COMS4995 Neural Networks and Deep Learning: Project Proposal

Xiaofei Gao
Columbia University
New York, NY 10027
xg2429@columbia.edu

Hellen Zhao
Columbia University
New York, NY 10027
hz2843@columbia.edu

Thea Zhu
Columbia University
New York, NY 10027
wz2636@columbia.edu

1 Introduction

In real-world classification tasks, especially in domains like ecology, medicine, or security, models often face unseen categories or shifts in data distributions. This project simulates such a realistic challenge by introducing domain shift and new sub-classes during testing. Transfer learning has become a critical technique in such settings, as it enables models to adapt knowledge from existing data to new, related tasks with limited labeled examples.

The core problem is to develop a model that can accurately classify both the super-class (bird, dog, reptile) and the sub-class of images, even when the sub-class distribution during testing differs from that during training, and potentially includes previously unseen sub-classes. This requires building models that generalize well, rather than overfitting to the training distribution.

This project is an opportunity to experiment with and understand transfer learning in a practical setting, particularly how it helps models adapt to distribution shifts and novel classes. It also offers a platform to test techniques like fine-tuning pre-trained models, data augmentation, and domain adaptation. We are interested in this task because it mirrors real-world machine learning problems where perfect, static training data is rare.

2 Related Work

Convolutional Neural Networks in Image Classification Convolutional Neural Networks (CNNs) are proved to be very effective in the image classification tasks. Reviewing the winners of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), we can see that almost all of them used CNNs. Currently, the populat CNNs with lower error rates are VGGNet, GoogLeNet/Inception V1-V4, ResNet, EfficientNet, etc [1]. Among them, Inception V4 reached a top-1 error rate of 16.4% on the ImageNet dataset [2], while EfficientNet V2 reached a top-1 error rate of 12.7% on the ImageNet dataset [3].

Transfer Learning Training a model from scratch is time-consuming and requires a lot of computing resources. By reusing a pre-trained model, we only need to fine tune the model on a smaller dataset to solve a new problem. There are various techniques we can apply during transfer learning, such as freezing certain layers in the pre-trained model, or fine tune the layers.

Transfer Learning in Image Classification In image classification, transfer learning is very helpful because there're a lot of available models that were pre-trained on ImageNet dataset. Kolesnikov et

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

al. (2020) introduced Big Transfer (BiT) [4], a simple recipe that combined selected components and achieved strong performance on over 20 datasets. There are many works have been done applying transfer learning into different image classification tasks, such as food image classification [5] and dog breed classification [6]. In particular, it is widely used in the medical image classification tasks. For example, Davila et al. (2024) evaluated eight fine-tuning strategies based on three pre-trained models across different types of medical images [7], and found the pre-trained model and fine-tuning method need to be chosen based on the target dataset.

3 Method

Our objective is to explore the potential that transfer learning can have on image super-class and sub-class labeling. In particular, we are interested in investigating different approaches to fine-tuning a preexisting network trained on a related dataset.

Our data will consist of images from 3 classes of animals, including bird, dog, and reptile. Each of these classes will also include their own sub-classes (such as hawk in bird, golden retriever in dog, etc.). One significant challenge we will need to overcome is the possibly differing distribution of classes between the training data and real world data (in this case testing data), along with the possible introduction of new sub-classes not present during training.

To combat these difficulties, we plan to employ various data augmentation techniques. We will utilize geometric transformations (rotation, flipping, cropping, translation, etc.), color adjustments (in brightness, contrast, saturation, jittering, etc.), and Gaussian noise injection to artificially expand our dataset and ensure we have sufficient training data in all classes.

For our pretrained model, we plan to choose an image classifier trained a source dataset that bears resemblance to our target dataset consisting of animal images. Consequently, we will use Inceptionv4 which was trained on ImageNet, a large-scale annotated dataset that itself contains many animal images.

The areas of fine-tuning we are interested in investigating include which (and how many) layers of Inceptionv4 to freeze. After training different models under this hyperparameter, we will evaluate and compare their performances based on accuracy, training speed (measured by number of iterations before convergence), and required compute power (measured by FLOPs usage). Furthermore, we will train a simple convolutional neural network from scratch to be used, along with the pre-trained Inceptionv4, as benchmarks for evaluating our fine-tune models.

In this way, our experiments will help us understand what sort of tradeoff there is between the effectiveness of transfer learning and computational resources.

References

[1] Chen, Leiyu & Li, Shaobo & Bai, Qiang & Jing, Yang & Jiang, Sanlong & Miao, Yanming. (2021). Review of Image Classification Algorithms Based on Convolutional Neural Networks. Remote Sensing. 13. 4712. 10.3390/rs13224712.

[2] Szegedy, Christian & Ioffe, Sergey & Vanhoucke, Vincent & Alemi, Alexander. (2016). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. AAAI Conference on Artificial Intelligence. 31. 10.1609/aaai.v31i1.11231. Review of Image Classification Algorithms Based on Convolutional Neural Networks. Remote Sensing. 13. 4712. 10.3390/rs13224712.

[3] Tan, M., & Le, Q. (2021, July). Efficientnetv2: Smaller models and faster training. In International conference on machine learning (pp. 10096-10106). PMLR.

[4] Kolesnikov, A., Beyer, L., Zhai, X., Puigcerver, J., Yung, J., Gelly, S., & Houlsby, N. (2020). Big transfer (bit): General visual representation learning. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16 (pp. 491-507). Springer International Publishing.

[5] Burkpalli, Vishwanath & Patil, P.C.. (2020). An efficient food image classification by inception-V3 based cnns. International Journal of Scientific and Technology Research. 9. 6987-6992.

[6] Agarwal, Ambuj & Kiran, Vidhu & Jindal, Rupesh & Chaudhary, Deepak & Tiwari, Raj. (2022). Optimized Transfer Learning for Dog Breed Classification. 10.13140/RG.2.2.35522.04807.

[7] Davila, Ana & Colan, Jacinto & Hasegawa, Yasuhisa. (2024). Comparison of fine-tuning strategies for transfer learning in medical image classification. Image and Vision Computing. 146. 105012. 10.1016/j.imavis.2024.105012.