Improving Deep Learning Model Over Multiple Runs

Prepared by	Xingyu Liu, xingyuliu@g.harvard.edu Lu Yu, luyu@g.harvard.edu Alex Leonardi, aleonardi@college.harvard.edu Chris Gilmer-Hill, cgilmerhill@college.harvard.edu
Prepared for	Jonathan Frankle, jfrankle@seas.harvard.edu

Background

One of the primary bottlenecks in machine learning pipelines is the time required to train the model; consequently, any innovations that improve time-efficiency in the training process have the potential to greatly expand the utility of machine learning as a tool across a broad range of applications. In particular, this project focuses on leveraging data from prior training runs to decrease the time required for subsequent model re-trainings, e.g. on newly gathered data.

One method explored in prior work to accomplish the goal of improving model inference time from past trained models is distillation, which leverages prior training runs to simplify the task of retraining the model, in effect "transferring" stored information from a pre-existing (and generally more complex) model to a new (and often simpler) one. Extending the process to the re-training of models, it may be possible to leverage a similar technique to transfer/compress stored knowledge from one or multiple previous model iterations into a new iteration incorporating newly-added data, without necessarily repeating the entire training process.

In addition to directly investigating improvements in the training process, this project will also seek to develop a framework for evaluating and comparing the time-efficiency of the model re-training techniques we explore. Such a framework has the potential to contribute significantly to future work in the broader field of improving the process of model training.

Problem Statement

This project entails exploring ways to reuse the computation that was invested in training these intermediate models to make training future models better. Specifically, it will entail (1) establishing a benchmark and metrics for testing out strategies for reusing this information, (2) developing and evaluating methods for reusing this information to improve metrics on the benchmark, and (3) putting forth recommended best practices for practitioners based on this research.

Resources Available

We think our strategy should be generally useful, not limited to a specific dataset. We plan to start with a simple image classification dataset, CIFAR-100, as our baseline and improve on it. After our method can work efficiently on CIFAR-100, we will try on another image classification dataset, like Caltech-101, to test our method's consistency and generality. Finally, if working well on image classification, we'll extend to more complex tasks — like segmentation, object detection, etc.

To sum up, the available resources are:

1. The **CIFAR-100** dataset is a opensource dataset that has 100 classes containing 600 images each. There are 500 training images and 100 testing images per class. The 100 classes in the CIFAR-100 are grouped into 20 superclasses. Each image comes with a "fine" label (the class to which it belongs) and a "coarse" label (the superclass to which it belongs).
2. Open-source tools built by MosaicML, like `Composer`, which contains a library of methods, and ways to compose them together for more efficient ML training.
3. AWS/MosaicML's cluster (if possible) for large-scale model training.

High-Level Project Stages

Broadly, we plan to organize our project into three major milestones:

1. **Landscape of Methods:** In our first milestone, we plan to explore various potential techniques for reusing results and information from previous runs to improve training time. We will test these techniques on relatively simple baseline problems without new data ingested, like five seperate runs on CIFAR-100 using ResNet-56.in order to simplify the model creation process. The evaluation framework would include metrics like the plot of time v.s. accuracy tradeoff, time spent on training runs, and how many steps the model takes to reach the full accuracy. The techniques we'll use for reusing past runs information to improve metrics on the benchmark would includes but not limited to: knowledge distillation, reloading past models and finetune on high-loss data, etc. The deliverable for this milestone will include the established benchmark and framework, and the results of our experiments on reusing past runs' information, hopefully allowing us to identify the most promising methods for further exploration.
2. **Deep Dive:** After selecting a few promising methodologies, we plan to test these methods on more complex settings. New data would be fed to the best model we find in the previous phase and we would evaluate the performance. It is possible that the methods we used in no-new-data setting would be less useful in the new-data settingand we'll seek more techniques to tackle such cases. After that, we would also explore different models on more complex tasks(e.g. segmentation, etc.) and compare the performance. The purpose of this deliverable would be to evaluate our methods on problems that are more emblematic of the actual use cases we see in the real world.
3. **Final Report:** The final report will compile our results into a single document with a recommended set of methodologies and use cases for reusing training results. At this point, we should be able to comment on the performance of the various methods across multiple baseline and complex issues to develop a set of final recommendations.

Project Timeline