UMamba-MoE : A Framework for Enhanced Image Segmentation

Pradeep Rajan

A dissertation submitted in partial fulfilment
of the requirements for the degree of
Master of Science in Artificial Intelligence
of the
University of Aberdeen.

Department of Computing Science

2024

Abstract

Medical image segmentation is required for accurate anatomical segmentations for disease diagnosis and treatment planning, one of the methods to achieved it is with the help of deep learning models. One of the traditional deep learning models that was widely utilized for Medical Image Segmentation was the Convolutional Neural Networks (CNNs), which produced great results. However, due to their local receptive field constraint, they were not able to capture global context information within an Image effectively. Hence, Transformers were introduced to overcome their drawbacks and limitations. The use of Transformers led to a significant improvement in the prediction, accuracy and metrics of the results. Even though certain advancements, such as the self-attention mechanism, were introduced, they were unable to capture long-range sequences effectively as anticipated. To address these challenges, U-Mamba was produced as a solution. This bi-directional state space model architecture is believed to outperform both CNN and Transformers. By implementing, Mamba is able to capture the relationships between the distant parts of an image.

The U-Mamba architecture is integrated with 'nnU-Net,' a combination of the semantic segmentation CNN model and Mamba network. The nnU-Net, with the help of a dataset fingerprint, nnU-Net automatically adapts its architecture for preprocessing, training strategies, and inference to the specifics of the dataset. The name suggests that it is based on a U-Net model. With the help of this integration, long-range sequences can be easily captured effectively.

Even though it produces the required results, the novelty model that we tried to proposed has helped improve the inference by increasing accuracy and other metrics such as Dice Similarity Coefficient (DSC) and Normalized Surface Distance (NSD). It can clearly observe a difference in value for DSC, which improved by at least 5%, and NSD by at least 2% for 2D images. The novelty introduced enriches the existing model 'U-Mamba' by enhancing the system's adaptability and performance across different datasets. This improvement is mainly facilitated by a concept known as the Mixture of Experts (MoE). The MoE deploys multiple experts, which are constructed around only one network, the Mamba network. When an expert of four is defined, about four Mamba networks are created and made to run in parallel, allowing us to achieve the desired outcome much effeciently. This novel strategy has attempted to introduce has facilitated the advancement of medical image segmentation, achieving significant advancements in medical image analysis through enhanced adaptability and accuracy across varied datasets.

Goals

For this dissertation, two main goals are strived to be achieved here,

- Reduce Computation Cost: The first goal is to reduce the computation cost significantly by reducing the required training steps. This is achieved with the help of the MoE approach. The MoE leverages parallel computational capabilities by concurrently processing inputs across multiple expert networks. This strategy not only accelerates the segmentation process but also increases the scalability by allocating resources dynamically.
- Enhance Segmentation Accuracy: The second goal is to improve the segmentation accuracy. Integrating MoE exclusively into the Mamba model helps identify intricate patterns and long-range spatial dependencies efficiently. This focused integration is designed mainly to focus on the Mamba model's generalization power, improving its performance and the accuracy of its segmentation outputs.