ELEC4840 Project: Skin Lesion Classification

Young, James Yang (20740589)

The Hong Kong University of Science and Technology

Abstract. This project tackles the 2016 ISIC Skin Lesion challenge, where images are classified as either benign or malignant. The project uses two methods, namely the Fourier Augmented Co-Teacher (FACT) framework and pseudo-labeling with semi-supervised learning. The results of these two methods are compared against a baseline model that uses supervised learning.

1 Introduction

Skin lesion classification plays a crucial role in dermatology and healthcare by enabling early detection, diagnosis, and treatment of various skin conditions and diseases. Accurate and timely classification of these lesions is imperative for dermatologists and physicians to make well-informed decisions and deliver optimal patient care. In this project, it aims to perform skin lesion classification with the ISIC 2016 challenge data [1] using domain generalization with the Fourier Augmented Co-Teacher (FACT) framework [2] and semi-supervised learning with pseudo-labeling. The results are compared against a baseline supervised learning model.

1.1 Literature Review

The first method for skin lesion classification uses semi-supervised learning with pseudo-labeling. Pseudo-labeling is a semi-supervised learning technique that involves assigning labels to unlabeled data based on the predictions of a trained model. As labeling data can be difficult, pseudo-labeling leverages the abundance of unlabeled data to enhance model performances. For example, Dzieniszewska et al. used pseudo-labeling with ISIC 2019 and 2020 data to enhance skin lesion segmentation on ISIC 2016 and 2017 skin lesion data [3].

The second method for skin lesion classification uses the FACT framework. As described by Xu et al. [2], FACT framework introduces a novel Fourier-based perspective for domain generalization. The key assumption is that the Fourier phase information carries significant semantic meaning and remains robust against domain shifts. To effectively capture the phase information, a novel Fourier-based data augmentation strategy called amplitude mix is developed, which interpolates between the amplitude spectra of two images. Additionally, a dual-formed consistency loss termed co-teacher regularization is introduced, encouraging consistency between predictions obtained from original and augmented images. Through extensive experiments on three benchmark datasets, the proposed method demonstrates state-of-the-art performance in the domain generalization task, highlighting its efficacy in addressing this challenging problem [2].

2 Methodology

2.1 Baseline Model

The baseline model only uses supervised learning. It uses ResNet-50 as the model architecture and is trained and evaluated on the ISIC 2016 skin lesion challenge data. The data has 900 training and 329 testing images, where for this project it is partitioned to 720 training, 180 validation, and 329 testing images. Before training, the images are augmented with random flipping, resized, and normalized. The Adam optimizer is used with a learning rate of 0.01 and a BCE loss function is used.

2.2 Semi-Supervised Learning with Pseudo-Labeling

The semi-supervised learning model uses pseudo-labeling. In this method, 2000 unlabeled data is taken from ISIC 2017 training data [5]. The model is first trained on 720 labeled data from ISIC 2016 training data. Then, the trained model is used to generate predicted labels on the unlabeled data. These pseudo-labels are then used alongside the labeled data to re-train the model. The Adam optimizer is used with a learning rate of 0.01 and a BCE loss function is used. The loss calculated uses a combination of the losses computed for both the labeled and pseudo-labeled data. It is evaluated on 329 images from ISIC 2016 testing data, same as the baseline model.

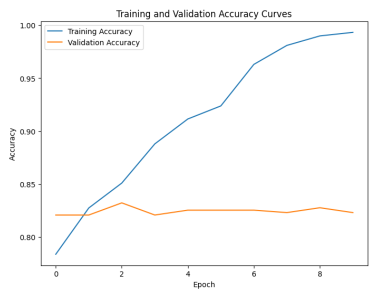
2.3 Domain Generalization with FACT Framework

The methodology for the FACT framework involves several steps. First, the dataset is augmented using Fourier transformations, where the amplitude spectrums of two images are linearly interpolated using the amplitude mix approach. This augmentation helps the model capture phase information, which is assumed to be invariant to domain shifts [2]. This way, the model can avoid overfitting to low-level features in the amplitude information. Then, a co-teacher regularization loss is used to enforce consistency between predictions on the original and augmented images. The co-teacher regularization method, within the context of the FACT (Few-shot Adaptive Classification Transfer) framework, involves training multiple co-teachers simultaneously. Each co-teacher makes predictions on unlabeled data and exchanges those predictions with the others. By encouraging consensus through a consistency loss, the co-teachers align their predictions and collectively improve their performance for few-shot classification tasks. The Adam optimizer is used with a learning rate of 0.01 and the cross-entropy loss function is used. The training data, validation, and test data are the same as the baseline model.

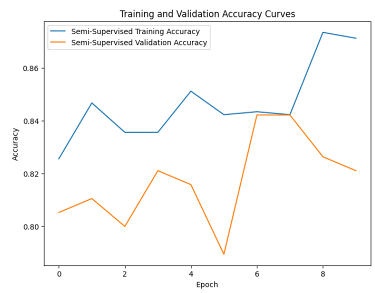
3 Experiment Results, Analysis, Future Improvements

3.1 Experiment Results

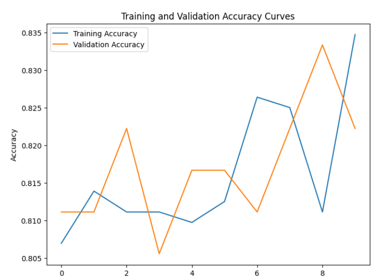
Figure 1 shows the training and validation curves for all 3 methods and Table 1 shows the testing results for all three methods.



(a) FACT Framework



(b) Pseudo-Labeling



(c) Baseline

Fig. 1: Training and Validation Accuracy Curves

Table 1: Result Comparison.

Method	AUC	Accuracy
Baseline Model	77.7%	79.8%
Pseudo-labeling	81.4%	82.1%
FACT Framework	71.4%	80.5%

As demonstrated in Table 1, the pseudo-labeling method was able to reach a higher AUC and accuracy score. However, whilst the FACT framework had slightly higher accuracy, it has a much lower AUC score.

3.2 Result Analysis

As shown in the experimental results, the pseudo-labeling method made a slight improvement to the baseline performance. By utilizing the large amount of unlabeled data available in the ISIC 2017 training dataset, pseudo-labeling enables the generation of additional labeled data through the assignment of pseudo-labels based on model predictions. This augmented labeled dataset can provide more diverse and representative samples for training, potentially enhancing the model's ability to generalize to unseen data and mitigating the limited labeled data problem.

However, the FACT framework saw lower overall performance which could be due to several factors. Since it only uses 720 training images from the ISIC 2016 data, the limited number of labeled examples in the dataset might not provide enough information. Furthermore, poor hyper-parameters or evaluation metrics may have hindered performance and due to time-constraints different parameters were not tested.

3.3 Future Improvements

For a more comprehensive analysis, more hyper-parameters should be tested. For example, the use of different epochs, learning-rates, loss functions should be tested. In the pseudo-labeling method, more testing should be done with different number of pseudo-labels and labeled data. For example, tests should be run with 1000 pseudo-label vs 2000 pseudo-label for a more comprehensive review. Also, different weights and confidence thresholds could be used for comparison. In the FACT framework, different evaluation metrics and hyper-parameters may have been a factor in its lower performance against the baseline model.

4 Conclusion

In conclusion, this project uses the FACT framework for domain generalization and pseudo-labeling for semi-supervised learning for skin lesion classification and compares these methods with a baseline supervised learning model. The project found that the supervised only baseline model achieved an AUC of 77.7% and an accuracy of 79.8%. The pseudo-labeling method achieved an AUC of 81.4% and an accuracy 82.1%, whilst the FACT framework achieved an AUC of 71.4% and an accuracy of 80.5%.

Although the pseudo-labeling method was able to gain small improvements to overall performance against the baseline model, the FACT framework was unable to improve from the baseline model. Further testing with different hyper-parameters, evaluation metrics, and other tweaks can be made to improve performances of both methods.

References

1. Gutman, David; Codella, Noel C. F.; Celebi, Emre; Helba, Brian; Marchetti, Michael; Mishra, Nabin; Halpern, Allan. "Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC)". eprint arXiv:1605.01397. 2016.
2. Xu, Q., Zhang, R., Zhang, Y., Wang, Y., Tian, Q.: A fourier-based framework for domain generalization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14383–14392 (2021)
3. A. Dzieniszewska, P. Garbat, and R. Piramidowicz, "Improving Skin Lesion Segmentation with Self-Training," in Cancers (Basel), vol. 16, no. 6, Mar. 11, 2024, pp. 1120. doi: 10.3390/cancers16061120. PMID: 38539454; PMCID: PMC10969337.
4. Qinwei Xu, Ruipeng Zhang, Ya Zhang, Yanfeng Wang, and Qi Tian. "A Fourier-based Framework for Domain Generalization." In: arXiv preprint arXiv:2105.11120 (2021).
5. Codella N, Gutman D, Celebi ME, Helba B, Marchetti MA, Dusza S, Kalloo A, Liopyris K, Mishra N, Kittler H, Halpern A. "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)". arXiv: 1710.05006 [cs.CV]