.

II. Literature Review

- **U-Net Architecture**: Ronneberger et al. introduced the U-Net, a convolutional network designed for biomedical image segmentation [1]. Its encoder-decoder structure with skip connections allows for effective feature extraction and spatial information preservation.
- **Squeeze-and-Excitation Networks**: Hu et al. proposed the Squeeze-and-Excitation (SE) block, which adaptively recalibrates channel-wise feature responses by modeling interdependencies between channels, enhancing representational power with minimal computational overhead [2].
- **Depthwise Separable Convolutions**: Chollet introduced depthwise separable convolutions in the Xception network, decomposing standard convolutions into depthwise and pointwise convolutions to reduce parameters and computation while maintaining performance [3].
- **Attention U-Net**: Oktay et al. developed the Attention U-Net, incorporating attention gates into the U-Net architecture to focus on relevant regions in medical images, improving segmentation accuracy without significant computational cost [4].

These studies highlight the potential of integrating attention mechanisms and efficient convolutional operations to enhance the performance and efficiency of medical image segmentation models.

III. Proposed Method and Experiments

1) **Dataset Selection**: We will utilize a publicly available medical image segmentation dataset, ensuring high-quality annotations for training and validation. Potential datasets include:
    - **ISIC 2018 Skin Lesion Analysis Dataset** (2D): A large collection of dermoscopic images used for skin lesion segmentation. This dataset is well-documented and widely benchmarked, making it ideal for developing and evaluating segmentation models for dermatology applications [5].
    - **BraTS (Brain Tumor Segmentation Challenge)** (3D): A dataset of multimodal MRI scans with expert-labeled segmentations of brain tumors. This dataset is useful for testing deep learning models on volumetric segmentation tasks [6].
2) **Model Architecture**:
    - **Baseline Model**: Implement the standard U-Net architecture as described by Ronneberger et al.
    - **Enhanced Model**: Integrate Squeeze-and-Excitation (SE) blocks into the U-Net architecture to recalibrate channel-wise features, aiming to improve focus on relevant features without significantly increasing computational cost.
3) **Optimization Techniques**:
    - **Depthwise Separable Convolutions**: Replace standard convolutions in the U-Net with depthwise separable convolutions to reduce the number of parameters and computational load.
    - **Quantization-Aware Training**: Apply quantization-aware training to prepare the model for deployment on edge devices, ensuring reduced model size while maintaining accuracy.
4) **Evaluation Metrics**:
    - **Accuracy Metrics**: Evaluate segmentation performance using Dice Similarity Coefficient (DSC) and Intersection over Union (IoU).
    - **Efficiency Metrics**: Assess computational efficiency by measuring Floating Point Operations Per Second (FLOPs) and inference time.
5) **Implementation Plan**:
    - **Framework**: Utilize PyTorch for model development due to its dynamic computation graph and extensive support for medical imaging applications.
    - **Training Strategy**: Employ data augmentation techniques to enhance model generalization and use early stopping based on validation loss to prevent overfitting.

- **Baseline Establishment**: Train the baseline U-Net model and record performance metrics for comparison.
- **Model Enhancement**: Integrate SE blocks and depthwise separable convolutions into the U-Net architecture, followed by retraining and evaluation.
- **Optimization Application**: Apply quantization-aware training to the enhanced model and assess any changes in performance and efficiency.

IV. Division of Labor

As this is an individual project, all tasks, including literature review, dataset preparation, model implementation, training, evaluation, and report writing, will be conducted independently.

V. Originality and Data Curation

This project aims to innovate by combining SE blocks and depthwise separable convolutions within the U-Net architecture, a configuration not extensively explored in existing literature. While the ISIC 2018 dataset is publicly available, we will perform data curation by preprocessing images and augmenting the dataset to enhance the model robustness.

VI. Scope and Feasibility

The project is designed to ensure the development of a functional segmentation model with demonstrable improvements in efficiency and accuracy. The integration of lightweight attention mechanisms and efficient convolutions is anticipated to yield measurable performance gains.

References

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv. (MICCAI)*, 2015, pp. 234–241. [Online]. Available: https://arxiv.org/abs/1505.04597

[2] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 7132–7141. [Online]. Available: https://arxiv.org/abs/1709.01507

[3] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 1251–1258. [Online]. Available: https://arxiv.org/abs/1610.02357

[4] O. Oktay et al., "Attention U-Net: Learning Where to Look for the Pancreas," *arXiv preprint arXiv:1804.03999*, 2018. [Online]. Available: https://arxiv.org/abs/1804.03999

[5] N. C. F. Codella et al., "Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC)," *arXiv preprint arXiv:1902.03368*, 2019.

[6] B. H. Menze et al., "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993-2024, 2015.