ELEC4010N Final Project Report

Project 1. Semi-Supervised Classification

1. Background Introduction

Semi-supervised learning holds significance in medical imaging, attributable to the scarcity of expert-annotated data, the diversity and intricacy of image patterns, and the presence of infrequent conditions. The objective of this approach is to decipher patterns and inherent uncertainty in an abundant set of unlabeled data, drawing from a smaller pool of labeled data.

In this project, the dermoscopic lesion dataset from the International Skin Imaging Collaboration (ISIC) 2016 challenge was used. Out of the 900 images available, the images were randomly partitioned into three subsets: a labeled training set consisting of 270 images, an unlabeled training set comprising 540 images, and a validation set of 90 images.

To ensure reliable results, this random partitioning process was replicated and tested a minimum of four times. It is important to note the existence of a class imbalance issue, with benign images significantly outnumbering malignant ones. To address this, we employed upsampling and augmentation techniques on the labeled training set, effectively increasing the number of images by multiple folds. Further measures were implemented to counteract the class imbalance issue during training.

2. Method Developments

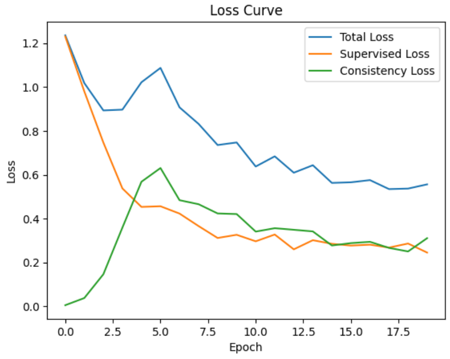
The Residual Network (ResNet-50), initialized with pre-trained "IMAGENET1K_V2" weights, a softmax output layer, and a dropout rate of 0.5, served as the baseline model for supervised binary classification. Subsequently, a Mean Teacher model was introduced. The ResNet-50 model functioned as the student model, while the teacher model was a deep copy of the student model at initialization.

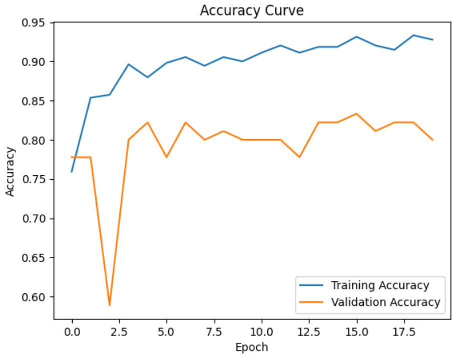
The student model processed the labeled training data to yield the supervised loss, while the teacher model processed the unlabeled training data to yield the consistency loss. The weights of the teacher model were updated by exponential moving average. The supervised loss was calculated using Binary Cross Entropy (BCE) focal loss, which merges BCE loss (for binary classification) and focal loss (to address class imbalance).

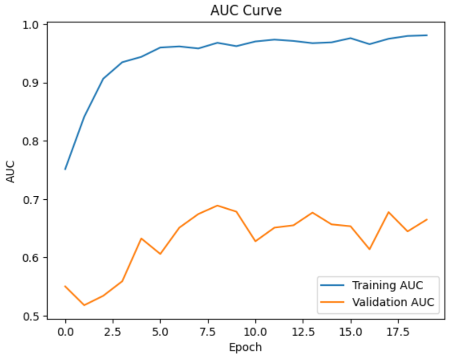
$$\text{BCEFocalLoss}\left(p_t\right) = -\left(\alpha_t\left(1-p_t\right)^{\gamma}\log\left(p_t\right) + \left(1-\alpha_t\right)p_t^{\gamma}\log\left(1-p_t\right)\right)$$

Modifications such as sigmoid ramp-up and variable momentum were applied to adjust the weight of the consistency loss. The gradual ramp-up for consistency loss aids in preventing an early decay in model performance during the initial epochs.

3. Results Analysis







The application of ramp-up mechanisms led to the convergence of the consistency loss to the supervised loss within approximately 5 epochs. Results were notably improved when the labeled training data was upsampled to a considerably high volume. After numerous tests, an approximate validation accuracy of **80%** and a validation (AUC) of **0.65** were attained, as compared with the baseline (ResNet50 + 270 labeled images) at a validation accuracy of **81%** and validation AUC of **0.54** after 20 epochs. Many different combinations of data portions and hyperparameters were tested. It is worth noting that the validation results do not change dramatically after certain epochs. Moreover, it should be noted that excessive balancing of the data through upsampling can have a counterproductive effect, potentially leading to a degradation in model performance.

4. Conclusion

The project demonstrated the effectiveness of semi-supervised learning in addressing challenges intrinsic to medical imaging, such as scarce expert-annotated data, complex image patterns, and rare conditions. The integration of a pre-trained ResNet-50 model with a Mean Teacher model facilitated the processing of both labeled and unlabeled data, enhancing the model's robustness and adaptability. Notably, addressing class imbalance through data augmentation and upsampling proved instrumental in improving model performance. These results highlight the potential of semi-supervised learning in enhancing the accuracy of disease detection and prognosis in medical imaging, a critical aspect in providing efficient, high-quality patient care.

Project 2. Domain Generalization on Fundus Images

1. Background Introduction

Deep neural networks have demonstrated challenges when handling out-of-sample distributions, which suggests a limited ability to generalize effectively to real-world scenarios. As a solution, domain generalization has been proposed. This technique is particularly vital in medical imaging, given the variability in imaging equipment across institutions, continual advances in medical imaging technology, and scarcity of domain-specific labeled data. These factors call for models that can effectively generalize across a range of imaging conditions, adapt to emerging technologies, and utilize data from various domains.

In this project, the fundus images segmentation dataset from the paper “Domain-Oriented Feature Embedding for Generalizable Fundus Image Segmentation on Unseen Datasets” (DoFE) was used. The dataset comprises a total of 1060 images sourced from four distinct domains: Domain 1 (101 images), Domain 2 (159 images), Domain 3 (400 images), and Domain 4 (400 images). The segmentation masks for each image include three categories - background, optic disk, and optic cup - necessitating a multi-class segmentation approach. To validate the model's domain generalization capabilities, we adopted a cross-validation-like approach where we trained the model over 15 epochs on three domains and tested it on the remaining one, repeating this process for four separate experimental runs.

2. Method Developments

The Fourier Augmented Co-Teacher (FACT) was implemented, following the methodology outlined in the paper "A Fourier-based Framework for Domain Generalization". It is basically the aforementioned Mean Teacher model fed by original data and Fourier augmented data. It adopted the Fourier transform technique for data augmentation, extracting both amplitude and phase components. The paper proposed Amplitude Swap (AS) and Amplitude Mixup (AM) strategies, and the AS strategy was deemed potentially overwhelming for the model's learning capacity. Therefore, the AM strategy was adopted, using linear interpolation to compute the amplitude of the augmented image.

$$\widehat{\mathscr{A}}\left(x_i^k\right) = (1-\lambda)\,\mathscr{A}\left(x_i^k\right) + \lambda\mathscr{A}\left(x_{i'}^{k'}\right)$$

The succeeding illustrations display two images before and after the Fourier augmentation. To initiate the process, images and masks were read as 3-channel images to accommodate the Fourier transform data loader. The masks contained three unique values corresponding to three distinct classes, which were subsequently utilized for one-hot encoding. Following the application of data augmentation on the training dataset, the outcomes of the Amplitude Mixup are represented in the images below.

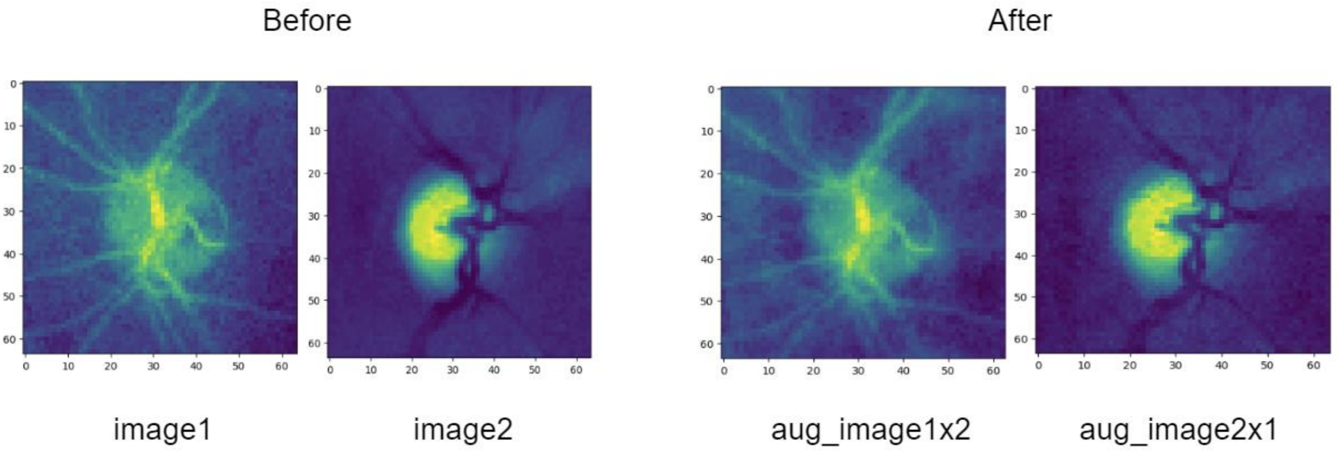


image1 image2 aug_image1x2 aug_image2x1

The central architecture was the Mean Teacher Model, composed of two submodels - the student and teacher models, both utilizing the U-Net architecture with outputs channels equal to 3 for the semantic segmentation task. Both the original and Fourier-augmented data were processed by the student and teacher models concurrently.

During the training process, the teacher model updated its parameters based on both its own parameters and those of the student model, after the computation of the loss. The total loss consisted of the supervised loss, calculated between the outputs and ground truth labels, and the consistency loss, determined between the outputs of the teacher and student models.

The supervised loss was quantified using the categorical cross-entropy loss function. Notably, after several experiments, it was empirically observed that the Dice Loss did not provide feasible convergence of the loss. Prior to the computation of the supervised loss, the ground truth labels were subjected to one-hot encoding, and the outputs underwent activation by the softmax function over the channel dimension. Next, the supervised loss (segmentation loss) between ground truth labels and the student model outputs was obtained.

$$\mathscr{L}_{\text{cls}}^{\text{ori}} = - y_i^k \log\Big(\sigma\Big(f\big(x_i^k, \theta\big)\Big)\Big)$$

$$\mathscr{L}_{\text{cls}}^{\text{ori}} = - y_i^k \log\Big(\sigma\Big(f\big(x_i^k, \theta\big)\Big)\Big)$$

The teacher model updated its weights by applying an exponential moving average of the student model's weights. To ensure consistency between original and augmented images, Kullback-Leibler (KL) divergence was implemented for co-teacher regularization. Comparisons were drawn between the original outputs of the student model and the augmented outputs of the teacher model, and conversely. This procedure promoted consistency in the outputs of both models throughout the training process and facilitated the achievement of comparable results between augmented and raw input data across both student and teacher models.

$$\mathscr{L}_{\text{cot}}^{a2o} = \mathrm{KL}\Big(\sigma\Big(f_{\text{stu}}\big(\hat{x}_i^k\big) / T\Big) \Big\| \sigma\Big(f_{\text{tea}}\big(x_i^k\big) / T\Big)\Big)$$

$$\mathscr{L}_{\text{cot}}^{o2a} = \mathrm{KL}\Big(\sigma\Big(f_{\text{stu}}\big(x_i^k\big) / T\Big) \Big\| \sigma\Big(f_{\text{tea}}\big(\hat{x}_i^k\big) / T\Big)\Big)$$

$$\mathscr{L}_{FACT} = \mathscr{L}_{cls}^{ori} + \mathscr{L}_{cls}^{aug} + \beta\Big(\mathscr{L}_{cot}^{a2o} + \mathscr{L}_{cot}^{o2a}\Big)$$

Finally, the total loss was calculated as the sum of the supervised loss and the weighted consistency loss.

Modifications such as sigmoid ramp-up and variable momentum were also used in this project. These methods not only serve to mitigate the issues arising from class imbalance but also enhance the performance of the teacher model.

3. Results Analysis

The metrics used were the Sørensen–Dice coefficient (Dice) and Average Surface Distance (ASD) due to their effectiveness in quantifying segmentation accuracy, particularly when small boundary shifts can cause significant outcomes. This project involved multi-class segmentation, with each class being mutually exclusive.

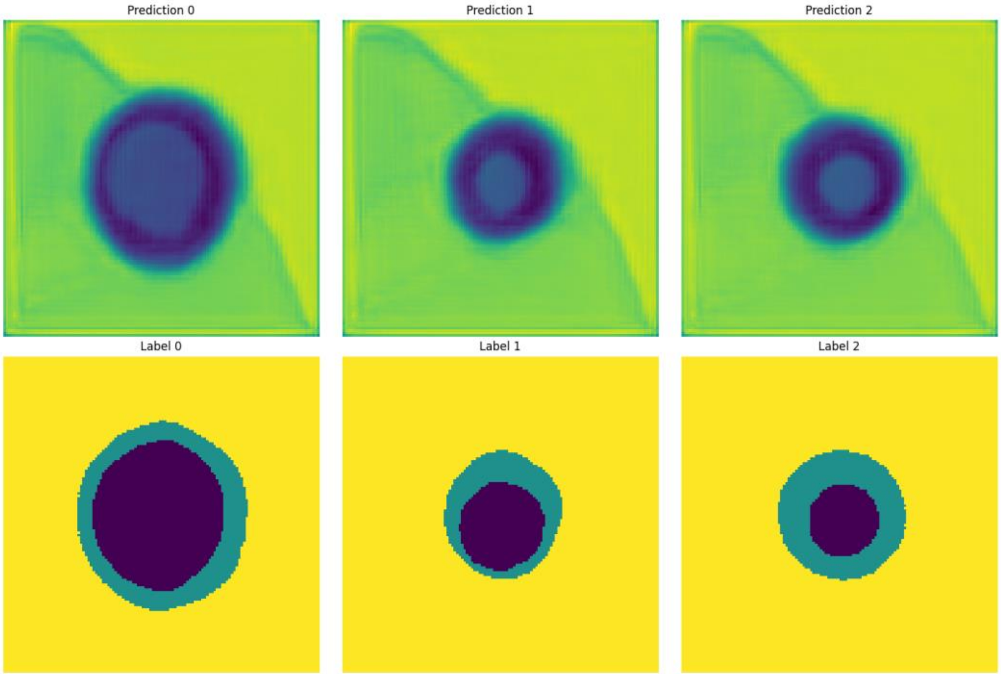
The Dice coefficient considers all predictions (True Positive, False Positive, True Negative, False Negative) and offers an accuracy metric suitable for medical imaging. It imposes penalties for both False Negative and False Positive predictions, thereby providing a balanced measure of segmentation performance. As previously stated, the ground truth labels were one-hot encoded to represent each class before being compared with the predictions. Subsequently, the total Dice coefficient was calculated.

$$\text{Dice} = \frac{2TP}{2TP + FP + FN}$$

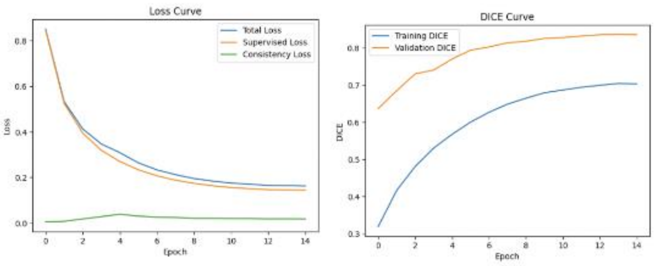
ASD is calculated individually for each region as a part of a binary segmentation task using the corresponding mask. The ASD for optic cup (OC) and optic disk (OD) is separately computed to reflect the model's performance on different subsets of imbalanced data.

$$\text{ASD}(A,B) = \frac{1}{|S(A)| + |S(B)|}\left(\sum_{a \in S(A)} \min_{b \in S(B)} \left\| a - b \right\| + \sum_{b \in S(B)} \min_{a \in S(A)} \left\| b - a \right\|\right)$$

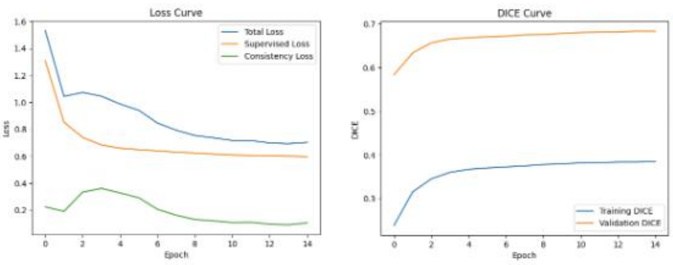
The following figures show some samples of the predictions and the labels.



Beyond the data imbalance issue, several exploratory experiments were conducted concerning the consistency loss. Performance was compared between the KL divergence and Mean Squared Error (MSE) loss, and it was determined that MSE loss yielded superior results as it calculates loss on the raw logits vector rather than computing loss on softened probabilities. This facilitates more effective learning for the student model. In the conducted experiment with a consistent dataset split, the model that utilized MSE loss outperformed the approach using KL divergence and yielded smoother loss curves under the same ramp-up initialization.



MSE Loss



KL Divergence

The training process was replicated across four separate runs. The results shown below demonstrate enhancements in the performance on the test dataset of the FACT model compared to the baseline model. The baseline model initialization is pure U-Net with categorical cross-entropy loss also trained over 15 epochs.

Train	Test	Model	Mean Test Dice	OC Test ASD	OD Test ASD
123	4	Baseline	0.5781	36.9649	27.7053
		FACT	**0.8730**	**7.4794**	**1.7167**
124	3	Baseline	0.6057	35.9788	24.8685
		FACT	**0.9039**	**6.2443**	**0.6492**
134	2	Baseline	0.6988	24.0777	15.9232
		FACT	**0.8527**	**8.2105**	**1.4624**
234	1	Baseline	0.6376	30.5020	21.3653
		FACT	**0.8996**	**5.3635**	**1.2530**

4. Conclusion

The implementation of Domain Generalization via the Fourier Augmented Co-Teacher (FACT) model markedly improved image segmentation. However, for this specific dataset, employing Dice Loss for backpropagation presented challenges. Therefore, instead of Dice loss, categorical cross-entropy loss was adopted as the primary loss function. Notwithstanding, the imbalanced optic cup data in the fundus dataset led to high error scores for the optic cup, as indicated above; the ASD of optic cup did not exhibit substantial improvement in the FACT model compared to the baseline performance. Future work may still necessitate the application of soft Dice loss as the loss function. Additionally, the results suggest that KL divergence, as proposed in the cited paper, may not be the optimal approach for the knowledge distillation architecture. The suboptimal result underscores the importance of continued investigation into alternative methods for further enhancing the performance of domain generalization in medical imaging.

Reference

- Tarvainen, A., Valpola, H. (2017). Mean Teachers Are Better Role Models: Weight-averaged Consistency Targets. Improve Semi-supervised Deep Learning Results. *arXiv:1703.01780*
- Wang, S., Yu, L., Li, K., Yang, X., Fu, C.-W., Heng, P.-A. (2020). DoFE: Domain-oriented Feature Embedding for Generalizable Fundus Image Segmentation on Unseen Datasets. IEEE Transactions on Medical Imaging. (https://github.com/emma-sjwang/Dofe)
- Xu, Q., Zhang, R., Zhang, Y., Wang, Y., Tian, Q. (2021). A Fourier-Based Framework for Domain Generalization. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). (https://github.com/MediaBrain-SJTU/FACT)
- Laine, S., & Aila, T. (2017). Temporal Ensembling for Semi-Supervised Learning. International Conference on Learning Representations (ICLR). *arXiv:1610.02242*
- Kim, T., Oh, J., Kim, N., Cho, S., & Yun, S. (2021). Comparing Kullback-Leibler Divergence and Mean Squared Error Loss in Knowledge Distillation. Proceedings of International Joint Conference on Artificial Intelligence (IJCAI). *arXiv:2105.08919*