COSE474-2024F: Final Project Proposal
"Comparing Fine-Tuned Legacy Architecture with Non-Fine-Tuned Modern Architecture Model: ResNet-18 vs ConvNeXT-Tiny on CIFAR-10"

Jaehoon Han, 2022320101

1. Introduction

ResNet, a CNN architecture used for image classification, was introduced by Microsoft Research in 2015. It has been continuously developed and many architectures are introduced such as ViT and Swin Transformer, etc. In 2022, Meta AI Research introduced ConvNeXT, a state-of-the-art(SOTA) architecture.

In this research, we explore the effect of model architecture and fine-tuning on performance on specific tasks. This research is helpful to selecting the model for other researches whether fine-tuned legacy architecture or pre-trained modern architecture.

2. Problem definition & chanllenges

ConvNeXT shows higher performance than ResNet. However, we cannot know exactly how fine-tuning is impacts the performance. We're going to estimate the effect of model architecture and fine-tuning.

To comparing, we will use fine-tuned ResNet-18 and only pre-trained(non-fine-tuned) ConvNeXT-Tiny models.

3. Related Works

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778. https://doi.org/10.1109/cvpr.2016.90.

Liu, Z., Mao, H., Wu, C., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A ConvNet for the 2020s. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 11966-11976. https://doi.org/10.1109/CVPR52688.2022.01167.

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks?. ArXiv, abs/1411.1792.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. ArXiv, abs/2010.11929.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 9992-10002. https://doi.org/10.1109/ICCV48922.2021.00986.

4. Datasets

Due to limited resource, we use CIFAR-10 which is a widely used dataset consisting of low-resolution images.

5. Schedule & Roles

- Do fine-tuning ResNet-18 with some or all classes of CIFAR-10.
- Estimate accuracy of models.
  - Fine-tuned ResNet-18 with 2 classes of CIFAR-10
  - Fine-tuned ResNet-18 with 5 classes of CIFAR-10
  - Fine-tuned ResNet-18 with all(10) classes of CIFAR-10
  - Non-fine-tuned ConvNeXT
- Comparing results.
- Write a paper