Anatomy-Aware Unsupervised Detection and Localization of Retinal Abnormalities in Optical Coherence Tomography

Supplementary Material

Table 6. Segmentation performance (mean Dice / mIoU) over 327 RETOUCH images using different reconstruction error metrics. The weighted metric corresponds to $\mathcal{E}(x, \hat{x}) = 0.6|x - \hat{x}| + 0.4(1 - \text{SSIM})$.

Model	Weighted	L1	SSIM	MSE
		VQGAN		
Dice	**0.1449**	0.1200	0.1003	0.1199
mIoU	**0.0823**	0.0670	0.0575	0.0670
		VQVAE		
Dice	**0.1846**	0.1574	0.1099	0.1573
mIoU	**0.1073**	0.0925	0.0623	0.0924
		ConvNeXt		
Dice	**0.1846**	0.1574	0.1099	0.1573
mIoU	**0.1073**	0.0925	0.0623	0.0924
		Ours		
Dice	**0.1998**	0.1795	0.1572	0.1786
mIoU	**0.1173**	0.1058	0.0962	0.1052

6. Reconstruction Metric Analysis for Anomaly Segmentation

This appendix provides a detailed analysis of the impact of different reconstruction discrepancy measures on anomaly localization performance. While the main paper reports segmentation results using our weighted reconstruction formulation, here we compare several commonly used pixel-wise error metrics, including ℓ_1, mean squared error (MSE), structural dissimilarity (1SSIM), and a weighted combination of ℓ_1 and SSIM defined as

$$\mathcal{E}(x, \hat{x}) = 0.6|x - \hat{x}| + 0.4(1 - \text{SSIM}(x, \hat{x})).$$

Table 6 reports mean Dice and mIoU scores over all 327 RETOUCH test images with available ground-truth annotations. The comparison is conducted across multiple reconstruction-based architectures, including VQGAN, VQVAE, ConvNeXt, and different variants of our Tri-VQGAN model.

This analysis serves two purposes. First, it demonstrates how the choice of reconstruction discrepancy metric influences segmentation quality, highlighting the trade-off between pixel-level intensity differences and structural similarity cues. Second, it justifies the use of our weighted formulation by showing its consistent and competitive performance across model variants.