CS 7643 Project Proposal - Fall 2022

Team Name: Team Mix & Fix & Match

Project Title: Exploring and Extending *ixMatch Semi-Supervised Learning

Project Summary: State of the art gains in deep learning fields can be attributed in large part to rich datasets. Development on architectures, optimizers, and loss functions is important and valuable, but those can't do anything on their own without informative labeled datasets. This represents a limiting factor in model development, since high quality labeled datasets are human labor intensive and overall expensive to produce and curate compared to unlabeled datasets. Unsupervised learning is one option, but forgoing labels entirely severely limits options. Exploring semi-supervised learning, where a mixture of labeled and unlabeled data is used, is an avenue that allows networks to potentially learn more from a lot less labeled information in the long run. This presents interesting challenges for model structure, loss terms, and other architecture choices that can't necessarily be solved by just throwing more parameters and training epochs at the problem. To this end, our project will explore some state of the art semi-supervised learning techniques and attempt to profile their capabilities in great detail.

Approach: We plan to re-implement the two holistic semi-supervised learning approaches: MixMatch [1] and FixMatch [2] in Pytorch. We plan to use the authors' Tensorflow code as reference. We will implement a modular system for hot-swapping different pre-trained models, different augmentation techniques (MixUp, Weak Augmentations, Strong Augmentations) & other MixMatch & FixMatch adjustments. This modular implementation will help us experiment with various combinations of consistency regularization (predictions not affected by perturbations) & pseudo/proxy labeling techniques (confident predictions on unlabelled data).

We want to run the experiments on a pretrained Wide ResNet-28 model against CIFAR-10 & CIFAR-100 databases. As an extension, we plan to use other common image classification models (VGG) and databases (SVHN & STL-10). Our main experimental goal is to observe the percentage error with varying amounts of labeled data and compare against the full supervised benchmark. We plan to assess various hyperparameter optimizations that were explored across [1],[2],and [3], including learning rate decay, weight decay, and momentum. Accuracy will also be observed for a variable number of labeled examples (250, 500, 1000, 2000, 4000) from the CIFAR-10 & CIFAR-100 datasets. From here, performance across the datasets will be explored to assess whether state-of-the-art performance can be achieved as claimed for MixMatch and FixMatch while including/excluding MixUp. For research on generalizability, we will also experiment with smaller base model architectures, such as "Resnet-18", to see how well the results hold. As a stretch goal, we hope to explore the performance of audio data well, as mentioned in the Mixup [3] analysis of the Google speech commands dataset.

As extensions, we will theorize and develop a hybrid alternative to the Mix/FixMatch algorithms to see if there is an effective way to combine the best of both scenarios. If successful, it will be

subjected to the same experiments and compared with all results. An additional experiment to compare the different algorithms we will conduct as an extension if time and resources allow are ablation studies, mentioned in both *ixMatch papers [1][2] to test how a network performs when parts of it are removed to investigate which layers, neurons, and parameters are most important to the process.

Ethical Implications: There are no direct ethical implications for this project.

Resources and related work: Although deep neural networks have been shown to be extremely powerful, there exist a few weaknesses such as overfitting the training data [1] and the need for massive labeled training datasets, which can be costly [2]. Overfitting is largely problematic for two reasons: the model is unable to generalize its findings to slightly different data and an increased sensitivity to adversarial examples. In the past, attempts to reduce the issue of overfitting generally required experts to generate augmented data in the vicinity of real training data entries. By including these augmented entries in training, the generalizability of the model could be increased, but this method required expert knowledge and was normally dataset dependent. Zhang et al [3] tried to address these issues with mixup, a data-agnostic augmentation approach. mixup employs a simple linear combination of two randomly selected training data points to generate augmented data points in order to expand the overall training dataset. With mixup, the generalization error was improved for image, speech, and tabular data compared to other state-of-the-art methods at the time.
The next major breakthrough in overcoming these issues came from Berthelot et al [1] with MixMatch. MixMatch utilizes a modified version of mixup that approaches the problem with semi-supervised learning, generating augmented examples for batches of labeled and unlabeled training data with predicted low-entropy labels, reducing the need for massive labeled datasets. The authors showed that MixMatch had a 4-fold improvement in error rate to other methods and successfully improved unlabeled prediction confidence, while increasing generalizability and robustness. The most current method in the semi-supervised learning space is FixMatch by Sohn et al [2]. FixMatch addresses the aforementioned problems with a pseudo-labeling approach. First, the model predicts the labels of unlabeled examples on a weakly-augmented version of the image. Then, the model tries to predict those given labels on a strongly-augmented version of the same images, which is used to calculate the unsupervised loss of the model. The authors show FixMatch performs better than other existing methods, while being a much simpler approach. Overall, all of these methods have explored how we can utilize labeled and unlabeled data to improve the generalizability of our models, improve robustness against adversarial examples, and reduce the need for massive labeled-datasets.

[1] “MixMatch: A Holistic Approach to Semi-Supervised Learning”, Berthelot et al.
[2] “FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence”, Sohn et al.
[3] “mixup: Beyond Empirical Risk Minimization”, Zhang et al.

Datasets:

- Images

        - CIFAR-10 and CIFAR-100
            - https://www.cs.toronto.edu/~kriz/cifar.html
        - SVHN
            - http://ufldl.stanford.edu/housenumbers/
        - STL-10
            - https://cs.stanford.edu/~acoates/stl10/
- Speech Recognition
    - https://www.tensorflow.org/datasets/catalog/speech_commands

Team Members: John Dugan, Caleb Goertel, Andrew Price, Vimal Venugopal