Université Paris Cité

TP Self-Supervised Learning Tasks

Yacine FLICI M2 VMI FA

1 Experimental Setup

Experiments were done in Python and PyTorch. The dataset CIFAR-10 was primarily used for training and evaluating the models, and other datasets like Pascal VOC and FashionMNIST. **Source Code**

2 Pretext Tasks Overview

The following self-supervised learning tasks were tested. These representations can later be used in downstream tasks such as classification and segmentation.

- **Rotation Prediction**
- **Context Encoder)**
- **SimCLR (Contrastive Prediction)**

Each of these pretext tasks enables the model to learn features that can later be used in downstream tasks such as:

- **Classification**: Using the learned representations, a simple classifier can be trained on top to categorize images into different classes.
- **Segmentation**: Representations learned through tasks like Context Prediction and SimCLR provide spatial and structural information beneficial for segmenting images into meaningful parts.

3 Model Evaluation on CIFAR-10

After loading each model on the respective pretext tasks, we evaluated their performance on a downstream classification task using CIFAR-10. A linear classifier was applied to frozen representations learned by each model and recorded the accuracy.

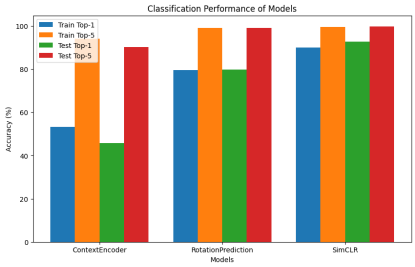

Figure 1: Comparaison of Classification for different pretext on classification

3.1 Modifications to the DeepUL Library

To add the FashionMNIST dataset into the ‘deepul‘ library, we made several adjustments to the dataset loading functions and pipeline configurations to ensure compatibility with this dataset. Specifically, we:

- Updated the dataset loader to download and preprocess FashionMNIST images, converting them to a compatible format for each pretext task.
- Modified transformation functions to handle the 28x28 dimensions of FashionMNIST, resizing images as needed for each specific task.

These modifications allowed us to effectively evaluate trained models on CIFAR-10, testing the generalization of representations learned from different pretext tasks. We found that the **Context Prediction** and **SimCLR** tasks transferred but did not perform well to FashionMNIST, while other tasks could not be directly transfer.

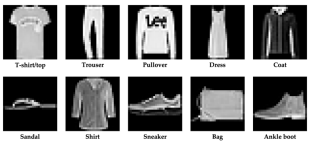

Figure 2: FashionMNIST images

3.2 Evaluation Results on FashionMNIST

The two pretext tasks, Context Prediction and SimCLR, demonstrated weak results on the downstream classification task with FashionMNIST. The following bar chart shows the performance of these tasks in terms of classification accuracy on FashionMNIST.

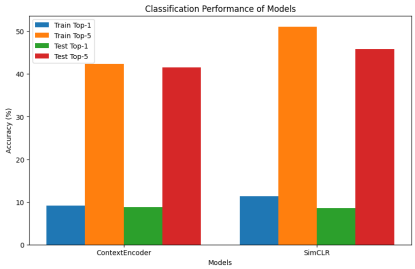

Figure 3: Classification accuracy on FashionMNIST

The results indicate that **SimCLR** performed slightly better than **Context Prediction**, suggesting that contrastive learning techniques like SimCLR can be effective for learning transferable representations on FashionMNIST.It also suggests that datasets should be at least similar to have transferable learning

4 Implementation Details

For training the new pretext task **Context Prediction** implemented the following main steps in the code:

1. **Dataset Preparation**: Created new Dataset class that devides the image into 9 patches and then chooses the central patch and another random patch to be used along with its position label

2. **Model Training**: provided a simpler code in the Jupter notebook to train the pretext ask on the dataset Pascal VOC. Chose this dataset because its images are big enough to be separted into patches

3. **Evaluation**: After training, each model was evaluated by training a linear classifier on the learned representations and testing on the downstream classification task.

5 Training

My training did not go so well, problem with small data, it resulted in non convergence. Solution would be to augment the data

(a) Original Image

(b) Central Patch

(c) Random Patch label number 0

Figure 4: Visualization of the Original Image, Central Patch, and Random Patch used in the Context Prediction task Dataset Pascal VOC.

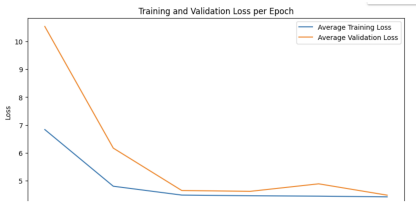

Figure 5: Classification accuracy on FashionMNIST

6 Conclusion

This project demonstrates the effectiveness of self-supervised learning for image representation learning. Each pretext task provided spatial or structural information to the model, improving its performance on downstream tasks.