DEEP LEARNING FINAL PROJECT FOR MULTI-LABEL RETINAL DISEASE CLASSIFICATION

Moiz Khan

Oulun Yliopisto

ABSTRACT

In this project Multi-labeled Classification of Diabetic Retinopathy (DR), Glaucoma (G) and Age-related Macular Degeneration (AMD) on the ODIR dataset using transfer learning was tackled. The training set had 800 significant class imbalance (DR: 517, G: 163, AMD: 142 positives). In this project a number of techniques were explored such as fine-tuning strategies, imbalance-handling losses, attention mechanisms (SE and MHA) and stronger backbones. EfficientNet-B0 was the primary model used in this project. Focal Loss significantly outperformed BCE and Class-Balanced Loss. Squeeze-and-Excitation and Multi-Head Attention improved feature focus and resulted in F1 scores surpassing reference scores . The best model (EfficientNet with Focal Loss) achieved an onsite average F-score of 0.854, exceeding the reference baseline of 0.804 by 0.05.

Index Terms— Transfer learning, multi-label classification, retinal disease, focal loss, attention mechanisms, class imbalance

1. INTRODUCTION

In the medical field privacy is a great concern, along with the cost of annotation and experts needed. The ODIR dataset contains images for three diseases, with strong positive class imbalance. The project explored and systematically evaluated transfer learning strategies, loss functions for imbalance, attention modules and advanced techniques to improve performance.

2. METHODS

2.1. Transfer Learning

We explored and compared three fine-tuning modes on EfficientNet-B0 and ResNet18 which included no fine-tuning, frozen backbone with classifier training and full fine-tuning. I was found that EfficientNet outperformed ResNet18 in all modes.

2.2. Loss Functions

The training distribution was DR: 517, G: 163, AMD: 142 positives. Seeing this it was obvious that there is a significant class imbalance so we implemented Focal Loss and found the best parameter for gamma to be $\gamma = 1.3$ and Class-Balanced Loss beta $\beta = 0.9$. Focal Loss improved the score significantly and achieved balanced improvement while CB Loss overfitted to minority classes.

2.3. Attention Mechanisms

To enhance feature selection we added attention mechanisms to EfficientNet backbone. Specifically we used Squeeze-and-Excitation (SE) blocks and Multi-Head Attention (MHA).We Inserted SE blocks at multiple stages of the feature extractor with a reduction ratio of 16. This encouraged our Neural Net to emphasize on more informative features while suppressing less informative features, with minimal additional computational cost. Additionally, we also explored a global Multi-Head Self-Attention module with 4 heads.

2.4. Additional Strategies

To further improve robustness and generalization, we employed strong data augmentation during training of the data, which included: random horizontal flips, random vertical flips, rotations, affine transformations and color jittering. These augmentations helped to stop overfitting and to better represent variation in real world retinal images. At the time of inference, test-time augmentation (TTA) was also applied by averaging predictions from the original image and the image's horizontally flipped version. Additionally, per-class decision thresholds were set and experimented with on the offsite test set to further improve average F1-score.

Beyond CNN models we also explored a transformer model approach by using Swin Tiny. Swin-Tiny processes images in small regions and gradually combines information across the image, allowing it to learn local details and the general overall structure. This is useful for retinal images where disease signs can appear in different areas. All models were initialized with pretrained weights from ImageNet.

3. RESULTS

Model	DR (P / R / F1)	Glaucoma (P / R / F1)	AMD (P / R / F1)	Avg F1	Onsite Avg F1
EfficientNet (No FT)	0.746 / 0.650 / 0.695	0.577 / 0.612 / 0.594	0.246 / 0.773 / 0.374	0.554	0.604
EfficientNet (Frozen)	0.844 / 0.814 / 0.829	0.719 / 0.469 / 0.568	0.579 / 0.500 / 0.537	0.645	0.744
EfficientNet (Full, BCE)	0.842 / 0.836 / 0.839	0.698 / 0.612 / 0.652	0.556 / 0.682 / 0.612	0.701	0.809
+ Focal Loss	0.909 / 0.786 / 0.843	0.755 / 0.816 / 0.784	0.469 / 0.682 / 0.556	0.728	0.854
+ Class-Balanced Loss	0.700 / 1.000 / 0.824	0.245 / 1.000 / 0.394	0.110 / 1.000 / 0.198	0.472	0.438
+ SE Attention	0.901 / 0.843 / 0.871	0.765 / 0.796 / 0.780	0.519 / 0.636 / 0.571	0.741	0.839
+ MHA	0.904 / 0.879 / 0.891	0.889 / 0.653 / 0.753	0.625 / 0.682 / 0.652	0.765	0.821
Swin-Tiny	**0.910 / 0.880 / 0.895**	**0.851 / 0.722 / 0.780**	**0.620 / 0.610 / 0.615**	**0.763**	**0.826**

Table 3.1

Table 3.1 summarizes the performance of different models. Among thetested models, full fine-tuning of EfficientNet significantly improved performance over no fine-tuning or a frozen backbone. Adding attention mechanisms boosted the results: Squeeze-and-Excitation improved overall balance, while Multi-Head Attention achieved the highest average F1 for the offsite test (0.765) and the best AMD F1 for the offsite test (0.652), indicating strong performance on minority classes. Focal Loss improved F1 the most for the onsite test while for the offsite test it increased Glaucoma detection but slightly reduced AMD F1. Class-Balanced Loss led to very low AMD precision despite high recall. The Swin-Tiny transformer performed comparably to the best CNN with attention, achieving a balanced performance across all diseases and the highest onsite average F1 (0.826), showing good generalization. AMD F1 remained lower than the other classes due to the very small number of AMD samples, which limited the model's ability to balance precision and recall.

4. DISCUSSION

Comparing different deep learning techniques we see that full fine-tuning was important, as models without it performed poorly, especially on AMD. Among loss functions, Focal Loss worked best, improving the overall F1 and specifically glaucoma detection, while Class-Balanced Loss often over-predicted rare classes specifically AMD, giving very low precision.

Attention mechanisms helped all models: Squeeze-and-Excitation improved the score as a result of its channel-wise feature focus and provided stable gains, while Multi-Head Attention helped in capturing long-range retinal patterns, boosting AMD and DR performance. The Swin-Tiny transformer performed well across all classes and generalised best to new data. AMD performance remained lower than other diseases but compared to the CNN models performed significantly better. This highlighted the need for more data, data augmentation or hybrid CNN–Transformer models in future work.

5. CONCLUSION

Focal loss with efficientnet full fine-tuning provided the best balance for moderate class imbalance. Attention mechanisms enhanced feature representation, with se offering stable gains. The best model achieved 0.854 onsite f-score, exceeding the reference by 5%.