Computer vision project : Semi-supervised Image Classification

Devikalyan Das, Rushan Mukherjee
7007352, 7015520

March 11, 2022

Abstract

Deep learning has shown promising results in the domains of computer vision. But this can be possible when there is large volume of labelled data. When there is large amount of unlabelled samples, Semi-Supervised Learning (SSL) provides an effective measure to improves the models performance by leveraging information not only from labelled data but also from large amount of unlabelled data. In this project, we have implemented some of the SSL techniques for image classification. As part of the project, we have implemented Pseudo Labeling as first task, which is a standard SSL algorithm, followed by SSL using Virtual Adversarial Training (VAT) Loss as second task which is a recent technique compared to the first one. Then as part of the third task, we have implemented a state of the art technique called FixMatch as a benchmark. Then in-addition to the FixMatch, we have tried a technique to improve the performance as compared to the benchmark. We have proposed an enhanced model and a novel combined loss function that further acts on the similarity between differently augmented samples.

1 Introduction

The success of the deep learning techniques in various domains of Computer Vision have been attributed to the scalability and ability to perform better with large number of labelled dataset through supervised learning. But acquiring large labelled data is a costly affair especially in domains such as medical applications. A powerful approach to address this is Semi-Supervised learning which mitigates the requirement of labelled data by leveraging unlabelled data, meaning it learns from both labelled and unlabelled data.

In this project, We were tasked with implementing some common techniques of Semi Supervised Learning (SSL) for the task of semi-supervised image classification. Many SSL methods work on the principle of producing an artificial label for the unlabelled image data and perform the training of the model to predict the artificial label called pseudo label when the model is fed unlabelled images.

Our given task was to implement a standard SSL approach, followed by a state of the art method and finally a novel technique that improves on one of the existing algorithms. In this report, we first discuss the Pseudo-Labelling techniqueLee [2013] which firstly aims to generate pseudo-labels for unlabelled samples and then train the model to predict artificial labels from the unlabelled samples along with the model prediction on labelled data using a combined Loss Function is used which takes into account both the loss from the labelled data and the pseudo loss obtained above. The generation of pseudo-labels is based on confidence thresholds as these labels can be wrong.

Further, we move on to Virtual Adversarial TrainingMiyato et al. [2018] which is a regularization approach based on Virtual Adversarial Loss. The idea is explained as to add a minuscule amount of noise to our training images that are imperceptible. This method as mentioned by the authors is based on the adversarial training methodGoodfellow et al. [2015]. Adversarial Training is best suited for supervised approaches. To implement VAT algorithm we find a type of perturbation between the original and the noisy or adversarial image such that the KL Divergence is the maximum.

Lastly, for our challenge task we have tried to improve the model performance of FixMatch Semi-Supervised Learning algorithmSohn et al. [2020]. Our FixMatch algorithm was trained on a WideRes-

NetZagoruyko and Komodakis [2017] model. For the CIFAR-10 dataset we used a model depth and width of 28 and 2 respectively. On the other hand, for CIFAR-100 dataset we have a model depth of 28 and a model width of 8. Here, we used a Wideresnet model with larger filter kernels for efficient capturing of features and a combined novel loss function that forces similarity between the labels of different augmentation.

2 DataSet Informtion

Our task is to implement semi-supervised learning techniques on two specific datasets namely CIFAR-10 and CIFAR-100. Both of the above given datasets have 60000 images. 50000 of them are training images and the rest 10000 are testing images. CIFAR-10 has 10 classes where as CIFAR-100 has 100 classes under which labels exist in the dataset. With regard to CIFAR-10, we are using two cases with 250 labelled and 4000 labelled samples per class. Similarly, we have taken 2500 labelled and 10000 labelled samples per class for CIFAR-100 dataset.

3 Task 1: Pseudo-Labelling

3.1 A Brief Description

Pseudo-labelling is one of the standard semi-supervised learning algorithms that are in use today. Pseudo-labelling simultaneously trains a model in supervised approach with both unlabelled and labelled information.

In every epoch, we use the model's class prediction on unlabelled data as a label to train against. This is done by picking the class having the predicted probability as maximum which the model decides. For this we use the three given confidence thresholds namely 0.6, 0.75 and 0.95. During training, we use a loss which is combined of labelled loss and pseudo loss.

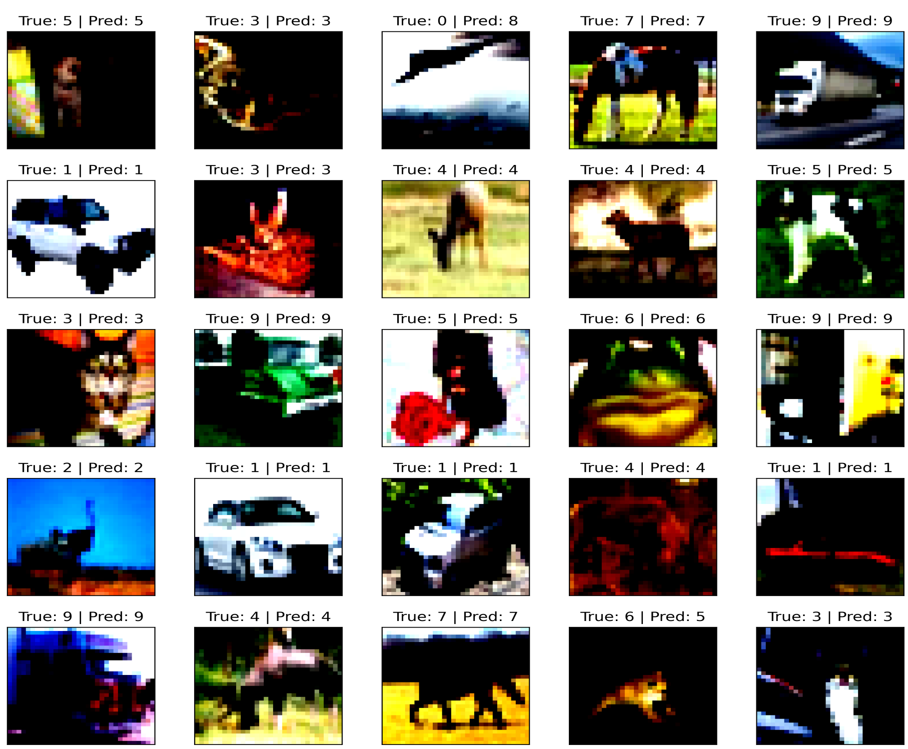

Figure 1: Test Samples for CIFAR-10 with 4000 samples

3.2 Model Parameters Used

The following model parameters were used to train the model for 0.6, 0.75 and 0.95 confidence thresholds. They are arranged in a tabular format as follows:

Parameter	Values
Model Depth	28
Model Width	2
Learning rate	0.03
Epochs Trained	100
Iterations per Epoch	1024
Batch Size(for both training and testing)	64

3.3 Experiment Results

We have trained our model using the given 3 confidence thresholds namely $0.95, 0.75, 0.6$. The observed error rates for these confidence thresholds for both CIFAR-10 and CIFAR-100 datasets along with their varying labelled samples are provided in below table. Some of our model predictions for CIFAR-10 with 4000 samples are provided in Figure number 1.

Dataset	CIFAR-10						CIFAR-100					
Samples	250			4000			2500			10000		
Thresholds	0.6	0.75	0.95	0.6	0.75	0.95	0.6	0.75	0.95	0.6	0.75	0.95
Error(percent)	73.45	71.67	66.71	20.48	20.43	19.37	70.86	68.47	72.58	46.33	46.11	45.08

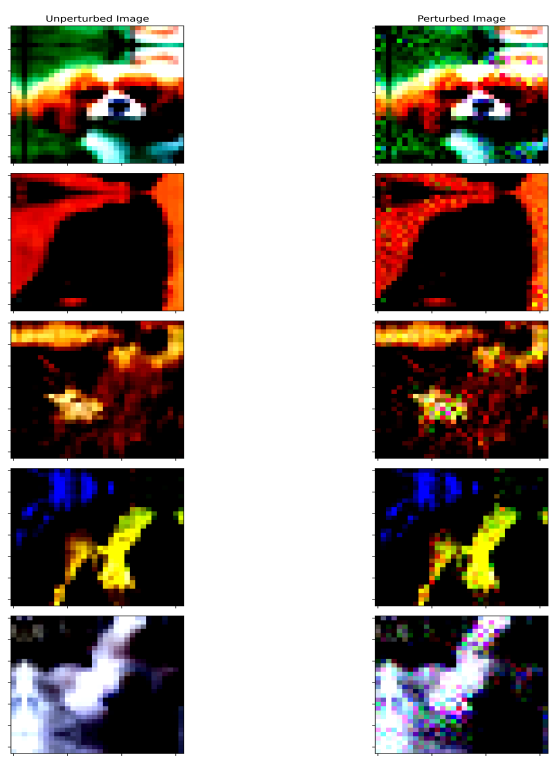

Figure 2: Unperturbed and Perturbed Images (computed as per the defined parameters)

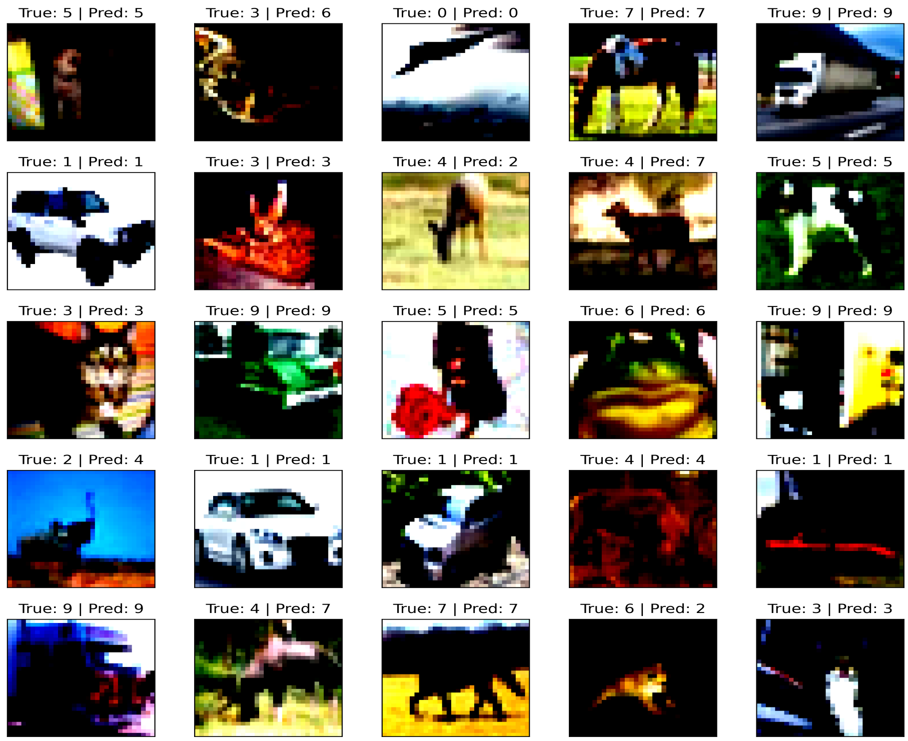

Figure 3: Test Samples for Task 2 for CIFAR-10 with 4000 Samples

4 Task 2: Virtual Adversarial Training

4.1 A Brief Description

The main goal for Task 2 is to implement VAT Loss. The VAT loss was implemented as per algorithm 3 and the training was as per algorithm 4 as per the provided instruction sheet. With the help of the VAT Loss we generate our perturbed samples from the unlabelled data and further used this Loss along with Cross Entropy Loss on labelled samples during the training phase. Next in line, we use Stochastic Gradient Loss for optimization. Figure 2 perfectly demonstrates the perturbed images that we have attempted in our project. We used the parameters as given in the below table for their generation.

4.2 Model Parameters Used

We trained our model for 250 and 400 samples for CIFAR-10 and 2500 and 10000 samples for CIFAR-100. The parameters that were used to train the VAT model are given below in a tabular format. They are:

Parameters	Values
EPS	10.0
xi	10.0
Iteration parameter for VAT Loss(iter)	5
Model Depth	28
Model Width	2
Learning rate(lr)	0.03
Epochs Trained	512
Iterations per Epoch	1024
Batch Size(for both training and testing)	64

4.3 Experiment Results

Some sample model predictions are provided in figure 3. Similarly, we describe the error rates observed with CIFAR-10 and CIFAR-100 datasets with the given specific number of samples. The values are in tabular form as follows:

Dataset	Number of Samples	Error(in Percent)
CIFAR-10	250	53.9401
	4000	20.1334
CIFAR-100	2500	74.3233
	10000	56.9368

5 Task 3: Challenge Task

We have attempted to improve some aspects of FixMatch which is a simple and state-of-the-art algorithm and is our benchmark for our challenge task. Our major areas of improvement

- We increased the kernel size of the Convolution Neural Network(CNN) kernels to 5X5 instead of the default 3X3. This enabled us to capture certain features missed by the smaller size kernels.
- We used a novel loss function which is linear combination of Cross Entropy Loss, Pseudo Label Loss and Cosine Similarity Loss functions. We used Cosine Similarity Loss because it tries to maximize the similarity between the predicted label for the strongly augmented samples and the pseudo label from the weakly augmented samples in order to achieve a much better loss minimization.

The Cumulative Loss function that we used can be visualized in a equation as below:

$$\textbf{Combined Loss} = \text{Cross-Entropy Loss}_L + \lambda_{UL} * \text{Cross-Entropy Loss}_{UL} + \rho_{CS} * \text{Cosine-Similarity Loss}_{UL}$$

where L is Labelled data , UL is unlabelled data and CS is the Cosine similarity parameter.

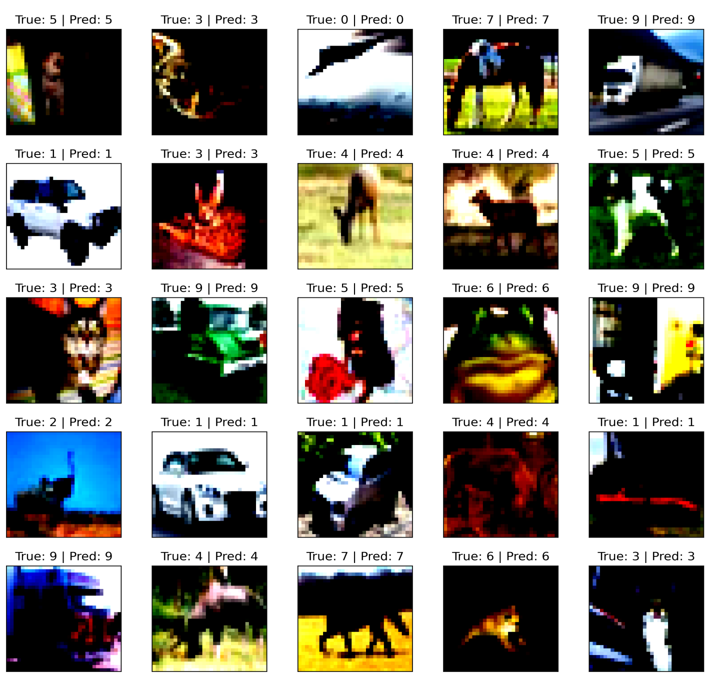

Figure 4: Test Samples for Task 3 for CIFAR-10 with 4000 Samples

5.1 Data Preparation

FixMatch uses extensive data augmentation for generating weakly augmented and strongly augmented labelled images/ samples. These are then used for training our models. To generate strongly augmented samples, we have taken into consideration random variations such as Auto-Contrast, Brightness, Colour, Contrast, Equalize, Identity, Posterize, Rotate and Sharpen. In addition we have also used ShearX, ShearY, Solarize, TranslateX, TranslateY, Horizontal Flip and Random Crop.

5.2 Model Parameters

We trained our model using a specific set of parameters as shown below:

Parameter	Values
ρ_{cs}	0.5
λ_{UL}	1.0
Model Depth	28
Model Width	2(For CIFAR-10) and 8(For CIFAR-100)
Learning rate	0.03
Epochs Trained	100
Iterations per Epoch	1024
Batch Size(for both training and testing)	64
Threshold	0.95

5.3 Experimental Results and Discussions

Here, we display a side by side comparison of FixMatch Semi-Supervised Learning algorithm and our new improved approach. The accuracy score of the 2 afformentioned models are given below as follows:

Dataset	CIFAR-10				CIFAR-100			
Sample Size	250		4000		2500		10000	
Accuracy	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
FixMatch	86.704	99.532	92.128	99.582	58.42	81.081	69.009	89.49
Our Model	86.499	99.124	**92.227**	**99.701**	53.553	77.747	66.869	88.037

6 Conclusion

The domain of SSL is rapidly progressing but these come at a cost of complicated algorithms with loss function. We have implemented a rather simple technique that challenges the state-of-the-art technique, but a lot of different techniques can be used such as attention mechanism, pre-trained models, novel loss functions and different optimizers could be used as future experiments.

References

Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples, 2015.

Dong-Hyun Lee. Pseudo-label : The simple and efficient semi-supervised learning method for deep neural networks. 2013.

Takeru Miyato, Shin ichi Maeda, Masanori Koyama, and Shin Ishii. Virtual adversarial training: A regularization method for supervised and semi-supervised learning, 2018.

Kihyuk Sohn, David Berthelot, Chun-Liang Li, Zizhao Zhang, Nicholas Carlini, Ekin D. Cubuk, Alex Kurakin, Han Zhang, and Colin Raffel. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *CoRR*, abs/2001.07685, 2020. URL `https://arxiv.org/abs/2001.07685`.

Sergey Zagoruyko and Nikos Komodakis. Wide residual networks, 2017.