DiffuPT: Class Imbalance Mitigation for Glaucoma Detection via Diffusion Based Generation and Model Pretraining
Supplementary Materials

A. Conditional Vs Unconditional generation

Two generative techniques were considered to generate glaucoma samples. First, we tried to generate glaucoma images only using unconditional generation. This can be achieved by training DM on glaucoma images only. However, sample quality was substandard due to the lack of enough glaucoma samples, thus we tried to generate samples conditioned on the class label. This can be achieved by training the DM on both glaucoma and non-glaucoma images. The class conditioning can be achieved during sampling using the equation 1. During sampling, we specify the target class and unconditional guidance scale $w = 3.0$ to control the quality of generation. The discrepancies in the results can be further noticed in the quality of the image produced. This can be clearly shown in Figure 1.

$$\tilde{\epsilon}_\theta(z_t, c) = (1 + w)\epsilon_\theta(z_t, c) - w\epsilon_\theta(z_t, \emptyset) \quad (1)$$

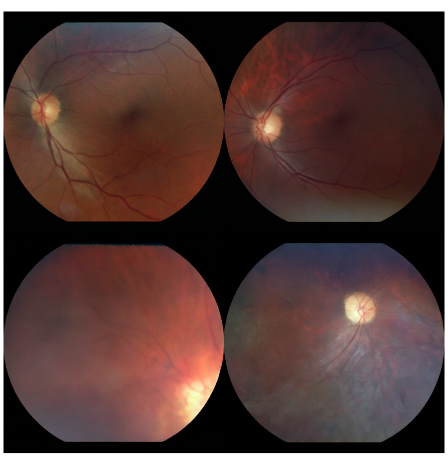

Figure 1. The top images are from conditional LDM and the bottom images show the unconditional generation.

B. Failed Attempts

We show failed attempts for fundus image generation. We attempted to generate images conditionally using Conditional VAE [1]. We tried different encoder-decoder architectures but none was successful. We show images of Conditional VAE in Figure 2.

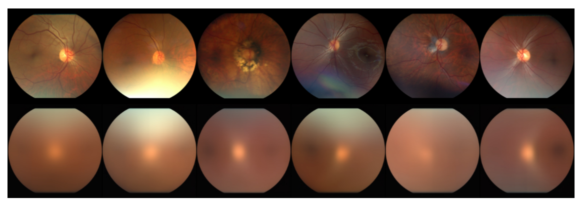

Figure 2. The top images are reconstruction images. The bottom images shows the generated samples.

C. Generation

We show more samples of the generation in Figure 3, MWGAN contains more artifacts than other models. LDM samples show the best quality and clearest images.

D. Results analysis

We show confusion matrices for normal Training, augmentation method, and diffuPT. The confusion matrices in Figure 4 show that diffuPT doesn't favor one eye over the other (left and right). One general trend noticed that there wasn't any clear bias of any generative model for the left eye over the right.

References

[1] Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning structured output representation using deep conditional generative models. *Advances in neural information processing systems*, 28, 2015. 1

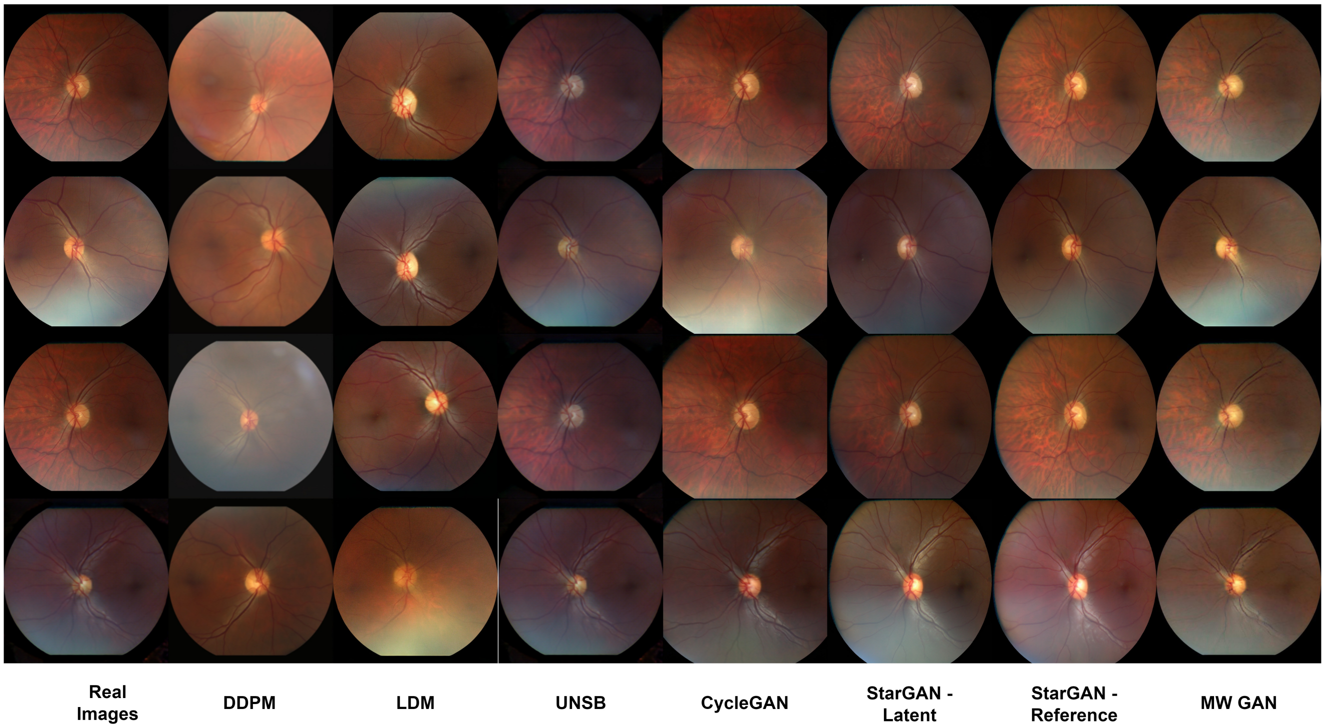

Figure 3. More Generated Samples.

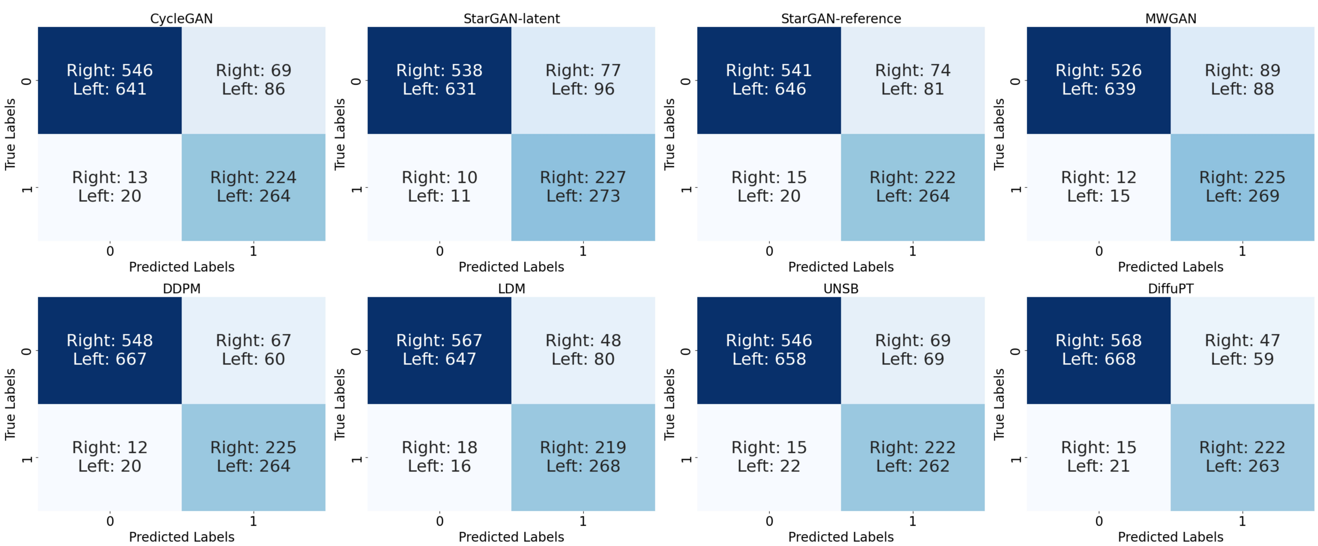

Figure 4. Confusion Matrices showing the right eyes separately from the left eyes.