Signal Processing in Practice
Assignment 08 Report

SAPTARSHI MANDAL — SR. No. 22925

March 2024

1 Image Classification

Initially, a ResNet18 model is trained using the *trainset1* subset and was subsequently evaluated on the CIFAR-10 dataset. Following this, another model was trained using the *trainset2* subset and subsequently assessed using the CIFAR-100 dataset.

Displayed below are several images representing various classes within the CIFAR-100 dataset. Both *trainset1* and *trainset2* were sourced from the CIFAR-100 dataset.

Notably, *trainset1* comprises 5000 images, while *trainset2* comprises a larger collection of 25000 images.

The Test Accuracies are given in Table.1

	Accuracy (CIFAR-100 test set)
trainset1	24.46%
trainset2	38.92%

Table 1: Accuracy VS Trainset

Observations and Possible Reasons:

1. Performance Discrepancy: There's a notable difference in the performance of models trained on *trainset1* and *trainset2*. The model trained on *trainset2* achieves a significantly higher accuracy (39.03%) compared to the one trained on *trainset1* (22.86%). This suggests that the larger training set (*trainset2*) provides the model with more diverse and representative data, leading to better generalization performance.

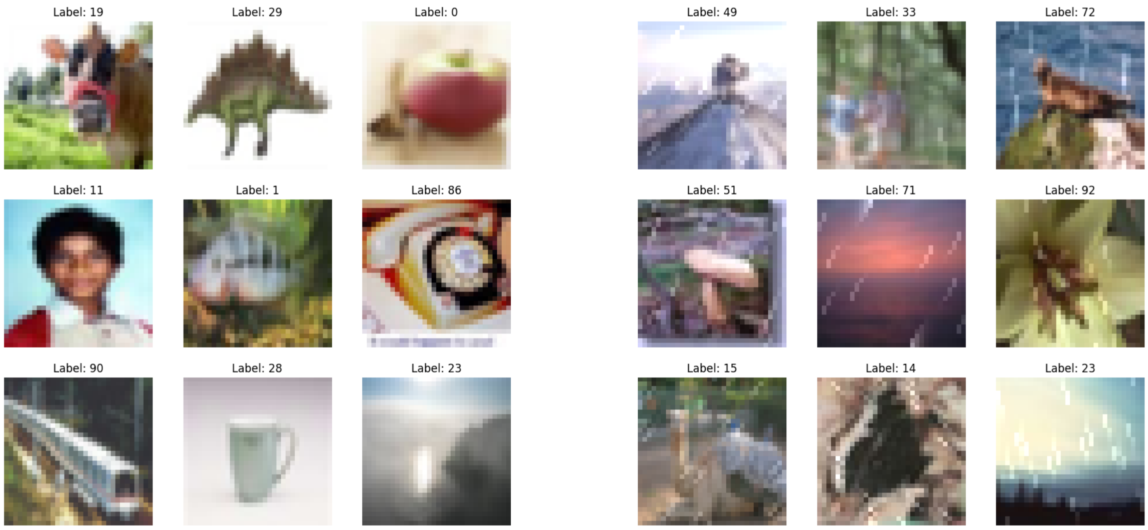
2. Impact of Dataset Size: The size of the training dataset seems to have a considerable impact on model performance. With *trainset2* being substantially larger (25000 images) than *trainset1* (5000 images), the model trained on the former has more data to learn from, resulting in improved performance on the CIFAR-100 test set.

2 Robustness against distribution shift

In this section evaluation is done on an additional test set, using the model in two modes, *train* and *eval* mode while varying the test batch size as 4, 8, 16, 32. The results are given in Table. 3 and Table. 2

This Additional Test dataset contains images which are from slightly different distribution than the images in CIFAR-100 Train and Test Datasets.

The difference in datasets are shown in the Figure.1



(a) CIFAR-100 Dataset (b) Additional Dataset

Figure 1: Comparison of CIFAR-100 Dataset (a) and Additional Dataset (b)

Batch Size	Trained on Trainset1	Trained on Trainset2
4	1.08%	1.10%
8	1.18%	1.30%
16	1.73%	1.86%
32	2.05%	2.10%

Table 2: Test Accuracies (Eval Mode)

Batch Size	Trained on Trainset1	Trained on Trainset2
4	6.69%	7.29%
8	12.18%	16.61%
16	15.79%	26.80%
32	19.81%	29.83%

Table 3: Test Accuracies (Train Mode)

Observations and possible explanations for the results:

- **Evaluation on Additional Test Set:**
  - Model accuracy ranges from 1.08% to 2.10% in eval mode and 6.69% to 19.81% in train mode, indicating poor generalization from CIFAR-100 to the additional test set.
  - Differences between the additional dataset and CIFAR-100, such as variations in lighting and object compositions, likely contribute to the observed lower accuracy.
- **Effect of Training Set:**
  - There is a noticeable difference in performance between models trained on Trainset1 and Trainset2. Models trained on Trainset2 generally achieve higher accuracy in both eval and train modes.
  - This suggests that the additional training data in Trainset2 leads to better generalization and higher performance on the additional test set.
  - The difference in performance between Trainset1 and Trainset2 highlights the importance of dataset size and diversity in training deep learning models.
- ***Train* and *Eval* Mode Effects:**
  - In *Eval* mode, the model uses batch normalization parameters learned during training, which may not be suitable for batches from the additional dataset due to its slightly different distribution. As a result, both models trained on *Trainset1* and *Trainset2* exhibit poor performance.
  - In contrast, *Train* mode allows the model to learn batch normalization parameters during evaluation, leading to better performance compared to *Eval* mode.
- **Batch Size Effects:**
  - Increasing the batch size can have implications for batch normalization. With larger batch sizes, the statistics computed for normalization (mean and variance) become more accurate representations of the entire dataset, leading to more effective normalization.

3 Self-Supervised Learning

The following steps are taken for performing Self-Supervised learning on *trainset2*:

1. A ResNet-18 model is pretrained for the CIFAR-100 dataset using trainset2 (without labels) in a self-supervised manner using the SimCLR objective.

2. For each batch of N images and its augmented version of N images resulting in $2N$ data points, where each image has one positive pair and $2(N-1)$ negative pairs, the loss function for a positive pair of examples is defined as:

$$l(z_i, \tilde{z}_i) = \log \frac{\exp(\text{sim}(z_i, \tilde{z}_i)/\tau)}{\sum_{k=1}^{N} 1[k \neq i] \exp(\text{sim}(z_i, z_k)/\tau) + \sum_{k=1}^{N} 1[k \neq i] \exp(\text{sim}(z_i, \tilde{z}_k)/\tau)}$$

3. The SimCLR loss on the batch can be obtained as:

$$L = \frac{1}{2N} \sum_{k=1}^{N} l(z_k, \tilde{z}_k) + l(\tilde{z}_k, z_k)$$

Some of the Augmented images are shown in Figure.2

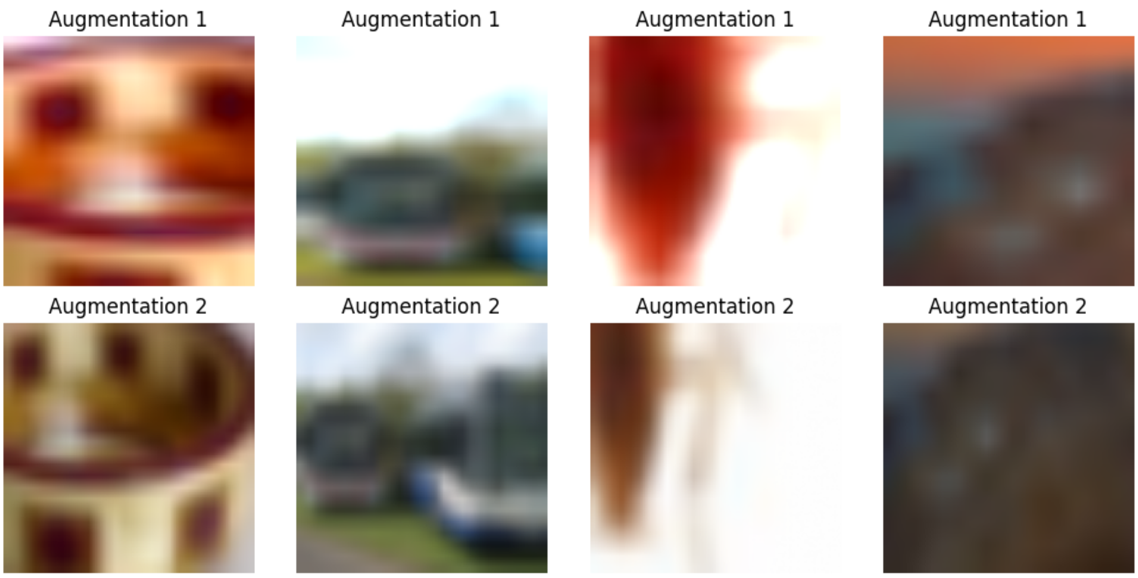


Figure 2: Augmented Images

Then the following is done: **Fine-tuning and Evaluation:**

1. The pretrained model is fine-tuned using trainset1 with Cross Entropy loss.

2. The accuracy of the finetuned model is evaluated on the CIFAR-100 test set.

The accuracy observed on the CIFAR-100 test set is given in Table. 4
Observations and Explanations

	Accuracy (CIFAR-100 test set)
trainset1	30.00%

Table 4: Accuracy By Using SimCLR and Fine Tuning

- **Image Classification Accuracy on 5000 Images**: 24.46%
- **Self-Supervised Learning on *trainset2* (25000 Images), Followed by Fine-Tuning with *trainset1* (5000 Images)**: Accuracy improved to 30%.

The increase in accuracy from 24.46% to 30% after performing self-supervised learning followed by fine-tuning can be attributed to several factors:

1. **Data Abundance**: Self-supervised learning utilizes a larger dataset of 25000 images, providing the model with more diverse and representative samples. This helps the model learn richer and more generalized features, leading to improved performance during fine-tuning.
2. **Feature Learning**: During self-supervised learning, the model learns meaningful representations or features from the unlabeled data. These learned features capture high-level semantics and underlying patterns in the data, making them more useful for the downstream classification task.
3. **Domain Adaptation**: Fine-tuning the model on the smaller dataset of 5000 images allows it to adapt its learned representations to the specific characteristics of the target dataset. This process helps the model focus on relevant features and refine its predictions for better performance on the target task.

In summary, the improvement in accuracy from 24.46% to 30% demonstrates the effectiveness of self-supervised learning followed by fine-tuning for enhancing the performance of image classification models, especially when limited labeled data is available.