Generating Synthetic Fundus Images and Masks of Pathological Myopia

EBUNOLUWA MAKINDE, University of Calgary
AYESHA SAEED, University of Calgary

Pathological Myopia (PM) is a severe eye condition and is a leading cause of blindness globally. The increasing rates of myopia highlight the need for early detection and intervention for PM, though this remains challenging due to the lack of large, well-annotated datasets for training AI models. This study investigates the potential of using generative models, such as Deep Convolutional Generative Adversarial Networks (DCGAN) and Denoising Diffusion Probabilistic Models (DDPM), to create synthetic Fundus images and corresponding segmentation masks. These models were trained on the PALM and ADAM datasets, with specific enhancements aimed at improving image quality and training stability. The DCGAN successfully generated Fundus images along with masks for the optic disc and atrophy, while the DDPM encountered resource limitations, which hindered its ability to produce realistic and high-resolution images. Our results indicate that synthetic image generation could be valuable for augmenting datasets for PM and enhancing AI model training, though further work is necessary to optimize these models for real-world clinical use.

CCS Concepts: • **Computer systems organization → Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

Additional Key Words and Phrases: datasets, generative AI, pathological myopia, fundus images

ACM Reference Format:
Ebunoluwa Makinde and Ayesha Saeed. 2018. Generating Synthetic Fundus Images and Masks of Pathological Myopia. *ACM Trans. Graph.* 37, 4, Article 111 (August 2018), 6 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Pathological Myopia (PM), a severe complication of progressive myopia, is a critical issue in ophthalmology and a leading cause of blindness globally, ranking as the seventh most common cause [Hemelings et al. 2021]. The prevalence of myopia is rapidly increasing, with projections suggesting it will affect half of the world's population by 2050 [Oliveira et al. 2024]. Early detection and intervention in PM are essential to mitigate its progression and prevent blindness.

The field of computer vision has witnessed rapid advancements in automating medical diagnoses, particularly with deep-learning models. However, a persistent challenge is the need for large, balanced, annotated datasets essential for training robust models. In the case of PM, this limitation hinders the reliability and clinical applicability of AI tools as there are not that many large datasets available. Augmenting datasets with synthetic images generated using advanced techniques, such as generative adversarial networks (GANs) and diffusion models, presents a promising solution to this problem.

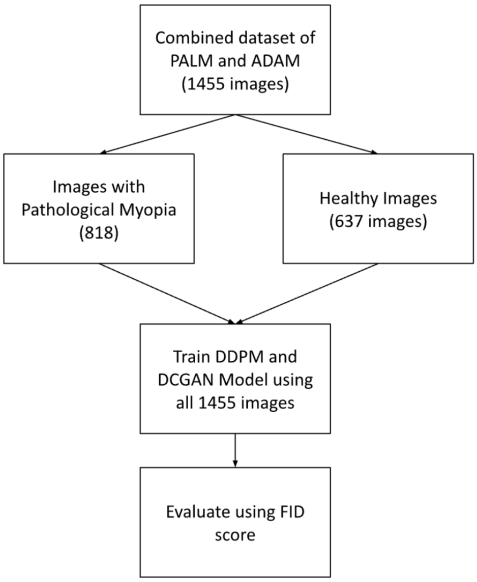

Fig. 1. Overview of our Methodology

Several approaches have already been proposed and performed to tackel the challenge of limited medical datasets, particularly in ophthalmology. Ahn et al. introduced FundusGAN, which generates synthetic fundus images and segmentation masks for diabetic retinopathy and macular degeneration [Ahn et al. 2023]. Similarly, models like StyleGAN2-ADA and R-sGAN have been effective in creating diverse, clinically relevant datasets for retinal diseases [Oliveira et al. 2024; Zhao et al. 2018]. However, while such models have been applied to diseases like AMD and diabetic retinopathy, applying these techniques to PM remains underexplored.

On the other hand, diffusion models have emerged as strong competitors to GANs in recent years, particularly in the realm of generating synthetic eye fundus images. Kim et al. explored the use of a denoising diffusion probabilistic model (DDPM) for generating fundus photographs using a small training dataset [Kim et al. 2024]. Additionally, Alimanov and Islam utilized a DDPM to generate both fundus images and vessel tree masks, addressing challenges related to manual annotation and data scarcity [Alimanov and Islam 2023]. Both studies have successfully produced high-quality fundus images across varying dataset sizes but have not explored the generation of masks.

Therefore, the goal of our project is to research into the feasibility of a novel approach to generating optic and atrophy masks tailored

Authors' addresses: Ebunoluwa Makinde, University of Calgary; Ayesha Saeed, University of Calgary.

0730-0301/2018/8-ART111 $15.00
https://doi.org/XXXXXXX.XXXXXXX

for PM. Our objective is to create a pipeline for the generation of high-quality realistic fundus images and corresponding segmentation masks that align with the characteristics of PM. The second is to have an improved dataset diversity, which can contribute to more balanced and comprehensive training datasets for automated PM detection tools.

2 DATASET

The dataset utilized in this project consists of two primary sources: the PALM dataset and the ADAM dataset, both obtained from the Zhongshan Ophthalmic Center at Sun Yat-sen University, China. These datasets are recognized for their extensive and high-quality ophthalmic data. There is some bias present in both datasets as they both come from Chinese patients which may not represent the greater world population. [Fang et al. 2022, 2024].

2.1 PALM and ADAM Dataset

The **PALM** dataset was released as part of the PAthological Myopia challenge which was held in conjunction with the International Symposium on Biomedical Imaging (ISBI) in 2019 [Fang et al. 2024]. The dataset contains 1200 fundus images, equally divided 1:1:1 subsets (training, validating, and testing) where each section had 400 images. This dataset primarily focuses on non-AMD (Age-Related Macular Degeneration) samples, as the disease of interest in this study is Pathological Myopia (PM) [Fang et al. 2024]. Other retinal diseases such as retinopathy, myopia, and glaucoma are also represented in these samples [Fang et al. 2024]. Delineations of the optic disc and fundus lesions and annotation of the fovea localization were done manually by seven ophthalmologists with an average of 8 years of experience and one senior ophthalmologist with more than 10 years of experience [Fang et al. 2024].

The **ADAM** dataset was introduced as part of the Automatic Detection Challenge on Age-Related Macular Degeneration (AMD), held as a satellite event during the ISBI 2020 conference [Fang et al. 2022]. The dataset comprises 1200 fundus images, equally divided into training, validation, and testing subsets. Manual pixel-wise annotations of the optic disc and five lesions (drusen, exudate, hemorrhage, scar, and other lesions) were provided by 7 ophthalmologists [Fang et al. 2022]. The coordinates of the fovea were also obtained by the 7 ophthalmologists and the final coordinate was determined by averaging the 7 annotations, and a senior ophthalmologist performed a quality check afterward [Fang et al. 2022]. Any errors observed in the final coordinates were modified using a method similar to the error processing method in the segmentation annotating process [Fang et al. 2022]. For the ADAM dataset, we only utilize the non-AMD samples in the dataset as our disease of interest is PM.

2.2 Dataset Preprocessing and Augmentation Techniques

To enhance the datasets' robustness and ensure optimal model training, comprehensive preprocessing and augmentation strategies were employed:

We curated a custom PyTorch dataset for both the PALM and ADAM datasets. Images without corresponding segmentation masks were excluded. The curated dataset is structured to return the image, its associated mask, and its label. Both datasets were combined into a single dataset to streamline the training process. All images were resized to a standardized resolution compatible with the input size of the model. Pixel intensity values were normalized to enhance contrast and ensure consistent data input for the model. An additional channel was appended to each image to enable the simultaneous generation of segmentation masks alongside the fundus images.

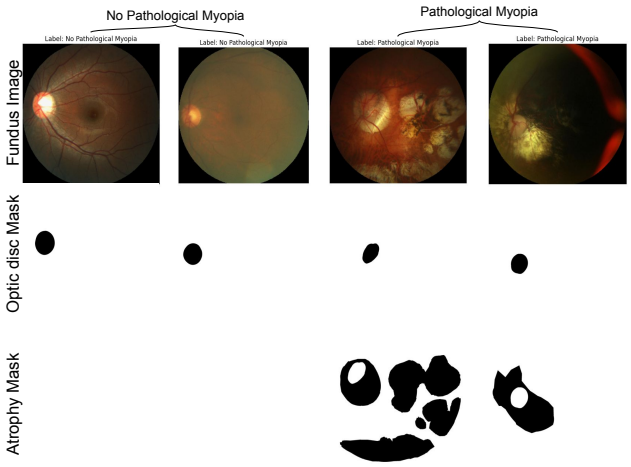

Fig. 2. Visualisations of Fundus images of patients with and without Pathological Myopia. Masks of the optic disc and atrophy present in disease patients are displayed (Only diseased patients have atrophy masks as it signifies the presence of disease).

3 MODEL ARCHITECTURE

As seen across literature, when exploring deep learning models for eye fundus image synthesis both GAN and diffusion methods are explored [Alimanov and Islam 2023; Kim et al. 2024]. Hence, we will be employing both GAN and diffusion-based models for our task.

3.1 Deep Convolutional Generative Adversarial Network (DCGAN)

DCGAN is a specialized form of GAN that employs convolutional layers in its generator and discriminator networks to synthesize high-quality images. Key features of DCGAN include:

(1) Fully convolutional generator with upsampling layers instead of dense layers.
(2) Fully convolutional discriminator with strided convolutions for downsampling.
(3) Batch normalization is used to stabilize training and improve convergence.
(4) ReLU activation in the generator and LeakyReLU in the discriminator for better gradient flow.

3.2 Denoising Diffusion Probablisitc Models (DDPM)

DDPM is a class of generative models that iteratively transform random noise into structured images through a diffusion process. Key features of DDPM include:

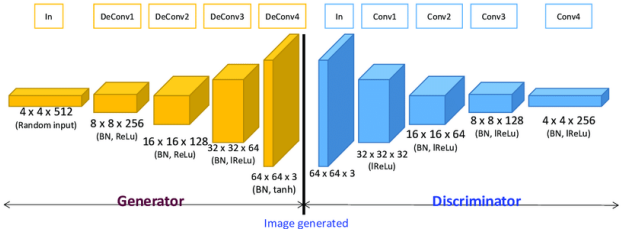

Fig. 3. Illustration of a Sample DCGAN Architecture [Suh et al. 2019]

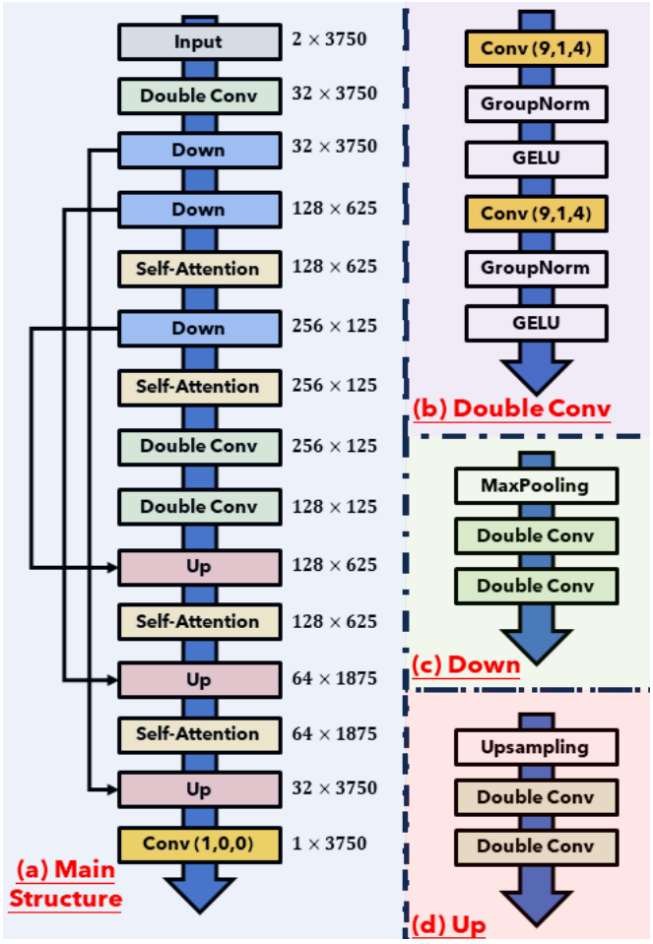

Fig. 4. Illustration of a Sample DDPM Architecture [Li et al. 2023]

(1) A forward process that gradually adds Gaussian noise to an image in T steps.
(2) Predicts and removes the noise step-by-step using a learning neural network (e.g. U-Net).
(3) Conditional generation capabilities using class labels.
(4) Improved stability and image quality with the incorporation of techniques like exponential moving average (EMA) for weights and self-attention mechanisms.

Table 1. DCGAN Model Hyperparameters

Metric	
Batch Size	32
Image size	360 by 360
Number of Channels	4 (3 RGB, 1 Mask)
Latent Dimension(nz)	100
Generator Features	32
Discriminator Features	64
Number of Epochs	100 and 130(with atrophy)
Learning Rate (lr)	0.0002
Optimizer	Adam
Beta1(Adam	0.5
Fixed Noise Vector	64 Samples

4 TRAINING

4.1 DCGAN

The DCGAN was modified from [Jain 2024] which was also used to generate synthetic fundus images of eyes. The training for DCGAN utilized PyTorch on Google Colab with T4 GPU specifications. The training process involved alternating updates between the Generator and the Discriminator in a two-player adversarial game. The Discriminator was trained to distinguish between the real and generated data whereas the Generator was trained to generate data that the Discriminator classifies as real. Each epoch consisted of multiple batches where gradients for the Discriminator and Generator were backpropagated and updated using their respective optimizers. Figure 1 outlines the hyperparameters used during the training.

The optimization of both networks was performed using the Adam optimizer with a learning rate of 0.0002 and a momentum term (Beta1) of 0.5 The Binary Cross-Entropy Loss (BCE Loss) function was employed to compute the loss for both real and fake data during Discriminator and Generator updates.

To ensure the stability of the training process and mitigate overfitting, a few techniques were employed. All weights in the networks were initialized using the Xavier initialization technique to ensure proper signal propagation. Batch Normalization was applied after each convolution layer to stabilize learning and normalize activations.

We conducted two different model runs. The first utilized a DCGAN to generate both eye fundus images and their corresponding optic disc masks. The second approach focused on generating eye fundus images using the optic disc and atrophy masks.

For training, we used concatenated images, where the masks and images were combined as described in the dataset section. Additionally, we modified the models to accommodate input and output channels, allowing them to process 4-channel or 5-channel images instead of the standard 3-channel format. Consequently, the output generated by these models was also in a 4-channel or 5-channel format.

The progress of the training process was evaluated using the Generator and Discriminator loss values. Discriminator Loss (D) measures the ability to classify real and fake images. Generator Loss (G) measures the ability to generate realistic images that fool

Table 2. DDPM Model Hyperparameters

Metric	
Image Size	64 by 64
Batch Size	8
Noise Steps (T)	300
Beta Start	0.0001
Beta End	0.02
EMA Decay Rate	0.995
Learning Rate	3e-4

the Discriminator. At regular intervals, the losses were printed for monitoring, and images were generated using a fixed noise vector to assess the Generator's progress.

4.2 DDPM

The implementation of our DDPM was modified from [Ahmadi 2023; Rampas 2024] although based on the original DDPM implementation [Ho et al. 2020]. Training of the DDPM model was conducted on Google Colab. The model was trained over more than 100 epochs using the combined dataset preprocessed and augmented as described earlier. The training loop iteratively optimizes the model using mini-batches of data to reduce the mean squared error (MSE) loss. The Exponential Moving Average (EMA) technique was employed to stabilize the model's performance over training steps. 2.1 Optimization Algorithm and Hyperparameters The optimizer used was AdamW (Adam with Weight Decay) which provides robust performance by incorporating weight decay directly into the parameter updates to regularize the mode. A constant learning rate of 3e-4 was used throughout the training process to maintain stability. The loss function, Mean Squared Error (MSE) measured the discrepancy between the predicted noise and actual noise added to images during the forward diffusion. Table 2 displays the hyperparameters.

At each iteration of the training procedure, a random timestamp t was sampled. The input image was corrupted using the forward diffusion process and the model predicted the added noise at timestep t. The MSE loss was calculated between the predicted and actual noise. Model weights were updated using the AdamW optimizer and the EMA weights were updated.

5 EVALUATION

With the implemented DCGAN, we were able to generate both fundus images and their corresponding masks. Specifically, we generated images with a resolution of 360 x 360 pixels, including optic disc masks and combined optic disc and atrophy masks. The results were evaluated after 130 epochs for optic disc masks and images, and 100 epochs for images, atrophy, and optic disc masks. To quantitatively assess the performance, we employed the Fréchet Inception Distance (FID) metric to distinguish the generated images by their respective masks. The results, summarized in Table 3, revealed that the FID for the model generating both atrophy and optic masks was lower than that of the model. FID score for the model generating both atrophy and optic disc masks was lower than the model generating only optic disc masks, this indicated better performance. Qualitatively, we visually evaluated the generated images (not the masks) and noted that the model producing both atrophy and optic masks yielded more accurate results. However, the model generating only the optic masks demonstrated superior performance in generating the optic masks.

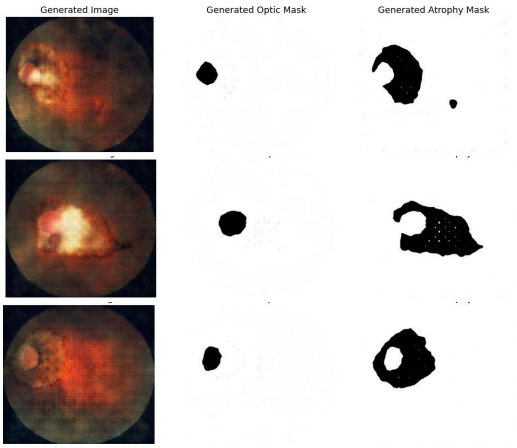

Fig. 5. Sample images generated by DCGAN; Images with Optic Disc and Atrophy Mask

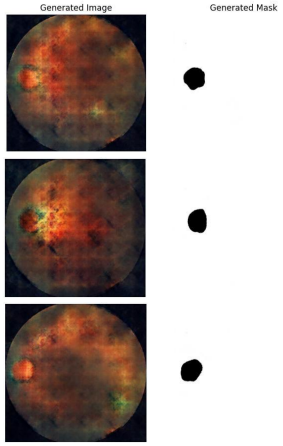

Fig. 6. Sample images generated by DCGAN; Images with Optic Disc Mask

Our DDPM model successfully generated fundus images and their corresponding masks at a resolution of 64 x 64 pixels. The low-resolution issue arose because the compute server we were using on Google Colab did not have enough RAM to perform the calculations required for the DDPM (Denoising Diffusion Probabilistic Model) when the image size exceeded a certain threshold. We tried methods such as reducing the batch size to mitigate this problem, but the RAM still proved insufficient to train the DDPM on

Table 3. Results of FID Score for DCGAN

Images Generated	Number of Epochs	FID Score
Image and Optic Disc Mask	100	156.24
Image, Optic Disc Mask, and Atrophy mask	130	167.67

images larger than a resolution of 64 x 64 pixels. This was achieved using a training set of approximately 1,071 samples derived from the ADAM and PALM datasets. Figure 7 showcases sample images created using the DDPM. We chose not to use FID as our evaluation metric for this model, as it was observed to perform poorly under the given conditions. Limitations in our results stem from insufficient computational resources, which restricted the generation of higher-resolution images and prolonged training durations.

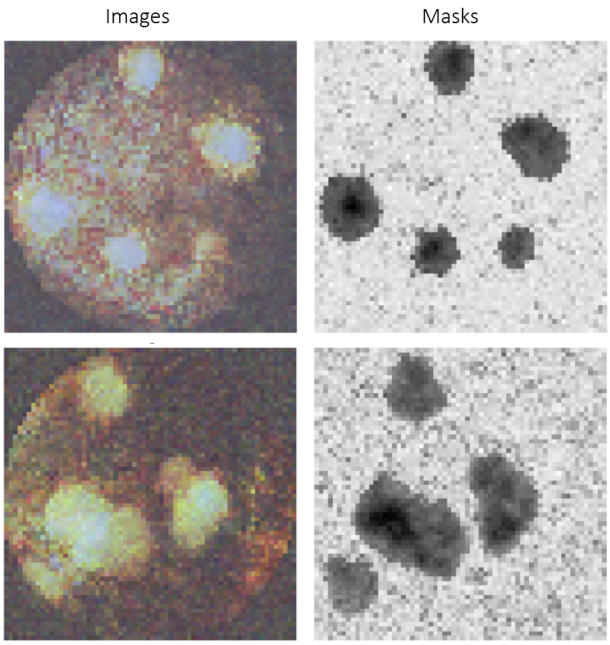

Fig. 7. Sample Images and Optic Disk Mask Generated from the DDPM

6 DISCUSSION

Our work demonstrates the feasibility of simultaneously generating masks and eye fundus images using deep learning methods for synthetic image generation. Specifically, we achieved this by training a model on 4- or 5-channel images, allowing it to learn from both the image and mask data simultaneously. While several studies have explored the use of generative AI technologies to produce high-quality eye fundus images, there remains a notable gap in models tackling the challenge of annotation limitations. Thus, our approach might serve as a launchpad for further research in this area. We explored the use of DCGAN and DDPM models for generating synthetic eye fundus images and their masks, particularly targeting the rare disease of pathological myopia (PM). Previous studies have successfully used generative models to produce medical images and masks, demonstrating close similarity to real images and enabling data augmentation in clinical imaging [Kim et al. 2024]. Our focus on PM is particularly pertinent, as it is a growing eye disease that has been underrepresented in studies compared to other conditions like glaucoma or age-related macular degeneration.

Using a DCGAN, we generated 1,455 synthetic images with optic disc masks and 818 synthetic images with both optic disc and atrophy masks. This was achieved without requiring manual or automatic segmentation, as the model generated masks concurrently with the images. To the best of our knowledge, this methodology is novel in its application to PM, a blindness-causing disease. This methodology can be applied to other computer vision medical applications. Despite demonstrating feasibility, our results are not yet optimal. The generated images still differ significantly from real images, as evidenced by relatively high FID scores. We attribute this to training limitations, including insufficient computational resources and training duration. We believe that extending the training duration could enhance the quality, as we consistently observed improvements in the model as training progressed.

Further research should aim to replicate our methods with extended training hours. Additionally, we recommend investigating post-processing techniques like style transfer, which can enhance the photorealism and coherence of the generated images, as suggested by prior studies [Saragih et al. 2024]. This step has the potential to significantly improve the realism and medical relevance of synthetic images, making them suitable for training segmentation models [Saragih et al. 2024]. To further explore state-of-the-art generative methods, we implemented a DDPM model. However, its performance was poor, primarily due to low-resolution input necessitated by RAM constraints and limited computational resources. Increasing the image resolution to 256 x 256 pixels and expanding the training dataset size could significantly enhance the performance of diffusion models. We recommend that future research address these limitations by utilizing hardware with higher RAM and GPU capabilities to enable more effective training and higher-quality results.

7 CONCLUSION

This project successfully explored the use of deep learning models, including DCGAN and DDPM, to generate synthetic fundus images and corresponding segmentation masks for Pathological Myopia (PM). The DCGAN model demonstrated strong performance in generating high-quality images and masks, with lower FID scores for generating both optic and atrophy masks compared to only optic disk masks. The DDPM model, while successful in generating fundus images, was limited by computational constraints, producing images at a lower resolution (64 x 64 pixels). In addition, the DCGAN model was able to generate images that looked more closely alike to the real images as compared to the DDPM model. The evaluation revealed that combining both optic disc and atrophy masks resulted in better-quality images.

This research contributes to advancing synthetic data generation in ophthalmology, particularly for PM, where large annotated

datasets are scarce. The ability to generate both images and segmentation masks opens avenues for increasing the accuracy of technology that automates the detection and analysis of PM. Future work will focus on improving computational efficiency and also gaining more computational resources so that the model can perform on higher image resolutions. Exploring hybrid models combining GANs and diffusion techniques may further improve image quality and mask accuracy.

REFERENCES

Raha Ahmadi. 2023. rahaahmadi/Diffusion-Model-Scratch. https://github.com/rahaahmadi/Diffusion-Model-Scratch original-date: 2023-08-29T08:05:16Z.

Sangil Ahn, Su Jeong Song, and Jitae Shin. 2023. FundusGAN: Fundus image synthesis based on semi-supervised learning. *Biomedical Signal Processing and Control* 86 (2023), 105289. https://doi.org/10.1016/j.bspc.2023.105289

Alnur Alimanov and Md Baharul Islam. 2023. Denoising Diffusion Probabilistic Model for Retinal Image Generation and Segmentation. arXiv:2308.08339 [eess.IV] https://arxiv.org/abs/2308.08339

Huihui Fang, Fei Li, H. Fu, Xu Sun, Xingxing Cao, Fengbin Lin, Jaemin Son, Sunho Kim, Gwénolé Quellec, Sarah Matta, Sharath M. Shankaranarayana, Yi-Ting Chen, Chuen-Heng Wang, Nisarg A. Shah, Chia-Yen Lee, Chih-Chung Hsu, Hai Xie, Baiying Lei, Ujjwal Baid, Shubham Innani, K. Dang, Wenxiu Shi, Ravi Kamble, Nitin Singhal, José Ignacio Orlando, Hrvoje Bogunovi'c, Xiulan Zhang, and Yanwu Xu. 2022. ADAM Challenge: Detecting Age-Related Macular Degeneration From Fundus Images. *IEEE Transactions on Medical Imaging* 41 (2022), 2828–2847. https://api.semanticscholar.org/CorpusID:246867031

Huihui Fang, Fei Li, Junde Wu, Huazhu Fu, Xu Sun, José Ignacio Orlando, Hrvoje Bogunović, Xiulan Zhang, and Yanwu Xu. 2024. Open fundus photograph dataset with pathologic myopia recognition and anatomical structure annotation. *Sci. Data* 11, 1 (Jan. 2024), 99. https://doi.org/10.1038/s41597-024-02911-2

Ruben Hemelings, Bart Elen, Matthew B. Blaschko, Julie Jacob, Ingeborg Stalmans, and Patrick De Boever. 2021. Pathological myopia classification with simultaneous lesion segmentation using deep learning. *Computer Methods and Programs in Biomedicine* 199 (2021), 105920. https://doi.org/10.1016/j.cmpb.2020.105920

Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems* 33 (2020), 6840–6851.

Mohit Jain. 2024. Natsu6767/DCGAN-PyTorch. https://github.com/Natsu6767/DCGAN-PyTorch original-date: 2018-12-13T15:32:36Z.

Hong Kyu Kim, Ik Hee Ryu, Joon Yul Choi, and Tae Keun Yoo. 2024. A feasibility study on the adoption of a generative denoising diffusion model for the synthesis of fundus photographs using a small dataset. *Discover Applied Sciences* 6, 4 (2024), 188.

Huayu Li, Xiwen Chen, Gregory Ditzler, William D.S. Killgore, Stuart F Quan, Janet Roveda, and Ao Li. 2023. Sleep Stage Classification with Learning from Evolving Datasets. https://doi.org/10.36227/techrxiv.23730546.v1

Guilherme C. Oliveira, Gustavo H. Rosa, Daniel C.G. Pedronette, João P. Papa, Himeesh Kumar, Leandro A. Passos, and Dinesh Kumar. 2024. Robust deep learning for eye fundus images: Bridging real and synthetic data for enhancing generalization. *Biomedical Signal Processing and Control* 94 (Aug. 2024), 106263. https://doi.org/10.1016/j.bspc.2024.106263

Dominic Rampas. 2024. dome272/Diffusion-Models-pytorch. https://github.com/dome272/Diffusion-Models-pytorch original-date: 2022-06-07T11:11:55Z.

Daniel Saragih, Atsuhiro Hibi, and Pascal Tyrrell. 2024. Using diffusion models to generate synthetic labeled data for medical image segmentation. *International journal of computer assisted radiology and surgery* 19 (06 2024). https://doi.org/10.1007/s11548-024-03213-z

Sungho Suh, Haebom Lee, Jun Jo, Paul Lukowicz, and Yong Oh Lee. 2019. Generative Oversampling Method for Imbalanced Data on Bearing Fault Detection and Diagnosis. *Applied Sciences* 9, 4 (Jan. 2019), 746. https://doi.org/10.3390/app9040746

He Zhao, Huiqi Li, Sebastian Maurer-Stroh, Yuhong Guo, Qiuju Deng, and Li Cheng. 2018. Supervised Segmentation of Un-Annotated Retinal Fundus Images by Synthesis. *IEEE Transactions on Medical Imaging* PP (07 2018), 1–1. https://doi.org/10.1109/TMI.2018.2854886

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009