IE643: Deep Learning - Theory and Practice **July-Nov 2023**

End-term Project Report : Exploring pretraining for better finetuning on Brain MRI images

Team Name: Runtime (T)error *Team Members: 21D170021,21D171001*

Abstract

This project qualitatively compares between the performances of Self- Supervised Pretrained models over Brain MRI images. The comparsion is done for both the methods as well as the datasets used for the pretraining. Additionally, we introduce 3 new datasets to study how they help with the feature extraction and finetuning with less labelled data. Finally, from the best performing models, we investigate regarding how the results would change if we would do a Cross-Domain Self- Supervised Pretraining for the model and then arrive at conclusions about the practicality of the practice in real life health care studies.

1 Introduction

The main goal of this project is to provide accurate results of the diagnosis of the Alzheimer's disease and what category a patient is in regarding this. An early diagnosis gives an exponentially higher chance of survival, easier treatment, and financial savings. The current issue faced is that there is not enough data available for training as medical images are extremely difficult to annotate, requires highly skilled people to make and is expensive. To solve this problem, the solution proposed is to implement Self Supervised Learning (SSL) on the data available to us. Self Supervised Learning is becoming increasingly more popular in the medical field due to this. Pretraining is done on various data sets, and fine tuning on a classification data set of brain MRI images and the results are compared to see which data set fared the best.

We provide a survey of existing literature in Section 2. Our proposal for the project is described in Section 3. We give details on experiments in Section 5. A description of future work is given in Section 7. We conclude with a short summary and pointers to forthcoming work in Section 8.

2 Literature Survey

To do the background study for our project, we had been through multiple papers which will be discussed in brief in this section. Most of these papers are from the related works of the reference paper we used for our project.

Firstly we had been through (He et al., 2020), where the problem solved was that due to the shortage of CT-Scans of patients positive for Covid-19, the authors of the paper created a dataset of such CT Scans and also proposed a 'Self-Trans approach' which involves a contrastive self-supervised pretraining followed by transfer learning. Figure 1 shows the whole training pipeline implemented in the project, using pretrained ImageNet model in the beginning, and then doing the Self-supervised pretraining on the CT-Scans dataset after that, and then doing the transfer learning over this. Another paper we had been through which has a similar pipeline to the previous, but still for a different end purpose was (Azizi et al,. 2022), where there is still the same principle, where we have large-scale supervised transfer learning with self-supervised learning and finally requires little task-specific customization.

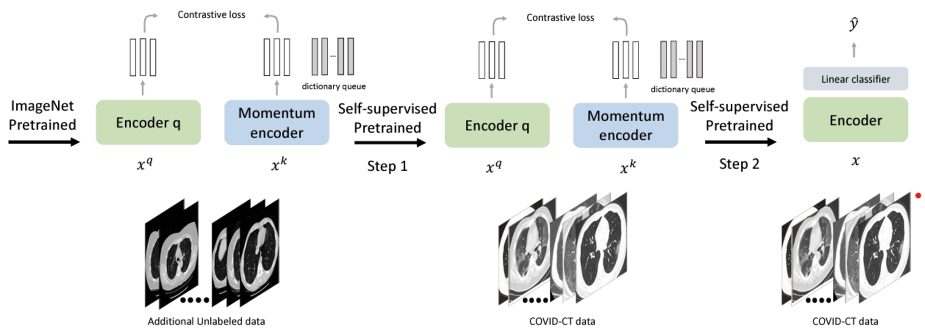

Figure 1: Entire training pipeline discussed in the paper [5]

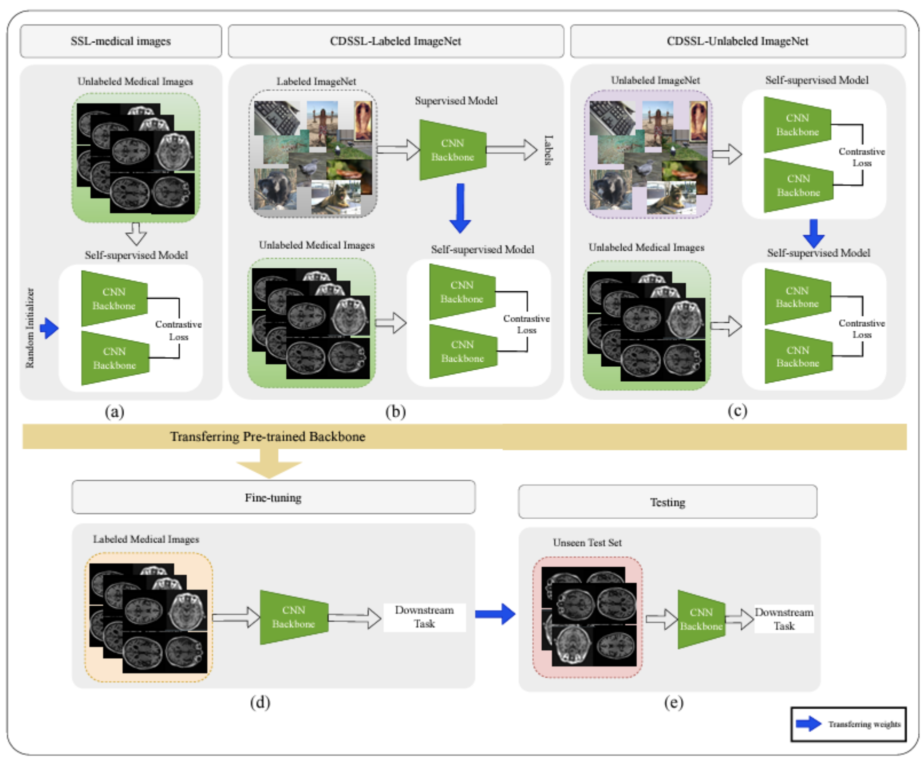

Figure 2: Entire training pipeline discussed in the paper [4]

Our project however draws major inspiration from a closely related paper on Cross Domain Self Supervised Pretraining (Dadsetan et al., 2023), and this is our reference paper. In their paper [4] , the authors perform many experiments to study how the pretraining over 2 different domains of images, namely Natural Images and Brain MRI images help in the final finetuning task of regression over CDR-SB scores (a metric to evaluate the severity of the Alzheimer's in the patient). Followed by which the metrics are compared to comment on the effectiveness of the different pretraining metrics, after which explanations are given for the trends noticed. This paper is quite fundamental to the motivation and idea for the modifications proposed by our team.

The fundamental idea behind the pipeline in this paper is that first the Backbone Encoder model (in this paper it is a ResNet50 while in our project it is a ResNet18), is first pretrained using the different pretraining methods and datasets, and finally that encoder is retreived to perform the fine-tuning on the Brain MRI images dataset for the regression task over CDR-SB scores.

As for the experiments conducted in this paper, Figure 2 explain them in a brief manner, where in each type of pretraining done, we can notice that the weights of these pretrained models are transfered on into the finetuning task, which is quite a key feature in our project.

3 Methods and Approaches

This project is going to study how pretraining on different datasets using the methods of SimCLR, BarlowTwins and SWaV will affect the finetuning task of classification over a dataset of Brain MRIs of alzheimer patients.

Before moving ahead, here is how each of the pretraining methods work:

SimCLR:

(SimCLR: A Simple Framework for Contrastive Learning of Visual Representations) (Chen et al., 2020c) is a self-supervised learning framework for learning representations from unlabeled data, particularly in the domain of computer vision. It was introduced by researchers at Google Research in a paper published in 2020. The key working principle in this framework is to send in augmented images (positive and negative pairs), and passing it along the pipeline to get representations over which contrastive learning is applied to bring similar instances closer together in the feature space and push dissimilar instances apart. The encoder that is used in the process will now be extracted to be used as the backbone for the finetuning task later on.

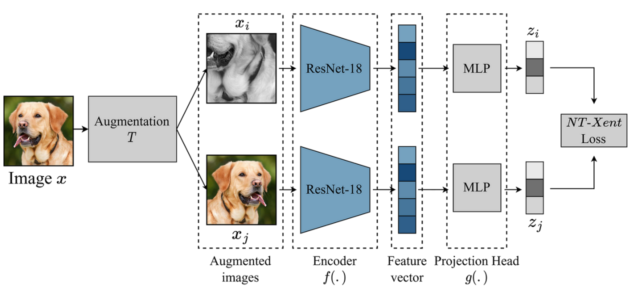

Figure 3: SimCLR pretraining pipeline

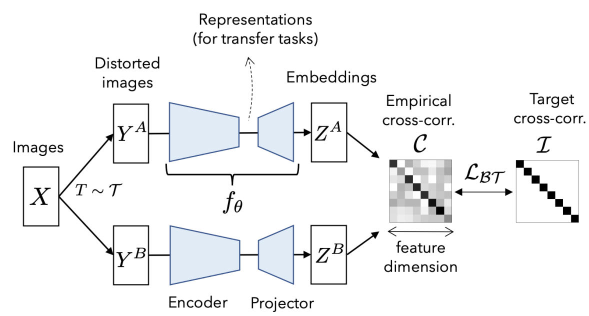

Figure 4: BarlowTwins pretraining pipeline

Barlow Twins: The Barlow Twins method (Zbontar et al., 2021) is based on a specific objective function that encourages the network to produce similar representations for similar inputs and dissimilar representations for dissimilar inputs. The objective function is designed to maximize the cross-covariance between the representations of augmented views of the same input and minimize the cross-covariance between representations of different inputs. The difference between SimCLR and BarlowTwins is that here we use a cross-covariance-based loss function, and not a contrastive learning approach. Now similar to the SimCLR method, we are going to be using the encoder block as the backbone for the fine-tuning tasks later on.

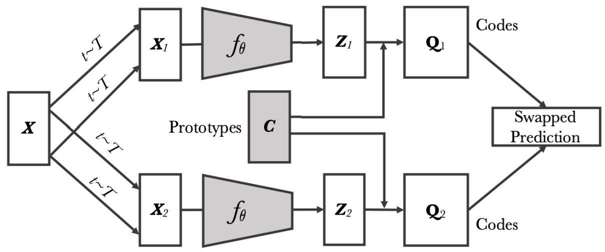

Figure 5: SwAV pretraining pipeline

SwAV:

The key idea behind SwAV (Caron et al.,2020) is to maximize the agreement between different augmentations of the same image while contrasting them with augmentations of other images. Unlike traditional contrastive learning methods (like in SimCLR), that rely on a fixed set of negatives for each positive, SwAV uses a "swapping" mechanism that allows it to create multiple negatives for each positive pair dynamically.

3.1 Work done before mid-term project review

For the mid term review, we had used pretrained models over Imagenet dataset which were found on the internet and then finetuned them over a dataset from Oasis Brains, where we did regression over the CDR score for finetuning. After the finetuning, we returned the R^2 metric to evaluate and compare the performance of the models.

3.2 Work done after mid-term project review

- Firstly instead of using pretrained models available online, we defined functions to do the entire pretraining locally, using the SimCLR, BarlowTwins and SwAV pretraining.
- Secondly, we changed the finetuning task from regression to classification over the CDR scores, since the CDR score were majorly categorical.As for the loss function now for the finetuning, we applied Cross Entropy Loss function.
- Thirdly, we used a bigger dataset for the finetuning task compared to the 209 images we had for finetuning before the mid term review. Currently the project is using 3 datasets of MRI inages from kaggle which is said to have data collected from different websites. The current finetuning dataset has 6400 images.
- Lastly, as for the modification to the original paper, we applied the pretraining pipeline to more Medical Image datasets to see how the results and metrics would change. The datasets include
  - Natural images and Brain MRI (originally discussed in the paper)
  - The ones we added in
    - Masked brain MRIs
    - CT scans of the body
    - X rays of chest

Finally we take the 2 different pretraining methods which perform the best, and do a cross Domain Self Supervised Learning over these 2 methods.

4 Data set Details

TinyImagenet

This dataset has been used as a replacement to the entire Imagenet for reproducability reasons. This dataset is to study how well natural images help the model to understand about the MRI images. This dataset has been accessed through a public Kaggle dataset called Imagenet-200

Brain MRI

For these images, we had used 3 different Kaggle datasets of Brain MRI images (details and links of which are present in the presentation slides), all three of which we imported into our Kaggle notebook to make one final big dataset of Brain MRI images.

Brain MRI masked

For this dataset, we found a kaggle dataset having the masked images but in .tif format, so for the preprocessing in this case, we had to convert the .tif files to .jpg, and uploaded these images on kaggle to make our own custom dataset, which is used in the project.

The next 2 datasets are to analyse how out of domain images, but still pertaining to the domain of medical images is going to have any effect to the results. We had to scrap out this data online.

CT scans

This dataset was also custom created by converting the .dcim images obtained from NBIA Cancer Images Archive (Link to the dataset is provided in the slides). These images were then used to make a custom dataset on kaggle, which was used for this task.

X ray of chests

We created 2 kaggle datasets for this finetuning, each of which were found online from the NIHCC collection of Chest X rays. These images were also in .dcim format, hence preprocessing was needed to convert the images to jpg, before finally making a custom dataset on kaggle for the images.

Preprocessing steps for the above datasets

For the pretraining processes, we have different preprocessing pipelines for each of SimCLR, BarlowTwins and SwAV methods. The following processes are applied to get the different augmentations of the images in all of the pretraining methods, and its hence an essential part of the pretraining to make the model more robust.

- Random resized crop
- Random horizontal flip
- Color jitter
- Random gray scale
- Gaussian blur
- ImageNet normalization

After all these transforms, a dataloader is created to feed the transformed images into the pretraining methods

4.1 Finetuning dataset

For the finetuning task, we are going to be using a Kaggle dataset which has the images of the Brain MRI divided into folders based on the category of the severity of the Alzheimer's disease (from which we decode the CDR values).

Preprocessing for the finetuning dataset

To these MRI scans from the dataset, we apply the following pre-processing methods:

- Converting the images to tensors
- Centre crop to resize the images to sizes of (224,224)
- Padding is added to the images

5 Experiments

In the pretraining using the proposed methods, different loss functions are used characteristic to the training algorithm, which were replicated from their respective papers and using open-source libraries (Lighlty SSL). Further, Stochastic Gradient Descent is used as the optimiser for the pretraining. We perform the experiments over the different pretraining datasets. Now we perform all the different pretraining methods over this dataset, and then use the backbone of these pretrained models to perform finetuning over the finetuning dataset. Then the metrics obtained from each of these models are compared to conclude on which pretraining method and pretraining dataset are the best for our particular downstream task.

In total, the following experiments are performed

- **Experiments on Natural Images and Brain MRI images:**

  These experiments were also conducted in the original paper and our results seem to agree with those from the paper, that the pretraining does help the model perform better on the downstream task.

- **Experiments on Masked Brain MRI images, CT scans and X ray images:**

  These experiments were our modification to the original paper to see how pretraining over this out-of-domain datasets would influence the results. However these datasets are not completely out of domain owing to the fact that they are still medical images, but not exactly of the same domain of the fine-tuning dataset (normal brain MRI images)

- **Experiments on Cross Domain Self Supervised Pretraining (CDSSP):**

  From the above experiments, we could conclude that the pretraining over the Natural Images and Brain MRI datasets produced the models with the best metrics. Morever, the barlow twins pretrained model has been performing the best quite consistently over all the experiments. Hence for this CDSSP experiment, we would be pretraining a Resnet-18 backbone with the Barlow Twins method over the TinyImageNet dataset, followed by then pretraining it over the Brain MRI dataset to study how the cross domain pretraining would influence the down stream task.

6 Results

6.1 Experiments on Natural Images and Brain MRI images

These datasets are what were prescribed by the paper.

Dataset used for Pretraining	SimCLR	BarlowTwins	SwAV	Untrained RESNET
Tiny IMAGENET	1.445	0.993	1.032	1.106
Brain MRI	2.329	1.008	1.050	1.106

Table 1: Results for the fine-tuning classification showing the average of log loss error for the test data. Comparison between the different pretraining methods are shown, and it can be noticed that with both the pretraining datasets, the Barlow Twins model performs better than the untrained resnet model.

6.2 Experiments on Masked Brain MRI images, CT scans and X ray images

Now doing the experiments over other medical datasets, this was the modification which we proposed for our project. Here the performance metrics of the different pretrained models

Dataset used for Pretraining	SimCLR	BarlowTwins	SwAV	Untrained RESNET
Masked Brain MRI Images	1.585	1.102	1.183	1.106
X Rays of Chest	1.700	1.098	1.191	1.106
CT SCANS	1.206	1.092	1.275	1.106

Table 2: Results for the fine-tuning classification showing the average of log loss error for the test data. It can be noticed that the pretrained models dont perform as good as the models trained over the original in domain dataset (brain mri), and extensive out domain dataset (natural images). It can however be noted that the barlow twins model still performs better than the untrained resnet model.

On to why the SimCLR model is performing quite poorly, its because the pretraining methods are heavily hyper-parameter sensitive, hence on getting the hyper-parameters wrong, it would lead to not reaching the optimal performance of the model. Hence we believe that the whole training process being the same, if we just increase the amount of validation on the hyper-parameters such as the learning rate, weight decay, number of epochs trained and batch size, we would be able to reach better metrics.

6.3 Experiments on Cross Domain Self Supervised Pretraining (CDSSP)

Here is the experiment we performed using the Barlow Twins pretraining over natural images and brain MRI since they gave the best metrics from the previous experiments.

Dataset used for Pretraining	BarlowTwins Pretraining	Untrained RESNET
Natural Images and then Brain MRI images	1.021	1.106

Table 3: Results for the fine-tuning classification showing the average of log loss error for the test data. Here we can see the Cross-Domain Self Supervised Learning over the Natural Images and the Brain MRI performs better than the untrained network

7 Future Work

Here are a few ideas that can be worked on over this project, to understand better about all the pretraining methods and how we can benefit from them.

- We could work around with the amount of data that is being used in the pretraining to quantitivaely analyse if more pretraining data is helping the model generalise faster.
- We could perform deeper validation of the pre-training and fine-tuning process by using different values of learning rates, number of epochs, weight decay and the batch size of the data loaders. Doing this ensures that we have the pre-training model which would perform to the best of its capability on limited fine-tuning data.
- After the finetuning of the final model is done, we can produce saliency maps using these to analyse which regions of the brain MRI were useful to to make out the severity of the patient. This can help the healthcare industry get a better understanding of the disease and how it affects the brain.

8 Conclusion

In the project, the problem of having very little high quality annotated image data for fine-tuning tasks was to be solved, particularly for the problem of Alzheimer's severity classification using the brain MRI scans. For this problem, we employed the idea of pretraining over a ResNet backbone which would then be transferred over to the fine-tunining task in hopes that it would give better performance metrics, basically signifying that it would learn from very limited annotated data.

In the experiment we performed, we do the pretraining using SimCLR, Barlow Twins and SwAV methods over datasets of Natural images, brain MRIs, masked brain MRIs, CT Scans and X Rays, and we consistently see that Barlow Twins Pretraining always beats the untrained ResNet, showing how pretraining can be beneficial in those areas such as healthcare where this is shortage of quality annotated datasets. Future directions of the project can be to extract the best performing models by fine-tuning the hyper parameters.

References

[1] Azizi, S., Culp, L., Freyberg, J., Mustafa, B., Baur, S., Kornblith, S., Chen, T., MacWilliams, P., Mahdavi, S.S., Wulczyn, E. and Babenko, B., 2022. Robust and efficient medical imaging with self-supervision. arXiv preprint arXiv:2205.09723.

[2] Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P. and Joulin, A., 2020. Unsupervised learning of visual features by contrasting cluster assignments. Advances in neural information processing systems, 33, pp.9912-9924.

[3] Chen, X., Fan, H., Girshick, R. and He, K., 2020. Improved baselines with momentum contrastive learning. arXiv preprint arXiv:2003.04297.

[4] Dadsetan, S., Hejrati, M., Wu, S. and Hashemifar, S., 2023, March. Robust Alzheimer's Progression Modeling using Cross-Domain Self-Supervised Deep Learning.

[5] He, X., Yang, X., Zhang, S., Zhao, J., Zhang, Y., Xing, E. and Xie, P., 2020. Sample-efficient deep learning for COVID-19 diagnosis based on CT scans. medrxiv, pp.2020-04.

[6] Zbontar, J., Jing, L., Misra, I., LeCun, Y. and Deny, S., 2021, July. Barlow twins: Self-supervised learning via redundancy reduction. In International Conference on Machine Learning (pp. 12310-12320). PMLR.