Reproducing Researchers' metrics with AutoML

Chau Yuan Qi,
Intern@ST Autonomous Solutions

Abstract

We present a training procedure that automatically selects the optimizer and hyperparameters used in model training in order to reproduce and improve upon the metrics quoted by researchers. In this experiment, we used ImageNet1k as the train and validation data, and we managed to obtain a validation accuracy of 67.100 after only 65epochs of training (15hours on NVIDIA RTX A5000) in comparison with the researchers' 69.774 on resnet18.

1. Introduction

The training procedure is separated into three phase, i.e. Automatic Optimizer Search, Hyperparameters Tuning and Full Training. This training procedure is done under the assumption that whichever variable that performs well over a performance metric (validation accuracy in this case) after an initial fixed period of time, will continue to perform well in the future. This is to reduce the amount of uncertainty in selecting multiple variables where the number of possible combinations grows very quickly. The goal of this experiment is to present a procedure that is able to reproduce results quoted by researchers in the shortest amount of time possible, consistently.

2. Automatic Optimizer Search

Automatic Optimizer Search is done by training the model on a dataset for a fixed number of steps, and the optimizer that returns the minimal loss value will be selected. The hyperparameters for all training are arbitrarily selected and fixed at learning_rate=0.001, weight_decay=0.0001, and momentum=0.3. The details are shown in Figure 1. In this case, AdamP returns the lowest loss value and is selected.

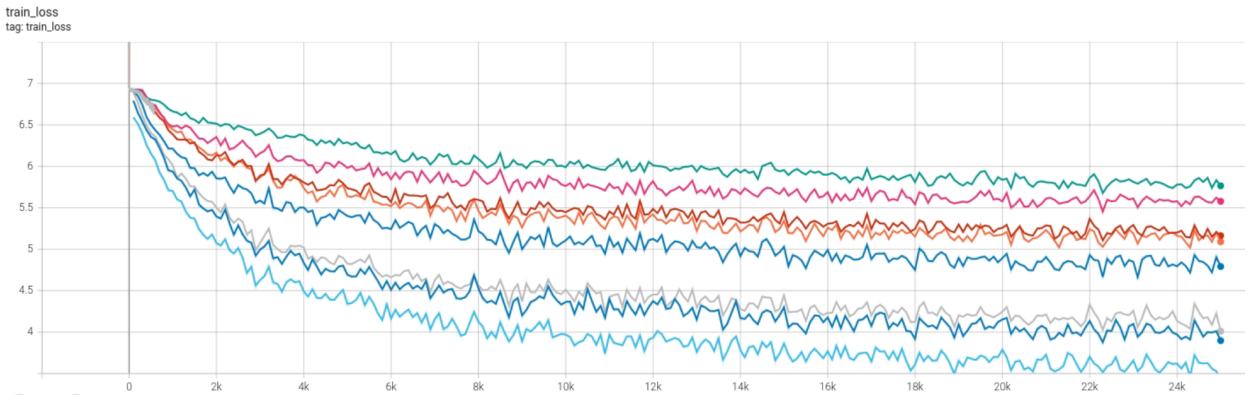

Figure 1: Training models on different optimizers

2.1 Hyperparameters Tuning

The selection of hyperparameters is done using Random Sampling from a pre-defined search space in Figure 2. The process is done using resources of 1 GPU/trial and 16 CPU Workers/trial with a batch size of 1024. The training for each set of hyperparameters tuning trial is scheduled using the ASHA algorithm that sets 3 rungs at epoch 1,3 and 9. The ASHA algorithm will stop training of trials at each rung if their performance is lower than other trials at the same rung. The set of hyperparameters that return the greatest validation accuracy will be selected. In this case, out of the 32 trials, only 4 remain in the final rung, and the best trial (trial 19) with a validation accuracy of 0.4319 is chosen. The hyperparameters of trial 19 were learning_rate=0.0012, weight_decay=0.0131, and momentum=0.7728.

Hyperparameters	Distribution	Range
learning_rate	loguniform	0.0001 - 0.1
weight_decay	loguniform	0.001 - 0.05
momentum	uniform	0 - 1

Figure 2: Search space for hyperparameters

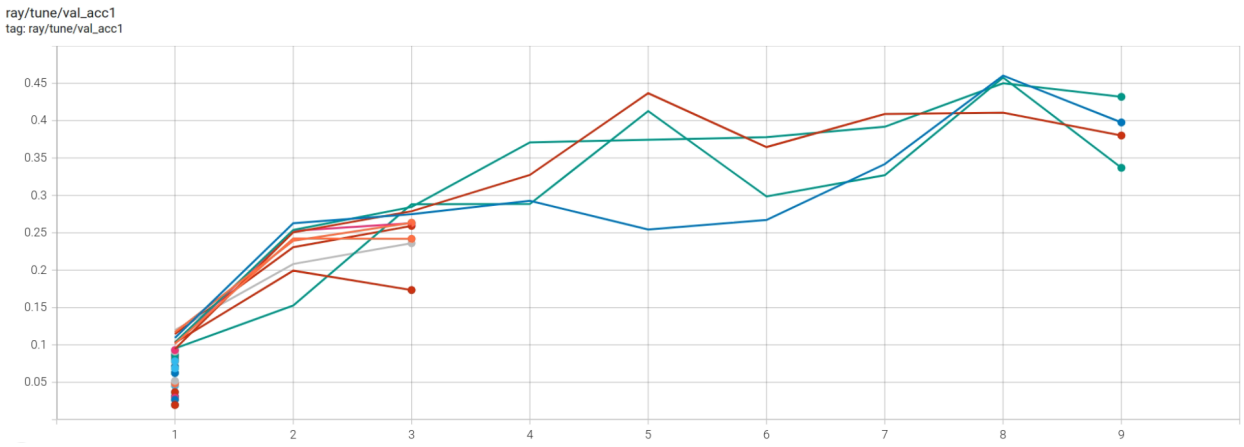

Figure 3: Validation accuracy against number of epochs during hyperparameters tuning

2.3 Full Training

By using the optimizer and hyperparameters selected from step 1-2, the model is trained for 200 epochs. This is done with a step scheduler with a decay epoch of 30 and decay rate of 0.1. After 65 epochs of training, we are able to achieve a validation accuracy of 0.671 as shown in Figure 4.

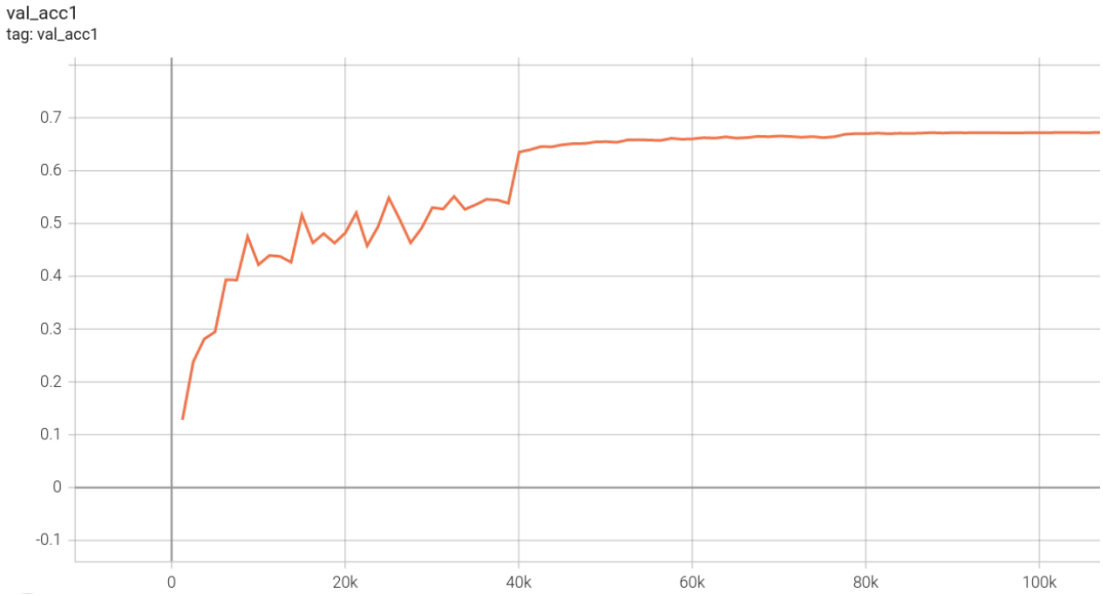

Figure 4: Validation Accuracy against training steps

3. Conclusion

The presented method provided a step-by-step procedure in order to train a new model and achieve good results consistently without having to manually tweak any variables. This is done in three steps, i.e. Automatic Optimizer Search, Hyperparameters Tuning and Full Training. After 65 epochs of training, we are able to achieve 97% of the researchers' quoted results. We believe that given enough time for training, reaching similar if not better results is definitely possible.