TransResNet: Integrating the Strengths of ViTs and CNNs for High Resolution Medical Image Segmentation via Feature Grafting

Muhammad Hamza Sharif
muhammad.sharif@mbzuai.ac.ae
Dmitry Demidov
dmitry.demidov@mbzuai.ac.ae
Asif Hanif
asif.hanif@mbzuai.ac.ae
Mohammad Yaqub
mohammad.yaqub@mbzuai.ac.ae
Min Xu ✉
xumin100@gmail.com

Mohamed Bin Zayed University
of Artificial Intelligence, UAE

Appendix

In this supplementary document, we provide more information about the dataset we used, additional results from the experiments, that we did not mention in paper due to page limitation constraints. We believe this supplementary information will help the scientific community to understand and reproduce our conducted research in a better way.

A Datasets

We have used ten datasets for three different segmentation tasks: **(a)** skin lesion segmentation (2 datasets), **(b)** retinal vessel segmentation (3 datasets), and **(c)** polyp segmentation (5 datasets). An overview of all these publicly available dataset in mentioned in the Table 1.

Dataset	Average Resolution	Train Samples	Test Samples
Skin lesion			
ISIC-2016	1468×1070	900	379
PH2	766×575	0	200
Retinal Vessel			
HRF	3504×2306	30	15
IOSTAR	1024×1024	20	10
CHASE	999×960	20	8
Polyp			
Kvasir	618×539	838	100
ClinicDB	384×288	612	62
ColonDB	574×500	0	380
Endoscene	574×500	0	60
ETIS	1225×966	0	196

Table 1: An overview of dataset used in the paper for three segmentation tasks.

B Probability Correction Strategy (PCS)

We have applied Probability Correction Strategy (PCS) [1] during inference to improve the final prediction. During the training pipeline, the number of negative samples (background pixels) is greater than the number of positive samples (foreground pixels), which leads the model to produce an unsharp and noisy output. We enhance the final output through logits re-weighting. In PCS, we count the number of samples of each class (positive and negative pixels) before sigmoid function and normalize the logits of each class with the total count of corresponding class. Figure 1, demonstrates the visualizations of predicted mask before and after applying PCS.

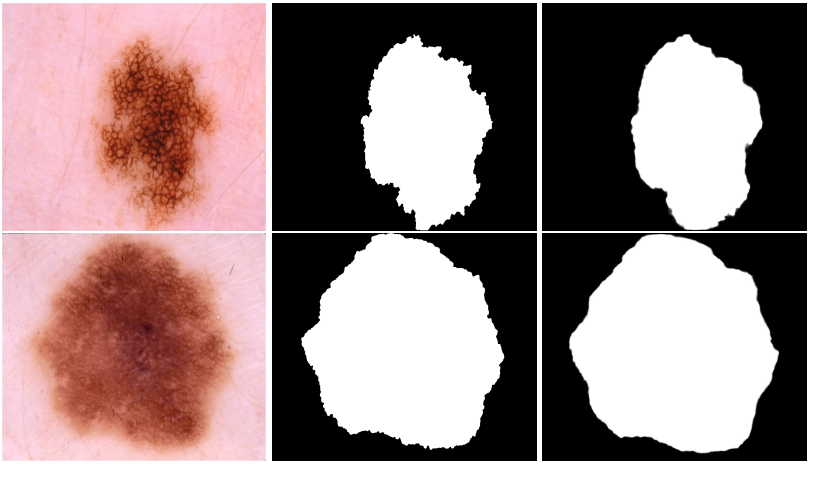
IoU= 0.835 *IoU*= 0.892 *IoU*= 0.839 *IoU*= 0.902

Predicted Mask without PCS **Predicted Mask with PCS**

Figure 1: Visualizations of predicted mask with PCS. The left side two images shows the predicted mask without applying PCS while right side represents the sharp mask after applying PCS. The IoU scores represents that performance increases after applying PCS.

C Additional Qualitative Results

We have provided the additional qualitative results of our method for all three segmentation tasks. Fig. 2, shows the qualitative results of skin lesion segmentation, while Fig. 3 and Fig. 4, illustrate the visualization of retinal vessel and polyp segmentation tasks respectively.



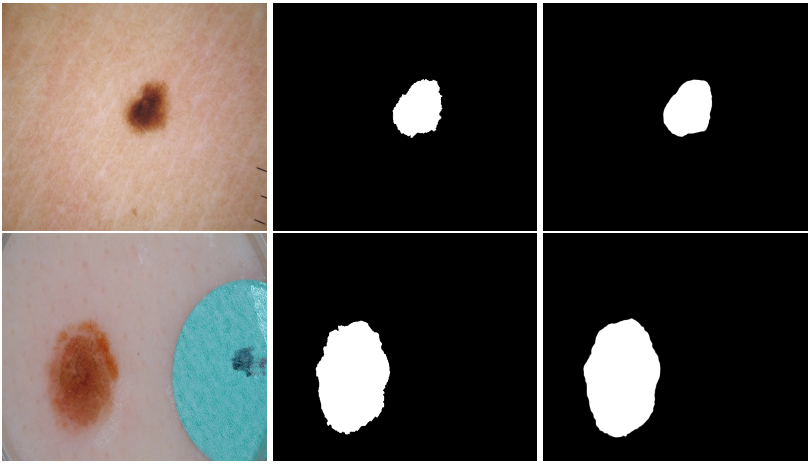


Figure 2: Visualizations of skin lesion segmentation

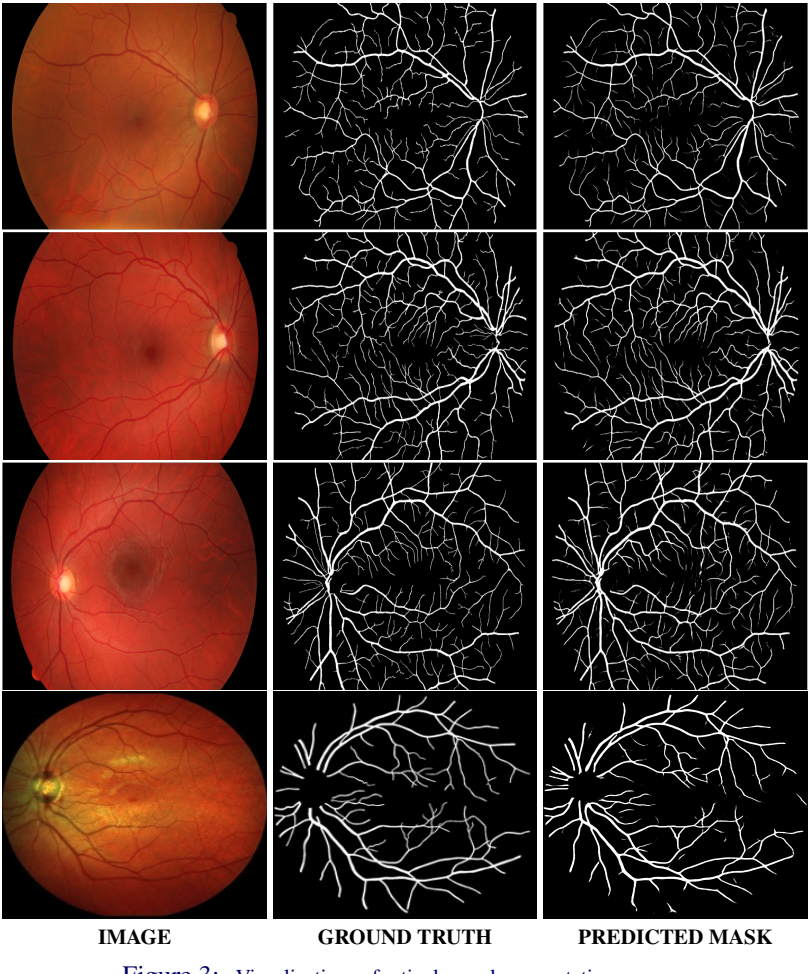


Figure 3: Visualizations of retinal vessel segmentation

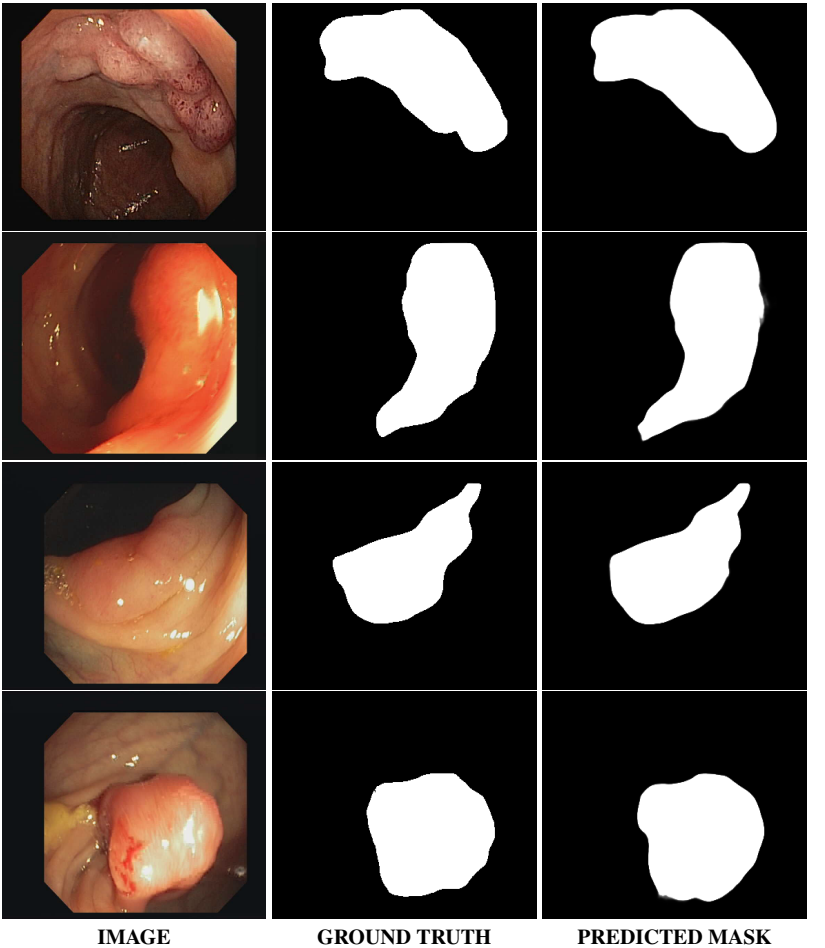


Figure 4: Visualizations of polyp segmentation

References

[1] Jun Wei, Yiwen Hu, Ruimao Zhang, Zhen Li, S Kevin Zhou, and Shuguang Cui. Shallow attention network for polyp segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 699–708. Springer, 2021.