

Enhanced Corneal Endothelial Cell Segmentation via Frequency-Selected Residual Fourier Diffusion Models

Tianyang Wang^{†1}, Xiaofei Nan^{†2}, Yunze Wang³, Yuhang Yan¹, Zhenkai Gao¹, Jingxin Liu^{*1}

¹Xi'an Jiaotong-Liverpool University, Suzhou, China

²Zhengzhou University, Zhengzhou, China

³University of Edinburgh, Edinburgh, UK

Abstract—Segmenting corneal endothelial cells in conditions like Fuchs endothelial dystrophy (FED) is challenging due to guttae obscuring cell details and complicating imaging, compounded by labor-intensive manual annotations and a lack of large annotated datasets. To address these issues, we introduce a novel two-stage framework using Denoising Diffusion Probabilistic Models (DDPMs) for generating corneal endothelial cell image training pairs. In the first stage, our method generates synthetic endothelial labels noise, used in the second stage to create high-resolution corneal images. We also present the Fourier Residual Block with Frequency Selection (FRB-FS), which enhances important high-frequency details for clearer textures and edges while suppressing irrelevant low-frequency components. This is the first application of diffusion models to corneal endothelial cell segmentation. Extensive experiments and ablation studies on two benchmark datasets validate the effectiveness of our framework.

Index Terms—Medical Image Analysis, Denoising Diffusion Probabilistic Model, Semantic Segmentation

I. INTRODUCTION

Accurate segmentation of medical images not only plays a critical role in disease diagnosis and treatment monitoring, but also remains a challenging task due to the complexity of anatomical structures and significant variations in imaging conditions [1]–[3]. In the case of Fuchs Endothelial Dystrophy (FED), the presence of guttae not only further complicates the segmentation of the corneal endothelium but also significantly affects the overall accuracy [4]. Guttae, which are collagen deposits resembling droplets on Descemet’s membrane, severely obscure cellular features, thus reducing the efficacy of traditional segmentation approaches [5]–[7]. Automated segmentation techniques often exhibit poor reliability, especially when applied to healthy corneal images. Moreover, manual annotation is not only time-consuming and expensive, but also the limited availability of annotated datasets further intensifies the challenge.

With the advancements in computer vision, automatic segmentation algorithms for corneal endothelial cells have become increasingly important, aiming to replace labor-intensive manual methods [8]–[10]. Traditional approaches that rely on frequency and morphological analysis often struggle under complex conditions such as those presented by Fuchs endothelial dystrophy [11]–[13]. While deep learning models, exemplified by the U-Net [14] and newer structural adaptations [15], offer promising directions, their efficacy is significantly hampered by the need for extensive, accurately labeled training

data. To mitigate this issue, recent studies have explored the use of generative adversarial networks (GANs) to synthesize realistic corneal endothelium images, thereby artificially expanding the available training datasets [16]. Despite their potential to transform the field, GAN-based methods still face critical challenges such as pronounced image noise and substantial training instability, including issues with non-convergence [17]–[19].

In this paper, we propose a novel framework leveraging Denoising Diffusion Probabilistic Models (DDPMs) to generate training image pairs for corneal endothelial cell segmentation. We also introduce a Fourier Residual Block with Frequency Selection (FRB-FS) that enhances image quality by focusing on high-frequency components essential for detailed textures and sharp edges. Our framework requires only a limited number of real corneal endothelial cell label images for training, yet it efficiently generates an unlimited number of high-resolution, high-quality image pairs from noise, significantly reducing the reliance on large annotated datasets. To the best of our knowledge, this is the first work utilizing DDPMs for corneal endothelium image synthesis.

Our contributions are as follows: (1) We propose a novel two-stage Denoising Diffusion Probabilistic Model (DDPM) framework for generating paired synthetic training data for corneal endothelial cell segmentation. (2) We introduce the Fourier Residual Block with Frequency Selection (FRB-FS) to enhance high-frequency details and reduce irrelevant low-frequency components, leading to improved image quality. (3) We validate our approach through extensive experiments on two benchmark datasets, demonstrating superior performance compared to existing methods.

II. METHODOLOGY

Overview. The proposed framework for generating corneal endothelial cell image pairs follows a two-stage process: label synthesis and image generation. First, endothelial cell labels are generated from random noise, which are then transformed into high-resolution images. This approach effectively creates realistic, high-quality image pairs for training, as shown in Fig.1.

A. Corneal Endothelium Label Synthesis

In corneal endothelium label synthesis, the model performs unconditional DDPM to generate synthetic endothelial labels

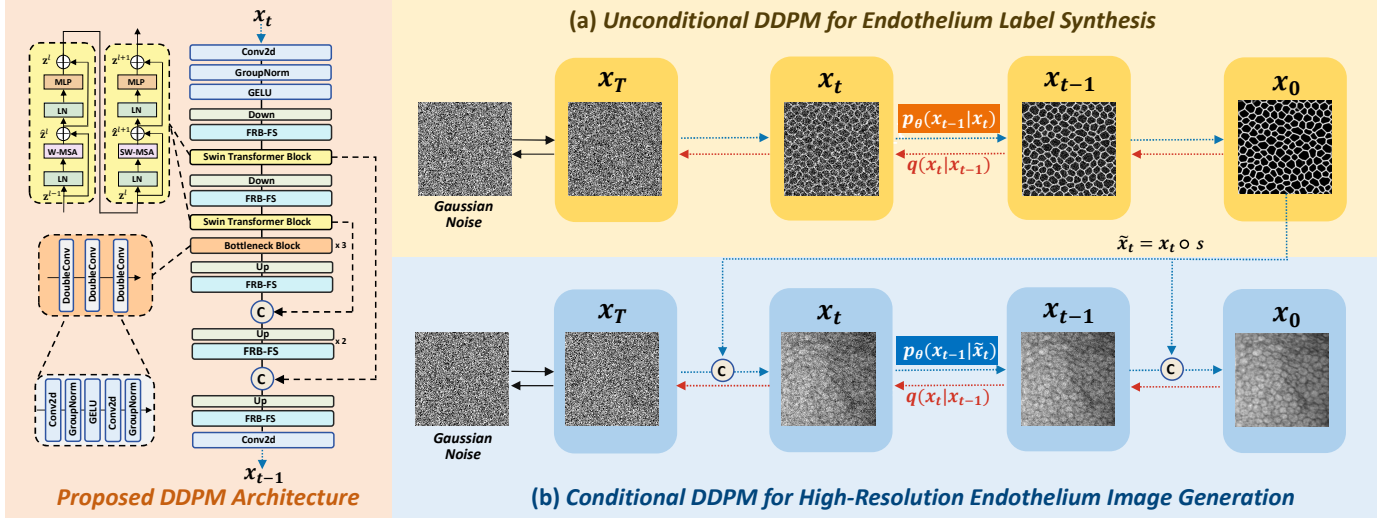


Fig. 1: Overview of our framework with (a) Unconditional DDPM for generating endothelium labels and (b) Conditional DDPM for generating high-resolution endothelium images.

from random Gaussian noise. The DDPM generates samples in two stages: the forward process and the reverse process. In the forward process, Gaussian noise is gradually added to the input image through a series of diffusion time steps T . At each step t , the image x_t depends only on the previous step x_{t-1} , making this a Markovian chain. The noise is added using the variance schedule β_t , defined as:

$$q(x_{1:T}|x_0) = \prod_{t=1}^T q(x_t|x_{t-1}) \quad (1)$$

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t I) \quad (2)$$

where $\mathcal{N}(\cdot; \cdot, \cdot)$ represents a Gaussian distribution.

In the reverse process, the model progressively denoises the image x_T step-by-step using the neural network $p_\theta(x_{t-1}|x_t)$ to approximate $q(x_{t-1}|x_t)$, ultimately generating the synthetic endothelial label x_0 . The reverse process is defined as:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \quad (3)$$

To improve the efficiency of this reverse diffusion process and to accelerate the convergence of the model, we utilize a cosine noise scheduler [20] instead of the linear one used in the original DDPM work [21] as it has shown to lead to faster convergence. The final step to sample an image from complete noise input x_T is:

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(x_t, t) \right) + \sqrt{\beta_t} z \quad (4)$$

where $\epsilon_\theta(x_t, t)$ is the predicted noise, and if $t \neq 1$, then $z \sim \mathcal{N}(0, I)$; otherwise, z is 0. This process repeats from $t = T$ to $t = 1$ to obtain x_0 .

B. High-Resolution Corneal Endothelium Image Generation

In high-resolution corneal endothelium image generation, the DDPM process is dependent on previously synthesized

endothelium labels. The conditional DDPM architecture concatenates the synthesized label with Gaussian noise at each step, guiding the image generation process. The process begins with the synthesized corneal endothelial label s , which is concatenated with the Gaussian noise input x_t at each diffusion step t , represented as $\tilde{x}_t = x_t \circ s$. Then the reverse process in the conditional DDPM progressively denoises this concatenated input \tilde{x}_t to generate the high-resolution image. The denoising step is defined by $p_\theta(x_{t-1}|\tilde{x}_t)$. This procedure generates high-resolution images along with corresponding labels, creating a substantial set of synthetic data for further training and evaluation.

C. The Architecture of DDPM

As shown in Fig. 1, the model begins with the input image x_t , undergoing initial feature extraction through a 2D convolutional layer, followed by normalization, and GELU activation. Features are then downsampled and passed through a Swin Transformer block [22], which uses Window Multi-head Self-Attention (W-MSA). After another downsampling and Swin Transformer block, the model enters a bottleneck phase with three convolutional blocks to refine features. During upsampling, two critical concatenation operations occur: first, the input image x_t is concatenated with the label along the channel dimension during the forward pass; second, skip connections between the downsampling and upsampling stages are utilized by concatenating the skip tensor with upsampled features along the channel dimension. Additionally, Fourier Residual Blocks with Frequency Selection (FRB-FS) are integrated at each upsampling and downsampling stage to enhance the model's capability to process fine textures through frequency domain operations, thereby improving the quality of image reconstruction. The model concludes by passing the upsampled features through a final 2D convolutional layer to generate the denoised output x_{t-1} .

The Swin Transformer block uses Window Multi-head Self-Attention (W-MSA) to compute self-attention within fixed windows, reducing complexity from quadratic to linear:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}} + B\right)V$$

where Q , K , and V are the *query*, *key*, and *value* matrices, d is the *query/key* dimension, and B is the relative position bias.

D. Fourier Residual Block with Frequency Selection

We introduce the Fourier Residual Block with Frequency Selection (FRB-FS), which, through frequency domain operations, enhances high-frequency components associated with blurred edges and fine textures, while also suppressing irrelevant low-frequency components. This dual function allows FRB-FS to generate denser and more accurate labels during DDPM sampling, thereby improving the representation of cell structures, as illustrated in Fig. 2. The process begins by applying a 2D Fourier Transform [23] to the input feature map $\mathbf{Z} \in \mathbb{R}^{H \times W \times C}$:

$$\mathbf{Z}_{\text{fft}} = \mathcal{F}(\mathbf{Z}) \quad (5)$$

where \mathcal{F} denotes the Fourier Transform. Subsequently, a ReLU activation is applied to both the real and imaginary parts of \mathbf{Z}_{fft} , which emphasizes useful high frequencies and reduces low-frequency noise:

$$\mathbf{Z}_{\text{real}} = \text{ReLU}(\text{Re}(\mathbf{Z}_{\text{fft}})), \quad \mathbf{Z}_{\text{imag}} = \text{ReLU}(\text{Im}(\mathbf{Z}_{\text{fft}})) \quad (6)$$

Following 1×1 convolutions, these components are recombined using the inverse Fourier Transform:

$$\mathbf{Z}_{\text{conv}} = \mathcal{F}^{-1}(\text{conv}(\mathbf{Z}_{\text{real}}) + i \cdot \text{conv}(\mathbf{Z}_{\text{imag}})) \quad (7)$$

where \mathcal{F}^{-1} represents the inverse Fourier Transform.

The process concludes by refining the real part of \mathbf{Z}_{conv} through a spatial convolution and integrating a residual connection to generate the output \mathbf{Y} :

$$\mathbf{Y} = \text{Conv}_{\text{spatial}}(\text{Re}(\mathbf{Z}_{\text{conv}})) + \mathbf{Z} \quad (8)$$

FRB-FS not only enhances boundary clarity and contrast by selectively emphasizing high-frequency details but also facilitates faster model convergence by inheriting the advantages of ResBlock [24].

E. Loss Function

We use the combined loss function for parameter optimization θ in the denoising network $\epsilon_{\theta}(\cdot)$, aimed at predicting noise, can incorporate both the L_1 [25] and MSE loss metrics. The combined loss function is defined as follows:

$$\mathcal{L}_{DM} = \mathbb{E}_{x_0, \epsilon, t} [\|\epsilon - \epsilon_{\theta}(x_t, t)\| + \|\epsilon - \epsilon_{\theta}(x_t, t)\|^2] \quad (9)$$

where x_t is computed as $\sqrt{\alpha_t}x_0 + \sqrt{1 - \alpha_t}\epsilon$.

III. EXPERIMENTS

A. Datasets

In this study, we employ two publicly available datasets from [26]: the Baerveldt and UT-DSAEK datasets, comprising

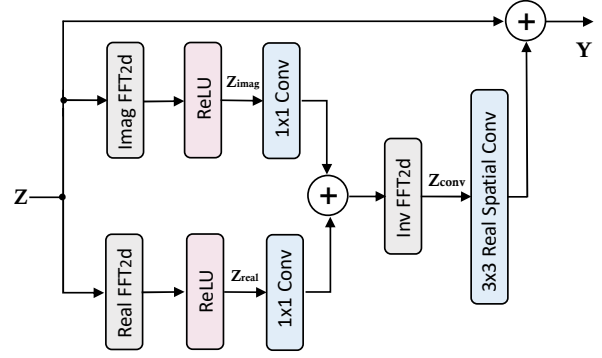


Fig. 2: Proposed Fourier Residual Block with Frequency Selection (FRB-FS).

a total of 1,203 corneal endothelial cell images acquired using a Topcon SP-1P microscope. These datasets cover both central and peripheral corneal regions, including cases with and without guttate. All images are 8-bit grayscale, with an original resolution of 240×528 pixels and have been manually annotated. For our experiments, we split the images into patches of size 256×256 pixels using nearest neighbour interpolation [27] to standardize the input for processing. Finally, we obtain 1,640 patches from the Baerveldt dataset and 766 patches from the UT-DSAEK dataset.

B. Implementation Details

All experiments were conducted on three NVIDIA GeForce RTX 4090 Ti GPUs, each with 64 GB of memory. Specifically, the U-Net [14] was trained using Binary Cross-Entropy (BCE) Loss [28] for segmentation tasks to assess the quality of the generated images. For the diffusion model training, the initial learning rate was set to 10^{-5} and was linearly decayed to zero over the course of 500 epochs to optimize the learning process. The Adam optimizer [29] was employed to adjust the weights efficiently. During the diffusion model training, a batch size of 32 was consistently used. Additionally, the noise schedule was linearly adjusted across 800 diffusion steps to incrementally denoise and reconstruct the images, ensuring both efficient and precise image processing during training and inference stages.

C. Results

We assess the performance of our proposed framework by comparing it to two well-established GAN-based image synthesis models, Pix2Pix [30] and Patch-GAN [16]. First, we evaluate the quality of the generated images using the Fréchet Inception Distance (FID) score. As shown in Table I, our diffusion-based framework achieves the lowest FID scores of 53.42 and 57.04 on the respective datasets, significantly outperforming the GAN models by a substantial margin.

To further evaluate the effectiveness of our method for corneal endothelium segmentation, we assess its performance on the Baerveldt and UT-DSAEK datasets. Our framework requires only the real labels for training, excluding the corneal endothelium images. We use 50% of the label data from each dataset to train our DDPM model, while the remaining 50%

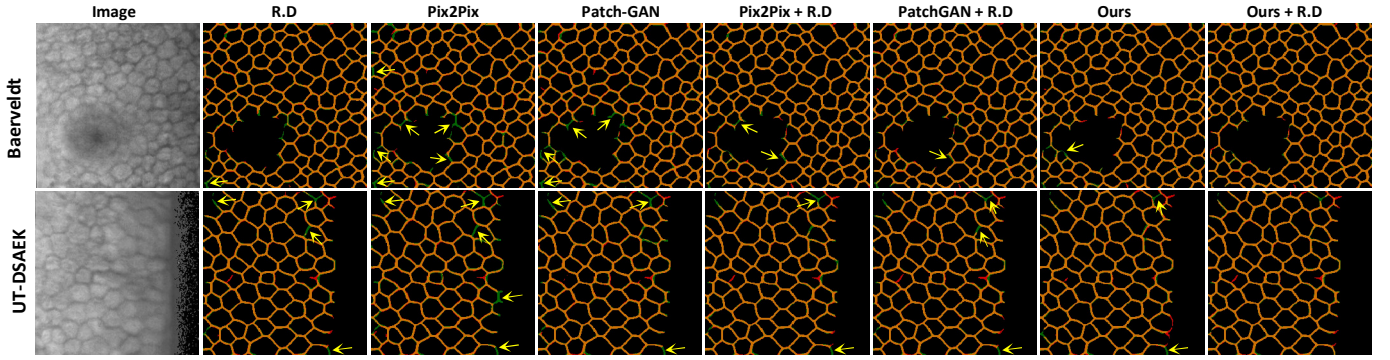


Fig. 3: Qualitative results on the Baerveldt and UT-DSAEK datasets are as follows: The prediction result is represented by a red line, while the ground truth is denoted by a green line, turning orange where they overlap. The yellow arrow indicates locations where the segmentation result does not align well with the ground truth.

TABLE I: Comparison of FID scores across different methods on Baerveldt and UT-DSAEK datasets.

Method		FID Score ↓	
		Baerveldt	UT-DSAEK
GAN-based	Pix2Pix [30]	142.72	157.94
	Patch-GAN [16]	68.02	71.15
Diffusion-based	Ours	53.42	57.04

data pairs are reserved for evaluating segmentation performance. Notably, GAN-based methods are unable to synthesize label images, and thus rely on the first-stage synthesized endothelial labels generated by our DDPM. In this experiment, U-Net is used as the baseline segmentation model. For each method, we train two models: one using only synthetic data, and the other incorporating 50% of the raw data pairs (+R.D) that were used to train our DDPM. Segmentation performance is then evaluated using the Dice Similarity Coefficient (Dice), Sensitivity (SE), Mean Accuracy (mAcc), and Modified Hausdorff Distance (MHD).

TABLE II: Comparison of U-Net [14] based corneal endothelial cell segmentation on benchmark datasets, using different training settings and performance metrics. The best and second-best results are in bold and underlined, respectively.

Test Data	Training Setting	Dice (%) ↑	SE (%) ↑	mAcc (%) ↑	MHD ↓
Baerveldt	Pix2Pix [30]	82.73 ± 1.98	80.16 ± 2.35	95.04 ± 1.77	0.582 ± 0.059
	Patch-GAN [16]	85.34 ± 2.13	83.22 ± 2.45	95.99 ± 1.84	0.435 ± 0.067
	Pix2Pix [30] + R.D	86.25 ± 1.82	83.95 ± 2.49	96.98 ± 1.79	0.294 ± 0.061
	Patch-GAN [16] + R.D	87.17 ± 2.19	84.76 ± 2.21	97.56 ± 1.72	0.195 ± 0.089
	Ours	86.68 ± 1.71	84.68 ± 2.35	97.63 ± 1.67	0.261 ± 0.084
	Ours + R.D	88.35 ± 1.91	85.52 ± 2.29	97.69 ± 1.86	0.132 ± 0.047
	R.D	86.66 ± 2.03	84.43 ± 2.42	97.35 ± 1.79	0.148 ± 0.072
UT-DSAEK	Pix2Pix [30]	81.25 ± 1.84	79.73 ± 2.46	94.91 ± 1.67	0.567 ± 0.069
	Patch-GAN [16]	84.37 ± 2.09	82.68 ± 2.31	95.19 ± 1.74	0.447 ± 0.063
	Pix2Pix [30] + R.D	85.28 ± 1.79	83.40 ± 2.48	96.05 ± 1.71	0.356 ± 0.078
	Patch-GAN [16] + R.D	86.46 ± 2.01	84.22 ± 2.41	96.78 ± 1.83	0.214 ± 0.087
	Ours	86.31 ± 1.93	83.90 ± 2.37	96.55 ± 1.68	0.284 ± 0.094
	Ours + R.D	87.24 ± 2.15	84.93 ± 2.29	97.49 ± 1.79	0.141 ± 0.053
	R.D	85.84 ± 2.04	83.76 ± 2.39	96.33 ± 1.72	0.162 ± 0.066

As shown in Table II, our proposed method, which combines synthetic data with raw data for training, sets a new benchmark by achieving state-of-the-art segmentation accuracy on both the Baerveldt and UT-DSAEK datasets. It significantly

outperforms the baseline GAN models, including Pix2Pix [30] and Patch-GAN [16], across all metrics. Notably, our method achieves the highest Dice scores of 88.35% and 87.24%, respectively. Furthermore, our approach demonstrates superior performance with the highest mean accuracy (mAcc) and the lowest Modified Hausdorff Distance (MHD), indicating more precise boundary delineation. Qualitative results in Fig. 3 illustrate that our method closely aligns with the ground truth, particularly in challenging regions where other methods struggle, as highlighted by the yellow arrows.

D. Ablation Study

We perform an ablation study on both datasets to verify the effectiveness of the FRB-FS module. As shown in Table III, incorporating the FRB-FS module significantly lowered the FID scores on the Baerveldt and UT-DSAEK datasets, reflecting improved image quality and closer alignment with real images. Moreover, the inclusion of the FRB-FS module reduced training time and accelerated model convergence.

TABLE III: Ablation study on the effectiveness of FRB-FS.

Dataset	FRB-FS	FID Score ↓	Training Time (hours) ↓
Baerveldt	×	67.26	19.53
	✓	53.42	18.41
UT-DSAEK	×	71.52	8.91
	✓	57.04	7.92

IV. CONCLUSIONS

In this study, we introduce a novel framework for generating corneal endothelial cell image pairs for segmentation tasks, utilizing DDPMs and the newly proposed FRB-FS module. Our method outperforms existing GAN-based models, demonstrating superior image quality and segmentation accuracy on the Baerveldt and UT-DSAEK datasets. The FRB-FS module further enhances image fidelity and reduces training time. Overall, our approach establishes a new benchmark in corneal endothelium image synthesis and segmentation.

REFERENCES

- [1] A. S. Panayides, A. Amini, N. D. Filipovic, A. Sharma, S. A. Tsaftaris, A. Young, D. Foran, N. Do, S. Golemati, T. Kurc *et al.*, “Ai in medical imaging informatics: current challenges and future directions,” *IEEE journal of biomedical and health informatics*, vol. 24, no. 7, pp. 1837–1857, 2020.
- [2] R. Yang and Y. Yu, “Artificial convolutional neural network in object detection and semantic segmentation for medical imaging analysis,” *Frontiers in oncology*, vol. 11, p. 638182, 2021.
- [3] C. L. Chowdhary and D. P. Acharjya, “Segmentation and feature extraction in medical imaging: a systematic review,” *Procedia Computer Science*, vol. 167, pp. 26–36, 2020.
- [4] T. Nakagawa, N. Okumura, M. Ikegawa, Y. Toyama, T. Nirasawa, F. Mascarelli, H. Vaitinadapoule, I. Aouimeur, Z. He, P. Gain *et al.*, “Shotgun proteomics identification of proteins expressed in the descemet’s membrane of patients with fuchs endothelial corneal dystrophy,” *Scientific Reports*, vol. 13, no. 1, p. 10401, 2023.
- [5] S. Watanabe, Y. Oie, H. Fujimoto, T. Soma, S. Koh, M. Tsujikawa, N. Maeda, and K. Nishida, “Relationship between corneal guttae and quality of vision in patients with mild fuchs’ endothelial corneal dystrophy,” *Ophthalmology*, vol. 122, no. 10, pp. 2103–2109, 2015.
- [6] Y. Q. Soh, G. S. Peh, S. L. Naso, V. Kocaba, and J. S. Mehta, “Automated clinical assessment of corneal guttae in fuchs endothelial corneal dystrophy,” *American Journal of Ophthalmology*, vol. 221, pp. 260–272, 2021.
- [7] A. M. Prada, F. Quintero, K. Mendoza, V. Galvis, A. Tello, L. A. Romero, and A. G. Marrugo, “Assessing fuchs corneal endothelial dystrophy using artificial intelligence–derived morphometric parameters from specular microscopy images,” *Cornea*, pp. 10–1097, 2022.
- [8] F. J. Sanchez-Marin, “Automatic segmentation of contours of corneal cells,” *Computers in biology and medicine*, vol. 29, no. 4, pp. 243–258, 1999.
- [9] S. Al-Fahdawi, R. Qahwaji, A. S. Al-Waisy, S. Ipson, M. Ferdousi, R. A. Malik, and A. Brahma, “A fully automated cell segmentation and morphometric parameter system for quantifying corneal endothelial cell morphology,” *Computer methods and programs in biomedicine*, vol. 160, pp. 11–23, 2018.
- [10] N. Joseph, I. Marshall, E. Fitzpatrick, H. J. Menegay, J. H. Lass, B. A. M. Benetz, and D. L. Wilson, “Deep learning segmentation of endothelial cell images using an active learning paradigm with guided label corrections,” *Journal of Medical Imaging*, vol. 11, no. 1, pp. 014006–014006, 2024.
- [11] A. Ruggeri, E. Grisan, and J. Jaroszewski, “A new system for the automatic estimation of endothelial cell density in donor corneas,” *British journal of ophthalmology*, vol. 89, no. 3, pp. 306–311, 2005.
- [12] L. Imre and A. Nagymihály, “Reliability and reproducibility of corneal endothelial image analysis by in vivo confocal microscopy,” *Graefes’ archive for clinical and experimental ophthalmology*, vol. 239, pp. 356–360, 2001.
- [13] E. Poletti and A. Ruggeri, “Segmentation of corneal endothelial cells contour through classification of individual component signatures,” in *XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013: MEDICON 2013, 25-28 September 2013, Seville, Spain*. Springer, 2014, pp. 411–414.
- [14] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.
- [15] Y. Zhang, R. Xi, L. Zeng, D. Towey, R. Bai, R. Higashita, and J. Liu, “Structural priors guided network for the corneal endothelial cell segmentation,” *IEEE Transactions on Medical Imaging*, 2023.
- [16] A. Kucharski and A. Fabijańska, “Corneal endothelial image segmentation training data generation using gans. do experts need to annotate?” *Biomedical Signal Processing and Control*, vol. 85, p. 104985, 2023.
- [17] T. Ueda, “Gan power devices: current status and future challenges,” *Japanese Journal of Applied Physics*, vol. 58, no. SC, p. SC0804, 2019.
- [18] H. Chen, “Challenges and corresponding solutions of generative adversarial networks (gans): a survey study,” in *Journal of Physics: Conference Series*, vol. 1827, no. 1. IOP Publishing, 2021, p. 012066.
- [19] D. Saxena and J. Cao, “Generative adversarial networks (gans) challenges, solutions, and future directions,” *ACM Computing Surveys (CSUR)*, vol. 54, no. 3, pp. 1–42, 2021.
- [20] A. Q. Nichol and P. Dhariwal, “Improved denoising diffusion probabilistic models,” in *International conference on machine learning*. PMLR, 2021, pp. 8162–8171.
- [21] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, “Deep unsupervised learning using nonequilibrium thermodynamics,” in *International conference on machine learning*. PMLR, 2015, pp. 2256–2265.
- [22] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, “Swin transformer: Hierarchical vision transformer using shifted windows,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 10012–10022.
- [23] M. Frigo and S. G. Johnson, “Fftw: An adaptive software architecture for the fft,” in *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP’98 (Cat. No. 98CH36181)*, vol. 3. IEEE, 1998, pp. 1381–1384.
- [24] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [25] Z. Dorjsembe, H.-K. Pao, S. Odonchimed, and F. Xiao, “Conditional diffusion models for semantic 3d medical image synthesis,” *Authorea Preprints*, 2023.
- [26] J. P. Viguera-Guillén, J. van Rooij, B. T. van Dooren, H. G. Lemij, E. Is-lamaj, L. J. van Vliet, and K. A. Vermeer, “Denseunets with feedback non-local attention for the segmentation of specular microscopy images of the corneal endothelium with guttae,” *Scientific reports*, vol. 12, no. 1, p. 14035, 2022.
- [27] M. Patil, “Interpolation techniques in image resampling,” *Int. J. Eng. Technol*, vol. 7, no. 3.34, pp. 567–570, 2018.
- [28] S. Jadon, “A survey of loss functions for semantic segmentation,” in *2020 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB)*. IEEE, 2020, pp. 1–7.
- [29] Z. Zhang, “Improved adam optimizer for deep neural networks,” in *2018 IEEE/ACM 26th international symposium on quality of service (IWQoS)*. Ieee, 2018, pp. 1–2.
- [30] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.