COSE474-2024F: Final Project Proposal
Hyperparameter Optimization for Transferring CLIP to a Specific Domain

JooYoung Park

1. Introduction

In this paper, we propose an approach to mitigate catastrophic forgetting by treating the indices of extracted layers from the pretrained model as hyperparameters. We investigate how freezing different layers impacts model efficiency and training cost, comparing the trade-offs involved in various scheduling techniques for these hyperparameters.

The selection of which layers to freeze during fine-tuning can be guided by multiple strategies. One approach involves comparing the similarity between the source dataset (pretrained model) and the target dataset. Another option is to calculate the gradients of the respective layers and analyze their impact on the models. Additionally, a mathematical model for layer permutation can be employed to further optimize fine-tuning. Each of these optimization strategies will be evaluated and compared in terms of performance and computational cost by the conclusion of this project.

2. Problem definition & chanllenges

Pretrained Vision-Language models, such as CLIP, have demonstrated significant potential for transfer learning applications. However, during fine-tuning on domain-specific tasks (e.g., medical imaging or autonomous driving), challenges like catastrophic forgetting arise. This phenomenon occurs when the model loses previously learned information during adaptation, leading to inefficiencies and undermining the purpose of transfer learning.

3. Related Works

According to previous research, transfer learning techniques in neural network architectures can leverage pretrained models to aid in the creation of models used in specific domains [1]. This approach can also be applied to CLIP, a computer vision model that utilizes the transformer architecture. By efficiently borrowing the low-level layers trained on a dataset significantly larger than IMAGENET, [2] it is expected to achieve good performance in new domains.

4. Datasets

The ChestX-ray14 dataset will be used as the target domain for transfer learning. We will compare the performance of various methods to extract optimized hyperparameters. Additionally, we will evaluate the efficiency of these methods using other datasets, such as LUNA16 and BRATS.

5. State-of-the-art methods and baselines

The baseline metric is obtained by testing the images from the dataset using the existing CLIP model without fine-tuning or transfer learning. In contrast, the SOTA method represents the metric achieved when full fine-tuning is performed, and we will also employ a re-purposing approach by naively removing only some of the top layers and fine-tuning the model.

References

[1]Yosinski, J., Clune, J., Bengio, Y., Lipson, H. (2014). How transferable are features in deep neural networks?. Advances in neural information processing systems, 27.

[2]Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... Sutskever, I. (2021, July). Learning transferable visual models from natural language supervision. In International conference on machine learning (pp. 8748-8763). PMLR.