Masked Autoencoders for Retina Blood Vessel Segmentation

John Yechan Jo, Qiwen Xiao, Wenbo An

Introduction

Retinal blood vessel segmentation is crucial in diagnosing various ocular and systemic diseases, including diabetic retinopathy and glaucoma. Traditional supervised learning approaches require extensive annotated datasets, which can be labor-intensive and expensive to obtain.

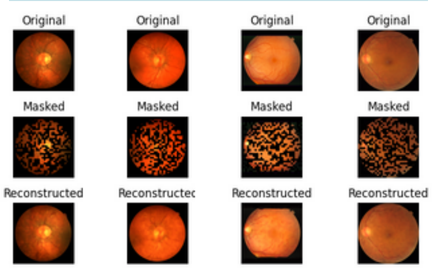
To address this challenge, we propose a **Masked Autoencoder (MAE)-based self-supervised learning approach** for retinal vessel segmentation. Additionally, we incorporate **Focal Tversky Loss** to handle class imbalance and **Learning Rate Reduction on Plateau** to adaptively adjust learning rates during training.

Methodology

Our approach consists of **MAE-based pretraining**, **vessel segmentation using the pretrained encoder**, and **adaptive training techniques to improve performance**.

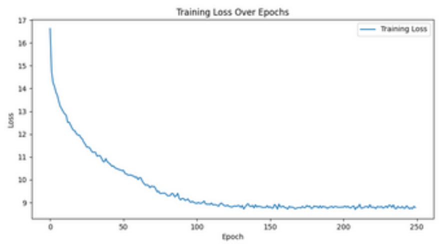
MAE for Feature Learning

- A **ResNet-inspired encoder** extracts meaningful hierarchical features.
- The decoder reconstructs **randomly masked patches** to enforce robust vessel representation learning.
- The **pretrained MAE encoder** is later fine-tuned for segmentation.



Retinal Vessel Segmentation

- The **pre-trained MAE encoder** serves as the backbone for a **segmentation model**.
- The decoder progressively **upsamples** latent representations to reconstruct vessel structures.
- **Dataset Structure**: Retinal vessel images from a structured dataset (train, test, validation).



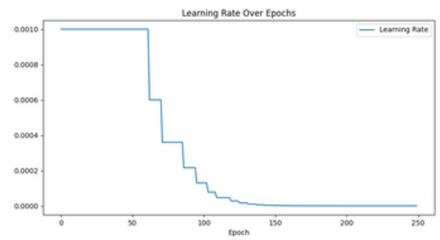
Focal Tversky Loss for Class Imbalance

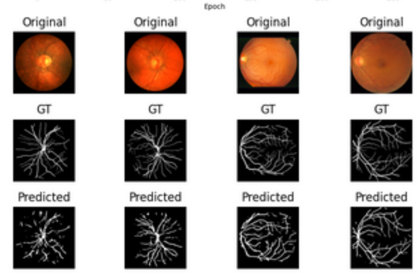
- Retinal vessel datasets are highly imbalanced, with vessels occupying a small fraction of the image.
- **Focal Tversky Loss** balances **false positives and false negatives** while focusing more on difficult-to-classify pixels.

$$L_{FT} = \sum_{c=0}^{C-1} (L_T^c)^{1/\gamma}$$

Learning Rate Reduction on Plateau

- Dynamically reduces the learning rate when validation loss plateaus.
- Prevents overfitting and ensures smoother convergence.





Future Work

Dataset Expansion & Domain Adaptation

- Augment training data and improve cross-domain generalization.

Enhancing Model Performance

- Incorporate **attention mechanisms** such as **Vision Transformers (ViTs).**

Clinical Application & Deployment

- Optimize for real-time processing in clinical settings.

Conclusion

This study demonstrates that **MAE-based self-supervised learning** effectively improves retinal vessel segmentation by learning meaningful vessel structures before fine-tuning. The use of **Focal Tversky Loss** helps handle class imbalance, ensuring better segmentation of thin and complex vessels, while **LR Reduce on Plateau** dynamically adjusts the learning rate to prevent overfitting and stabilize training. These techniques together enhance model robustness and generalization.

Future work will focus on **expanding the dataset**, **integrating attention mechanisms**, and **optimizing real-time deployment** for clinical applications. With further improvements, this approach has the potential to significantly contribute to **automated retinal disease diagnosis, aiding early detection and treatment planning in ophthalmology.**

Reference

1. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2022). Masked Autoencoders Are Scalable Vision Learners. arXiv preprint arXiv:2111.06377.
2. Salehi, S. S. M., Erdogmus, D., & Gholipour, A. (2017). Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks. arXiv preprint arXiv:1706.05721.
3. Prechelt, L. (1998). Early Stopping – But When? Neural Networks: Tricks of the Trade.
4. Wang, Z., Chen, J., & Zhang, H. (2023). Masked Autoencoder Pretraining for Medical Image Analysis: A Review and Benchmark. arXiv preprint arXiv:2312.05391.