Project Proposal

CS 598 - Deep Learning for Healthcare | Spring 2024
Project Members: Brian Betancourt, John Lewis, and Robbie Li

Paper Citation (APA)

Singh, P. & Cirrone, J.. (2023). Efficient Representation Learning for Healthcare with Cross-Architectural Self-Supervision. *Proceedings of the 8th Machine Learning for Healthcare Conference*, in *Proceedings of Machine Learning Research* 219:691-711 Available from https://proceedings.mlr.press/v219/singh23a.html.

1. General Problem

The motivation behind introducing CASS (Cross-Architectural Self-Supervised Learning) is two-fold: (1) existing state-of-the-art self-supervised learning methods have extreme computational requirements that make them inaccessible to most practitioners, and (2) the limited amount of data makes it infeasible to run smaller epochs with larger batch size to achieve the effectiveness outlined in these self-supervised learning methods.

Representation learning allows us to use self-supervision to learn useful priors by pretraining unlabeled images. This is crucially important because the medical imaging field suffers from a lack of available labeled data due to a variety of factors, such as the high cost of labeling data at scale because it generally requires domain-specific knowledge. In fields with limited data or high cost to produce labels, we can use self-supervision to help with the downstream learning process without the need for labels.

Existing state-of-the-art self-supervision techniques require significant computational resources, sometimes taking upwards of 20 GPU days to train. This provides a major hurdle in using them for pretraining unlabeled data. There have been attempts to use smaller epochs with larger batches during pre-training. However, larger batch sizes are not always feasible because of limited data availability. For example, existing techniques are trained with batch sizes of 1024 or larger, which can be larger than the entire medical imaging dataset. However, smaller batch sizes significantly reduce the existing methods' performance. It is clear that there is a need for a more computationally efficient representation learning method with a higher allowance for variance in epoch and batch size. CASS introduces a method that combines CNN and Transformers to learn predictive data presentations in limited data scenarios.

2. Approach

CASS helps solve our general problem by leveraging CNN and Transformer methods for efficient learning in a siamese contrastive method. Traditionally, contrastive self-supervision methods use different augmentations of images to create positive pairs. CASS leverages architecture invariance instead of using this augmentation invariance approach. Extracted representations of each input image are compared across two branches representing each respective architecture. By contrasting their extracted features, they can learn from each other on patterns they would generally miss. This helps provide more useful pre-trained data for the downstream learning method.

CASS's approach helps reduce the time complexity of pre-training in two major ways. First, augmentations are only applied once in CASS in comparison to twice in augmentation invariance approaches. Therefore per application CASS uses less augmentations overall. Second, there is no scope for parameter sharing in CASS because the two architectures used are different. A large portion of time is saved in updating the two architectures each epoch as opposed to re-initializing architectures with lagging parameters.

3. Hypothesis

Our inferred hypothesis (given it is not explicitly stated) from the research paper is that leveraging the CASS self-supervised learning approach will significantly improve the efficiency of representation learning in healthcare applications in scenarios which involve a lack of data or computing resources.

4. Ablations

Our intended ablations for this project involve reducing the number of pre-training epochs and batch sizes to understand their importance in the model performance. Furthermore, the paper in Appendix B.5 mentions additional ablation studies by applying (1) a softmax layer, (2) a sigmoid layer, and (3) neither sigmoid nor softmax layer. We plan to replicate the same ablation studies to validate the findings in the paper.

5. How We Plan to Access Data

The paper outlines 4 different datasets that were used to test the CASS method: (1) autoimmune diseases biopsy slides, (2) Dermofit dataset, (3) brain tumor MRI datase, and (4) ISIC 2019 dataset. We will make a best-effort to access all 4 datasets. We are confident that we will be able to access at least one of the datasets and in the case where one or more of the datasets is unavailable or difficult to get, we will focus on the datasets that we are able to access.

We can also explore the possibility of using the MedMNIST datasets that are used in the GitHub repository. If these datasets are sufficient to test our hypothesis, this would be an excellent option as the GitHub repository has code for the preprocessing of this data readily available.

6. Feasibility Discussion

The entire point of CASS is to improve computational feasibility of self-supervision, and it has been used on small datasets. It has also been proven to be better with smaller epochs and batch sizes than existing methods. While we have been able to access some of the datasets outlined above, there are a few, such as the Dermofit dataset, that seem to be more difficult to access.

The hardware used by the original researchers is quite advanced (NVIDIA RTX8000 GPU with 48GB video memory) and we may not have access to machines that match the specifications used, even with the availability of cloud computing resources. Therefore, we may need to run our model evaluations on a much smaller subset of the data, which could make it difficult to replicate the original findings.

7. Existing Code Statement

We intend to use existing code. Given the maturity displayed in the documentation of the Github repository, this will allow us to concentrate on our efforts on testing the hypothesis instead of attempting to re-create the model methodology from the paper, from scratch.

8. References

Singh, P. & Cirrone, J.. (2023). Efficient Representation Learning for Healthcare with Cross-Architectural Self-Supervision. *Proceedings of the 8th Machine Learning for Healthcare Conference*, in *Proceedings of Machine Learning Research* 219:691-711 Available from https://proceedings.mlr.press/v219/singh23a.html.

Singh, P. (2022). *Official PyTorch implementation of CASS.* [Software]. Retrieved from https://github.com/pranavsinghps1/CASS

Cheng, J. (2017, April 2). *Brain tumor dataset*. figshare. https://figshare.com/articles/dataset/brain_tumor_dataset/1512427

Tschandl P., Rosendahl C. & Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci. Data 5, 180161 doi.10.1038/sdata.2018.161 (2018)

Noel C. F. Codella, David Gutman, M. Emre Celebi, Brian Helba, Michael A. Marchetti, Stephen W. Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, Allan Halpern: "Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC)", 2017; arXiv:1710.05006.

Marc Combalia, Noel C. F. Codella, Veronica Rotemberg, Brian Helba, Veronica Vilaplana, Ofer Reiter, Allan C. Halpern, Susana Puig, Josep Malvehy: "BCN20000: Dermoscopic Lesions in the Wild", 2019; arXiv:1908.02288.

Jiancheng Yang, Rui Shi, Donglai Wei, Zequan Liu, Lin Zhao, Bilian Ke, Hanspeter Pfister, Bingbing Ni. Yang, Jiancheng, et al. "MedMNIST v2-A large-scale lightweight benchmark for 2D and 3D biomedical image classification." Scientific Data, 2023.

Jiancheng Yang, Rui Shi, Bingbing Ni. "MedMNIST Classification Decathlon: A Lightweight AutoML Benchmark for Medical Image Analysis". IEEE 18th International Symposium on Biomedical Imaging (ISBI), 2021.