Summary Report - Deep Learning CS F425

(Under the guidance of Professor Tirtharaj Dash)

Question - Supervised Learning, Self-Supervised learning (Barlow Twins Method), and Semi-Supervised learning on the STL-10 dataset.

Team Members -	Shrey Pandit	2019A7PS0138G
	Shreyas Bhat	2019B1A80969G (Auditing course)
	Jatin Singh	2019A7PS0085G
	Rachit Jain	2019A7PS0140G

This report is supported along with the code - https://colab.research.google.com/drive/18FrJs_Eox1E_uzWwX1dEdZ1Gh6fsYbwW?usp=sharing
And is done as part of the completion requirement for the course.

*(Code is well documented and is easy to understand)

The framework used for coding - **PyTorch**

About the Dataset -

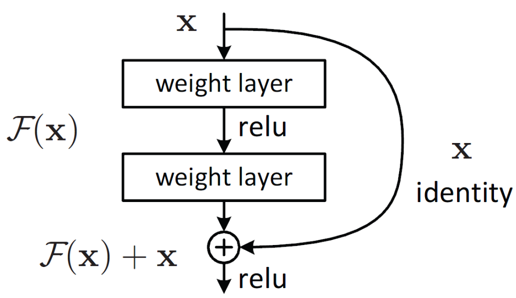
STL-10 - https://cs.stanford.edu/~acoates/stl10/

Split	Content	Number of samples
Train	Image, Label	5000
Test	Image, Label	8000
Unlabeled	Image	100,000

Image dimension (96x96x3) - RGB Images

1 - Training a supervised model -

We trained a Resnet Architecture, with 9 Conv blocks, having dropouts and batch norm. This model was defined and trained from scratch.



We trained it using Adam optimizer, LR of 0.001, Loss being cross-entropy, trained on GPU for 10 epochs.

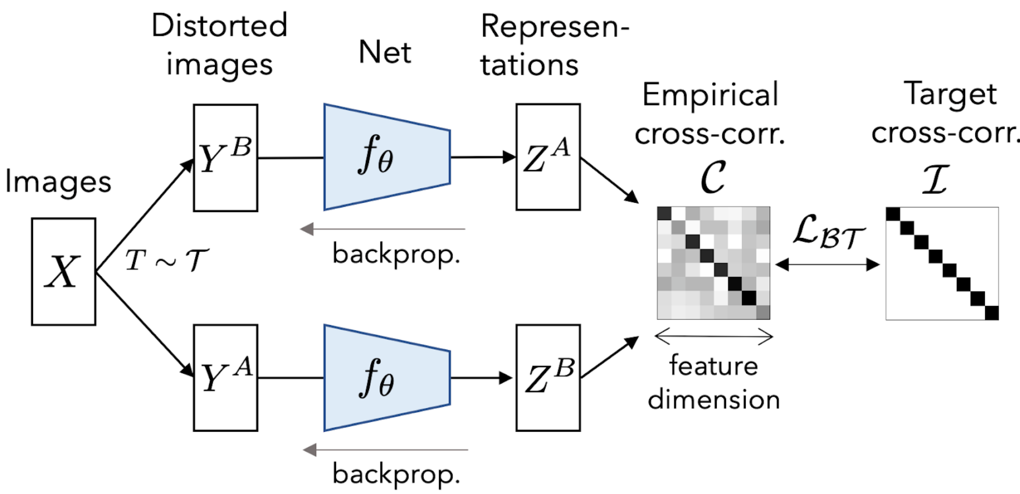
Evaluation metric being Accuracy, F1 score.

Metric	Score
Accuracy	59.2%
F1	71.42%

Note- This accuracy and F1 score could be easily increased by making the ResNet block deeper, (Something like ResNet 101) but that would not entirely justify the purpose of accuracy gain using the Semi-supervised model, done later. We also felt that using a relatively simpler model is much more practically feasible in case of any application to be made instead of super deep ResNet blocks.

2 - Training a Barlow twins model -

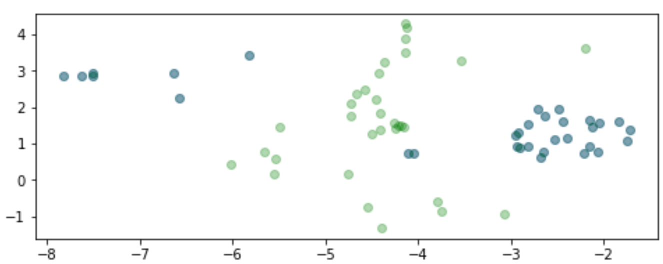
We used the same architecture of Resnet without the final fully connected layer, we tried the exact same configuration as mentioned in the original paper **Barlow Twins: Self-Supervised Learning via Redundancy Reduction** by Zbontar et. al. We used the fully connected layer of the decoder to be of the same size as 8192 for both the ResNet blocks and used the lambda as specified in the paper to be 5e-3.

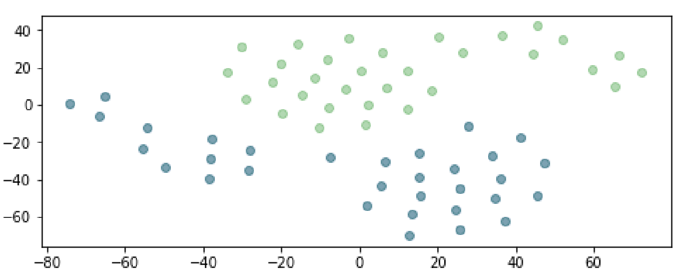


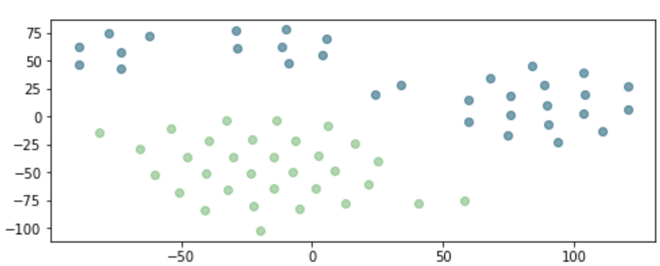
We trained it using Adam optimizer, LR of 0.001, Loss being cross-entropy, trained on GPU for 10 epochs.

We also plot the embeddings as they progress to change throughout the course of training -

Initial Epochs -

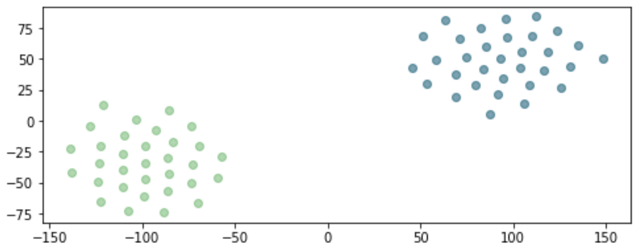






We can see how the initial embeddings were mixed up, and slowly start to separate as we train.

Final Epoch -



We can see that the embeddings are separating from each other due to the loss function penalizing them using the cross-correlation matrix.
Testing Loss also drops from around 2112.35 in the beginning to around 42.46 in the final epoch showing the training of the model is in the proper direction.

3 - Stacking a linear classifier over the Trained Barlow twins model weights

We train a linear classifier over the frozen weights of the Barlow Twins model, fine-tune it on the 5000 training samples, we find a training accuracy and F1 score of 37.21% and 39.55%. This lower accuracy could be justified by the fact that we trained the Barlow twins model for lesser epochs as per computational feasibility and Self-supervised approaches require much higher training time for properly implementing the contrastive loss-like method.

4 - Training Semi-supervised using Pseudo Labeling -

We also tried the pseudo labeling technique as a part of semi-supervised learning using a mix of labeled and unlabeled samples. For the initial training, we tried it on **50,000 unlabeled samples**, using the frozen weights of the model trained in part 1 (weights - https://drive.google.com/file/d/1-1zJpQBKEYUv-DEHiBFHjDVgLwmGiRLQ/view?usp=sharing for replicability), with a threshold value of **99% confidence** on the prediction to be taken as a sample point.

We append these data points with the original dataset and train the model to find the newly improved results.

Method	Unlabeled samples	Training + Unlabeled after 99% threshold	Accuracy	F1 Score
Supervised	0	5000	59.2%	71.42%
Self-supervised	50,000	5000+ 21,529 (after 99% threshold removal)	63.22%	73.95%

5 - Bonus - Trying out the different ratios of labeled and unlabeled samples -

We try out 4 different combinations of labeled and unlabeled splits for semi-supervised techniques.

Supervised	Self-supervised	Supervised: Self-supervised	Acc / F1
5000	0	1:0 (original supervised)	59.2% / 71.42%
5000	2000	1:0.4	60.21% / 71.92%
5000	5000	1:1	61.33% / 72.27%
5000	10,000	1:2	61.72% / 73.12%
5000	50,000	1:10	63.22% / 73.95%

Observation - We can observe from the above results that the performance of the model is directly proportional to the number of training samples we have. So we can say that increase in the number of training samples using the unlabeled dataset using the semi-supervised learning method results in an increase in model performance.

Contributions -

Shrey Pandit - Supervised Learning, Barlow Twins, Pseudo Labeling, Bonus part, Report
Shreyas Bhat - Barlow Twins, Pseudo Labeling
Jatin Singh - Report, Bonus part
Rachit Jain - Report, Bonus part