AutoML for CV

Astha Gupta

Gabe Cemaj

Idea

In recent years the idea of AutoML has become more prevalent and gained significant popularity. The concept of automating the process of picking and designing a ML pipeline is very attractive as this is often a very time consuming step. With the growth of ML and DL more and more decisions must be made to pick the correct combination of parameters to build a ML pipeline.

The space from which to pick extends well beyond tradition hyperparameters. Not only must we now pick an architecture, depth, width, activation function, and optimizations but there are also a large array of options to pick from for preprocessing and normalization steps. This large search space would benefit from an automated way of exploring it to find a combination that results in the best performance.

Challenges

Some of the main challenges we expect to run into while doing this work are related to the exponential number of knobs that can be tuned. Find a way to prioritize the parameters to tune to most efficiently improve performance while optimizing for cost. There are a lot of choices that can be made when picking a pipeline, especially when trying to explore the space from end to end, and picking what to optimize and by how much in each area will be a big challenge. And while there is a lot of work that has shown a lot of promise in subsections of a ML pipeline, little work has been done to try to automate the entire process.

Approach and Implementation

This project aims to string together many of the aspects of AutoML to build a grand unified solution to generate end to end ML pipelines. The idea is to combine neural architecture search, along with hyperparameter tuning and grander exploration of preprocessing combinations to automatically generate a full end to end pipeline.

To evaluate this project we will attempt to generate models to perform over a variety of standard benchmarking tasks in a few different domains. Domain we hope to explore:

Computer Vision (ImageNet, COCO, BraTS), NLP (SQuAD), and other domains (Click Logs). We will compare our auto generated models to the STOA in for each dataset and evaluate the cost benefit analysis of the extra computer needed to generate these automatically. We hope to answer the question of how one might best spend their resources when trying to build a full ML pipeline in a production system (comparing performance vs compute cost vs manual tuning costs).

Planned Demo

We aim to show off a few auto generated models and their performance in comparison to those in STOA. An important goal of this project is to be able to demonstrate high performance cross domain, to show that we can generate automated end to end pipelines for a variety of data types.

Previous Work

This project aims to combine many ideas from previous work. This paper CFG AutoML, proposed a methodology to explore AutoML piepliens in an end to end manner. We hope to take that and expand it with DARTS to explore the architecture space. By combining these two papers we hope to achieve a differentiable space that consists not only of the neural space but the entire pipeline. We also hope to combine this with a standard hyperparameter tuning algorithm to further tune the entire pipeline.