Lab Ass. 2.2: On the effect of rotation-based self-supervised learning

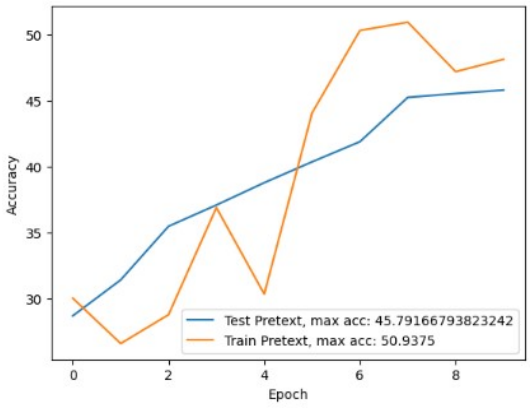
Alberto Hernández Ramos

***Abstract*— The objective of this lab assignment is to present the student with the transfer learning and image data augmentation techniques in the context an image classification task.**

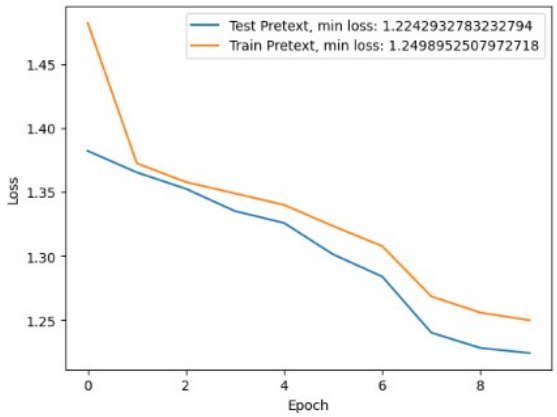
I. Training the pretext task

A. Graphical representations

In this experiment, training was conducted for the pretext task, which involved predicting the rotation of images. Regarding the results obtained, it has been observed that the maximum accuracy on the test set is 50.93%, compared to 45.79% on the training



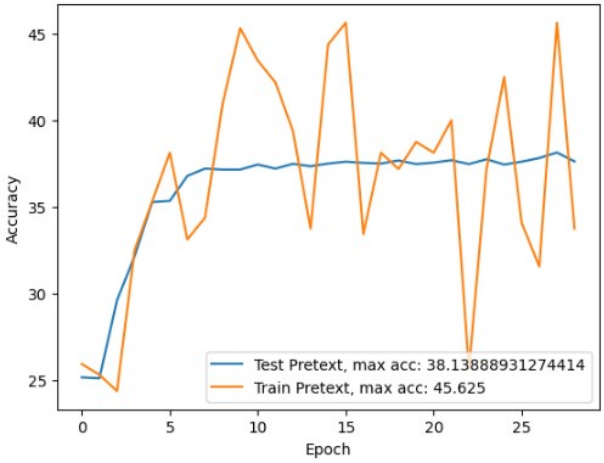
Regarding the obtained results, it was observed that the minimum loss in the test set is 1.2498 compared to 1.2242 in the training set.

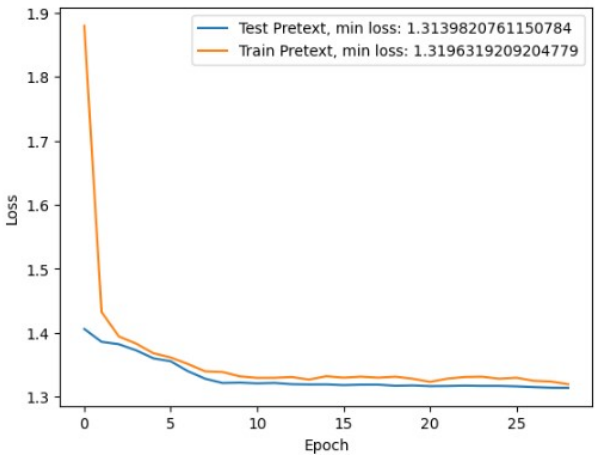


B. Discuss on the results

The results obtained in this experiment are far from ideal; therefore, I believe that several hyperparameter modifications could be made to try to improve the model. First, I think training with 10 epochs is too little, as can be seen in the previous graphs where the curves do not stabilize. I have changed it to 30 epochs. I have also modified the number of blocks in the Resnet model; now it is Resnet34, meaning num_blocks = [3,4,6,3]. This implies that there are 3 BasicBlocks of 64 channels, 4 of 128, 6 of 256, and 3 of 514.

After conducting the training of the new model, the results are quite surprising. The accuracy and loss metrics are significantly worse than in the previous case. This could be because I might be training a model that is too complex using too few data, assuming that the implementation of the model and other aspects has been done correctly.



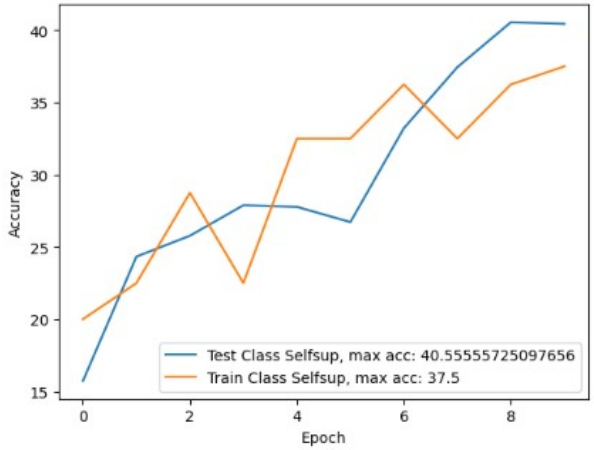


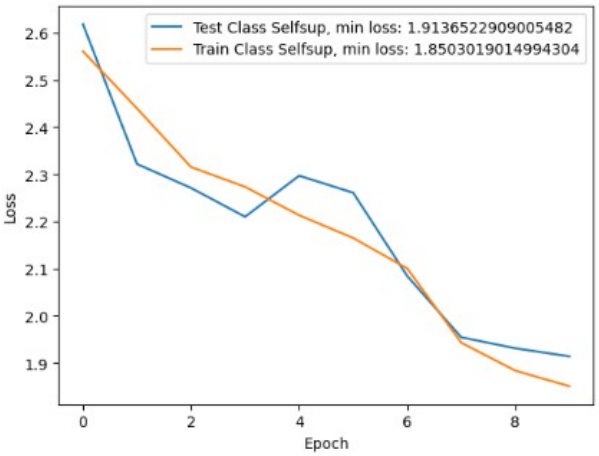
II. Training the classification task

For this section, we will classify a set of images. To do this, two approaches will be tested: first, training based on the weights calculated in the previous section using the pretext task (selfsup), or training from scratch.

A. *Selfsup*

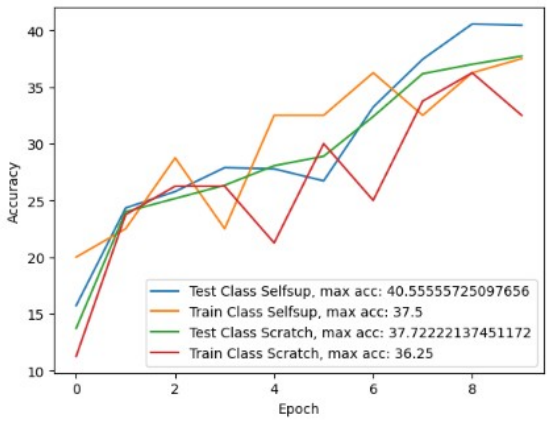
For this section, a maximum accuracy of 40.55% and 37.5% has been achieved on the training and validation sets, respectively, using the model and default hyperparameters from the lab assignment.



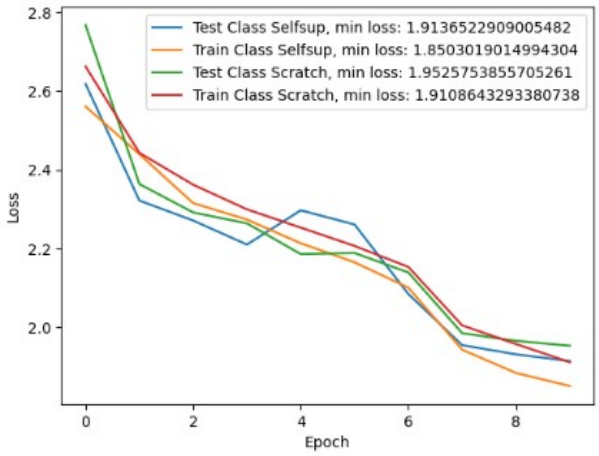


B. *Result Comparison*

Regarding the results of all the experiments conducted, it is observed that the best outcomes for the test sets are achieved using the SelfSup approach, reaching 40.55%, while for the case of training from scratch, a performance of 37.72% is attained, indicating a 3% difference in accuracy.



As for the loss graphs, they also favor the SelfSup approach, with a minimum loss of 1.911 compared to 1.952 for training from scratch

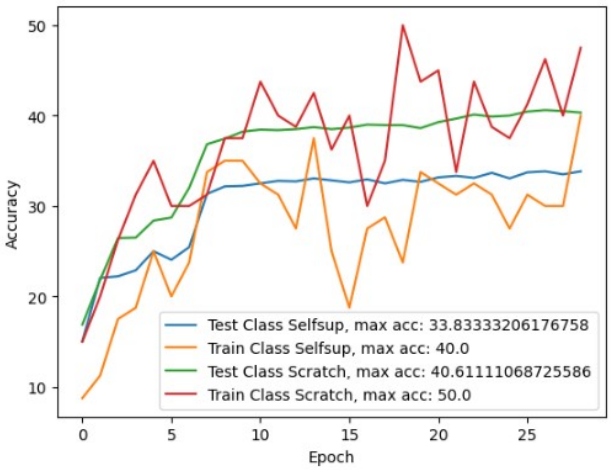


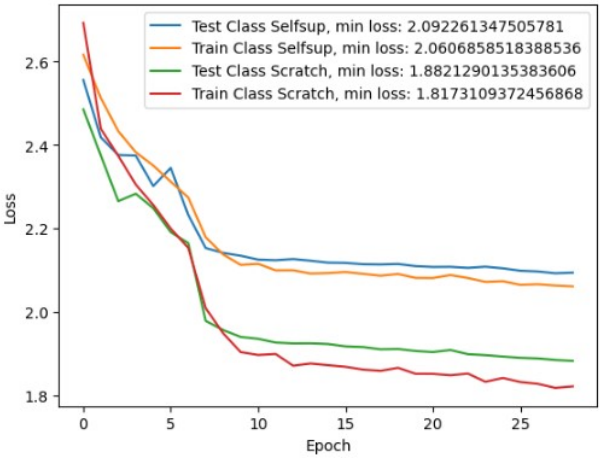
C. *Discuss the result*

Although the results are better for the SelfSup case, I believe the difference is not too significant, just a 3% margin. This is because the results of the pretext task are quite poor, so the SelfSup model is not benefiting from the transfer of learning from the pretext task (it transfers a very poor representation).

D. *Discuss the potential benefits of using different hyperparameters*

To discuss these results, I have decided to try training with the previously described setup (modifying the number of epochs to 30 and using ResNet34). In this case, the results are better when training from scratch than when transferring learning from the pretext task. This contradicts the self-supervised learning approach and demonstrates that the result from the previous section is not indicative of transfer learning functioning but rather a matter of chance.



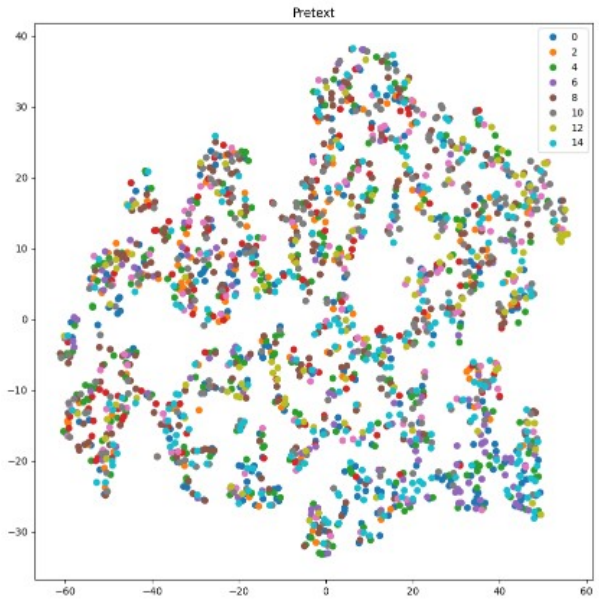


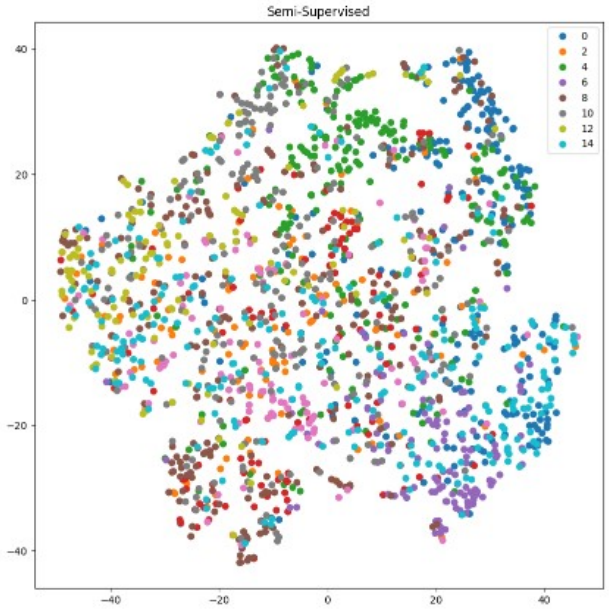
III. Model visualizations

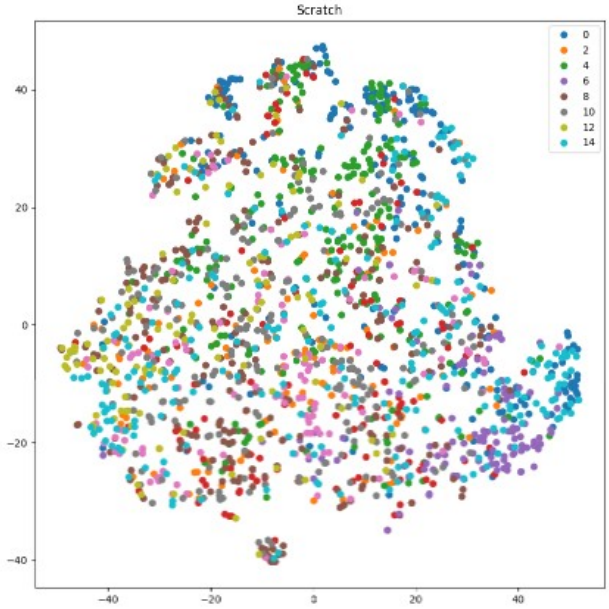
In this section, the T-SNE representations of the three trained models are displayed.

In all cases, it can be observed that separate regions based on their classes are not distinguished. This makes sense, given that, as shown in the previous results, the models perform poorly, and the representations obtained by the CNNs are quite inadequate.

In theory, if there were quality transfer learning from the pretext-task model to the classification model, a difference would exist between the model trained using the weights and the model trained from scratch.







IV. Conclusion

As a conclusion, I would say that the results have been quite poor overall, demonstrating that there is no transfer learning between models. Therefore, I am inclined to think that the implementation of the practice may not have been correct.