Enhancing Image Classification with a Hybrid CNN-Transformer Model: A Comparative Study of ResNet-18 and a Modified Architecture

Chinmaya Mathur[*]

February 2025

Abstract

In this thesis, we propose a Hybrid model that integrates the strengths of Convolutional Neural Networks (CNNs) and transformer encoders to enhance image classification. We specifically modify the ResNet-18 by replacing its 4th block with a transformer encoder which includes a multi-head self-attention layer and a position-wise feedforward network. This modification aims to leverage the transformer's ability to capture long-range dependencies and improve the feature extraction capability of the model.

On evaluating the performance of both the models on the CIFAR-10 dataset, we see that the Hybrid model performs slightly better than ResNet-18. The classwise accuracy analysis shows that the Hybrid model performs better for several classes like "airplane", and "dog" but shows a decrease in accuracy for classes like "cat". To understand the impact of architectural modification, we compare the weights of the first 3 blocks using a quantile-quantile (QQ) plot. The analysis shows that the weights remain largely similar in distribution but the magnitude changes with the Hybrid model having bigger weights.

We further analyze the significance of the changes in classwise accuracies using the Wilcoxon signed rank test that confirms the observed changes in accuracy are significant across all classes but the magnitude of change in medians of the difference in the accuracy of the two models is not big in all classes. Our findings support the integration of the transformer encoder into CNN architecture but we see that the performance of the model can still be increased by introducing regularization terms in the training. We can also explore different configurations using a transformer encoder and experiment with different datasets to generalize our results and further improve model accuracy.

---
[*]Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: mathurchinmaya@gmail.com. Supervisor: Chun-Biu Li.