FlowOCT: Wavelet Flow Matching for OCT-to-OCTA Translation

Ze Xiong[1] , **Dehui Qiu**[1*] , **Liguo Deng**[1] , **Longfei Zhou**[1] , **Zhetao Xu**[2] , **Fa Zhang**[2] , **Xiaohua Wan**[2]

[1]Information Engineering College, Capital Normal University, China
[2]School of Medical Technology, Beijing Institute of Technology, China

{1200604026, qiudehui, 1221001005, 1201001041}@cnu.edu.cn
{3220242843, zhangfa, wanxiaohua}@bit.edu.cn

Abstract

Retinal angiography provides critical vascular information, yet optical coherence tomography angiography (OCTA) acquisition remains slower and less accessible than conventional structural OCT. A key open question is: *how angiographic signals can be recovered?* To address the issues of artifacts and structural blurring in existing generative methods, we propose a wavelet flow matching-based model, FlowOCT. This method operates directly on OCT wavelet coefficients, learning frequency-specific velocity fields to map them to the OCTA wavelet distribution. To ensure anatomical accuracy of the vasculature, we introduce constraints such as depth-wise contrastive alignment and two-dimensional projection consistency. Given the sparse nature of vascular signals in OCTA data, a structure-focused loss function and corresponding evaluation metrics are designed. Experiment results show that FlowOCT outperforms existing methods in terms of both image quality and perceptual similarity. Downstream diagnostic tasks further validate its superior authenticity and generalization capability. Code is available at https://github.com/cnu-medilab/FlowOCT.

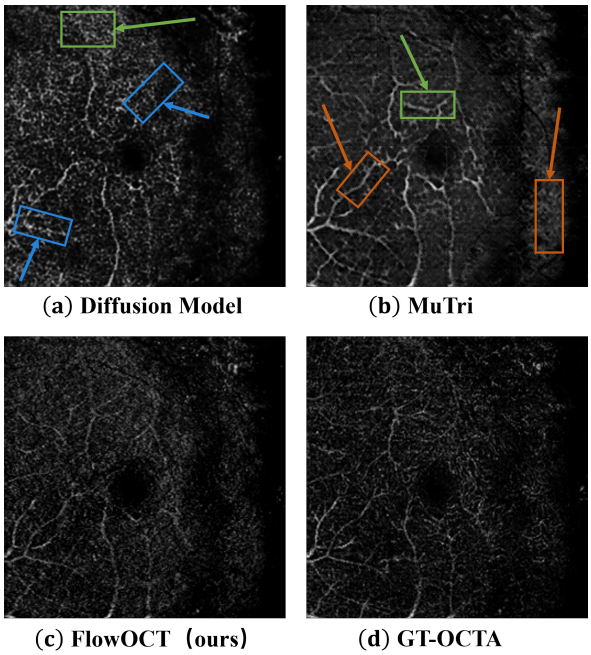

Figure 1: Visual comparison of generated OCTA images. (a) Results generated by the diffusion model; (b) Results generated by the MuTri; (c) Results generated by the proposed FlowOCT method; (d) Ground-truth OCTA image. Blue regions indicate vascular discontinuities (breaks), green regions indicate falsely generated vessels, and orange regions indicate vascular adhesion.

1 Introduction

Retinal diseases frequently present with both structural and vascular alterations. Structural optical coherence tomography (OCT) is routinely used to evaluate retinal morphology, while optical coherence tomography angiography (OCTA) provides complementary depth-resolved visualization of retinal blood flow and vascular organization [Huang *et al.*, 1991]. The combined interpretation of OCT and OCTA supports diagnosis, staging, and monitoring of retinal diseases by providing complementary structural and vascular perspectives [De Carlo *et al.*, 2015].

Despite its clinical value, OCTA requires more advanced instrumentation and remains slower to acquire, more sensitive to motion artifacts, and less widely available than OCT [Jia *et al.*, 2012]. In clinical practice, many diagnostic decisions are therefore made based solely on structural OCT [Pan *et al.*, 2025], without direct access to angiographic information. This limitation motivates a fundamental question: *Can clinically meaningful angiographic information be recovered from structural OCT?*

Recent learning-based methods attempt to synthesize OCTA from OCT using generative models [Ran and Cheung, 2021]. While these approaches can produce visually plausible angiograms, most advanced methods rely on stochastic sampling conditioned on OCT, which neglects key properties of retinal vasculature [Zhu *et al.*, 2025a; Zhu *et al.*, 2025b]. Angiographic signals are highly sparse, and clinically meaningful structures are dominated by low-amplitude,

*Corresponding author.

high-frequency components. Treating OCT merely as a conditioning signal makes such information easy to ignore, leading models to emphasize large vessels while failing to preserve fine capillary networks. Visual comparison in Figure 1 highlights distinct vascular artifacts in synthesized OCTA images from baseline methods. The image generated by the Diffusion model [Ho *et al.*, 2020] (a) exhibits vessel discontinuity and false vessel growth. Meanwhile, the result from the MuTri model [Chen *et al.*, 2025b] (b) suffers from false vessel growth and vessel adhesion. In contrast, our proposed FlowOCT method (c) effectively mitigates these artifacts, producing vascular structures that are morphologically closer to the ground truth (d). These systematic error patterns highlight the limitations of existing approaches in achieving anatomically plausible OCTA synthesis. As a result, existing generative methods often produce anatomically implausible outputs[Ran and Cheung, 2021; Courtie *et al.*, 2024]. Therefore, the OCTA images synthesized by existing methods often suffer from anatomical inaccuracies. Although visually subtle, such errors can introduce risks in clinical interpretation, as the integrity, connectivity, and morphological fidelity of vascular structures directly impact diagnostic conclusions[Coscas *et al.*, 2016].

Instead of treating OCT as a conditioning signal, we formulate OCT-to-OCTA synthesis as a noise-free flow-matching problem that starts directly from structural OCT and learns a deterministic velocity field to transport it toward the OCTA distribution. This formulation preserves patient-specific anatomy throughout generation and avoids artifacts introduced by stochastic sampling [Lipman *et al.*, 2023]. To faithfully recover vascular structure across spatial scales, we perform flow matching in a wavelet representation rather than in pixel space. This representation separates coarse anatomical layout from fine vascular detail and aligns naturally with the multi-scale organization of retinal vasculature.

Accurate angiographic synthesis further requires correct localization of vascular signals, which are confined to narrow axial bands and exhibit structured spatial organization [Garrity *et al.*, 2017]. To preserve this anatomy, we impose two complementary constraints. First, to prevent diffuse textures and vessel displacement across depth, we introduce retinal-depth contrastive alignment, which enforces depth-aware discrimination by aligning vascular signals to the correct retinal slice while remaining distinct from other depths. Second, to reflect how retinal imaging is interpreted clinically, we constrain synthesis through spatial projections to two-dimensional (2D) retinal layer views, ensuring that vascular spatial organization aligns with standard clinical assessment.

Contributions. We make the following contributions:

- We propose FlowOCT, a wavelet-based flow-matching framework that reformulates OCT-to-OCTA synthesis as a noise-free, anatomically faithful transport process from structural OCT wavelet representations to OCTA signals.
- To ensure anatomically and clinically meaningful vascular localization, we introduce depth- and spatial-aware constraints, including retinal-depth contrastive alignment (RDCA) and spatial retinal layer projection constraint.
- We introduce a Structure-Focused Objective that mitigates signal sparsity by emphasizing vascular regions, thereby improving capillary quality in training and establishing a correlated foreground evaluation metric.
- We demonstrate through image quality metrics and downstream diagnostic tasks that our method improves structural fidelity and clinical reliability.

2 Related Work

2.1 OCT-to-OCTA Image Translation

Cross-modality OCT-to-OCTA synthesis is challenging because of the intrinsic modality gap between static tissue structures and dynamic blood flow. Early 2D methods evolved from general frameworks like Pix2Pix [Isola *et al.*, 2017] to specialized architectures, including Texture-UNet [Zhang *et al.*, 2021] and AdjacentGAN [Li *et al.*, 2020b] for slice-wise B-scans, and MultiGAN [Pan *et al.*, 2022] for en-face projections. Despite their utility in quantifying retinal features [Badhon *et al.*, 2024; Wang *et al.*, 2025], 2D approaches inherently lose volumetric information and suffer from inter-slice discontinuity, leading to fragmented reconstruction.

With the availability of large-scale datasets like OCTA-500 [Li *et al.*, 2024a], recent research has shifted toward 3D volumetric translation. State-of-the-art frameworks, such as TransPro [Li *et al.*, 2024b] and XOCT [Khosravi *et al.*, 2025], utilize projection consistency and cross-dimension supervision, while MuTri [Chen *et al.*, 2025b] employs multi-view semantic alignment via VQ-VAE. However, these methods often overlook the inherent sparsity and multi-scale organization of vascular signals. Consequently, they struggle to recover fine capillary networks and frequently produce anatomical distortions, such as vessel adhesion, fragmentation, and spurious vessels, which compromise clinical reliability.

2.2 Diffusion and Flow Matching Models

Generative modeling has transitioned from GANs toward Diffusion Models (DMs) [Ho *et al.*, 2020; Song *et al.*, 2021]. Despite their success in medical reconstruction [Wu *et al.*, 2023; Müller-Franzes *et al.*, 2023], DMs face high computational costs due to the iterative nature of reversing stochastic paths. To improve efficiency, recent studies have conducted diffusion in compressed domains, such as latent spaces [Rombach *et al.*, 2022] or the wavelet domain [Phung *et al.*, 2023; Huang *et al.*, 2024], which leverages frequency decomposition to preserve structural sparsity and fine vascular details while reducing signal complexity.

To further accelerate inference, Flow Matching (FM) [Lipman *et al.*, 2023; Albergo and Vanden-Eijnden, 2023; Liu *et al.*, 2023; Tong *et al.*, 2023] has emerged as a simulation-free framework based on Continuous Normalizing Flows (CNFs). Recent progress has been done on medical imaging [Chen *et al.*, 2025a; Chen *et al.*, 2026; Jedynak, 2025] . By integrating wavelet representations with Flow Matching, our method unifies frequency-domain spatial efficiency with straight-line inference to ensure high-fidelity and rapid 3D OCT-to-OCTA translation.

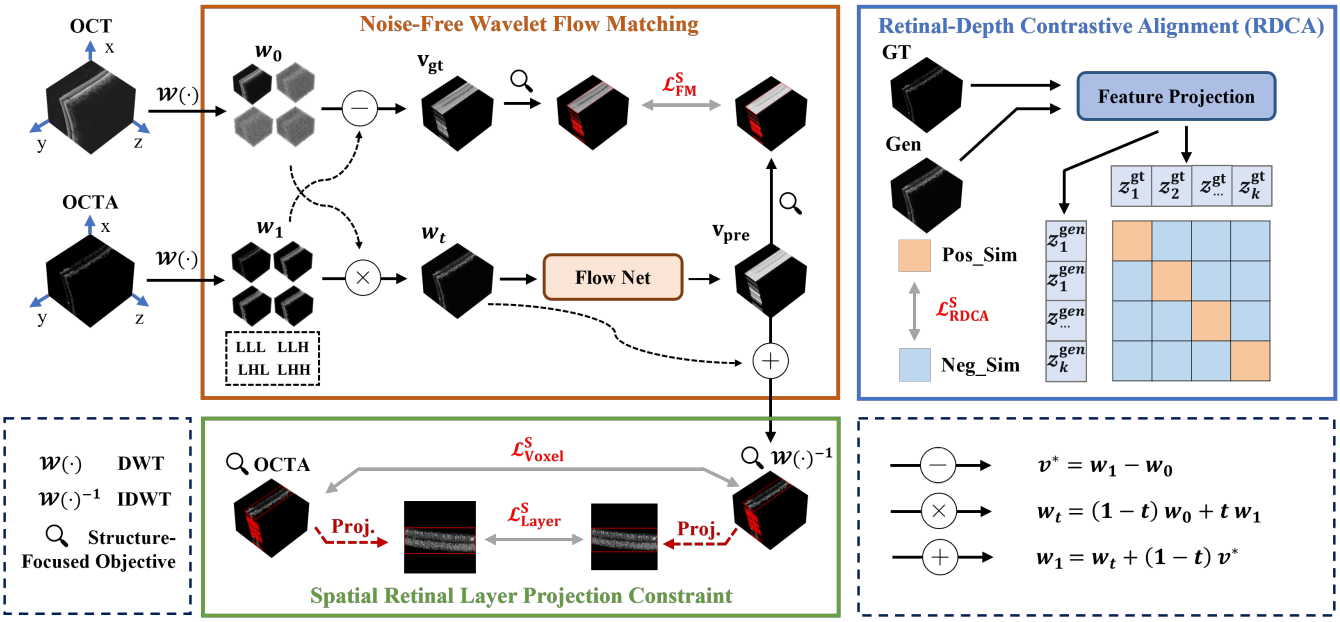

Figure 2: The overview of our proposed FlowOCT. FlowOCT primarily consists of four modules: 1) Noise-Free Wavelet Flow Matching, which serves as the core of the generative model; 2) Retinal-Depth Contrastive Alignment (RDCA), which constrains the depth distribution of vessels; 3) Spatial Retinal Layer Projection Constraint, which ensures vascular structural consistency through complementary voxel-wise and layer-projection supervision; 4) Structure-Focused Objective, which emphasizes vascular regions through mask-based weighting.

3 Methodology

As shown in Figure 2, our proposed FlowOCT integrates four complementary modules that jointly address the challenges of signal sparsity, depth specificity, and clinical consistency in OCTA synthesis, enabling anatomically faithful and clinically viable vascular generation.

3.1 Problem Setup

Let π_0 and π_1 denote the distributions of structural OCT and OCTA volumes, respectively. Both modalities are defined on a shared spatial domain $\Omega \subset \mathbb{R}^3$ with coordinates (h, w, z), where z indexes retinal depth. Given paired samples

$$\{(\mathbf{x}_0^i, \mathbf{x}_1^i)\}_{i=1}^N \sim (\pi_0, \pi_1), \tag{1}$$

our goal is to learn a conditional generative model that maps an observed OCT volume $\mathbf{x}_0 \sim \pi_0$ to a synthesized OCTA volume $\hat{\mathbf{x}}_1 \sim \pi_1$. N denotes the number of paired training samples.

3.2 Wavelet Representation

Angiographic signals exhibit strong multi-scale structure, with fine vessels encoded at high frequencies and global vascular organization at coarse scales. To explicitly represent this structure, we operate in the wavelet domain.

Let $\mathcal{W}$ denote 3-Dimension discrete wavelet transform (DWT). For a volume $\mathbf{x} \in \mathbb{R}^{H\times W\times Z}$, we compute

$$\mathbf{w} = \mathcal{W}(\mathbf{x}), \tag{2}$$

where

$$\mathbf{w} = \{\mathbf{w}^{(s,b)} \mid s = 1, \dots, S,\ b \in \mathcal{B}\}, \tag{3}$$

with s indexing wavelet scales and $\mathcal{B}$ denoting orientation subbands. The inverse transform recovers the original volume:

$$\mathbf{x} = \mathcal{W}^{-1}(\mathbf{w}). \tag{4}$$

3.3 Noise-Free Wavelet Flow Matching

We then define the generative process directly in wavelet space. Let

$$\mathbf{w}_0 = \mathcal{W}(\mathbf{x}_0), \qquad \mathbf{w}_1 = \mathcal{W}(\mathbf{x}_1) \tag{5}$$

denote the wavelet representations of OCT and OCTA volumes. We model a continuous flow governed by

$$\frac{d\mathbf{w}(t)}{dt} = \mathbf{v}_\theta(\mathbf{w}(t), \mathbf{w}_0, t)\,, \tag{6}$$

where $\mathbf{v}_\theta$ is a learnable velocity field conditioned on the source OCT wavelet coefficients $\mathbf{w}_0$.
The flow is initialized as

$$\mathbf{w}(0) = \mathbf{w}_0, \tag{7}$$

and integrated to $t = 1$, yielding synthesized coefficients

$$\hat{\mathbf{w}}_1 = \mathbf{w}(1). \tag{8}$$

The final OCTA volume is obtained via

$$\hat{\mathbf{x}}_1 = \mathcal{W}^{-1}(\hat{\mathbf{w}}_1). \tag{9}$$

Flow matching objective. Given paired samples $(\mathbf{w}_0, \mathbf{w}_1)$, we define a linear interpolation path

$$\mathbf{w}(t) = (1-t)\mathbf{w}_0 + t\mathbf{w}_1, \tag{10}$$

with target velocity

$$\mathbf{v}^* = \mathbf{w}_1 - \mathbf{w}_0. \tag{11}$$

The flow-matching loss is

$$\mathcal{L}_{\text{FM}} = \mathbb{E}_{t,\mathbf{w}_0,\mathbf{w}_1}\left[\|\mathbf{v}_\theta(\mathbf{w}(t), \mathbf{w}_0, t) - \mathbf{v}^*\|_2^2\right]. \tag{12}$$

3.4 Retinal-Depth Contrastive Alignment

Retinal vasculature is organized into depth-specific plexuses occupying narrow axial bands. To preserve this organization, we introduce Retinal-Depth Contrastive Alignment (RDCA) to enhance depth-specific vascular delineation and prevent axial diffusion and misalignment.

Given real and synthesized OCTA volumes $\mathbf{x}_1$ and $\hat{\mathbf{x}}_1$, we extract axial slices

$$\mathbf{x}_{1,z} = \mathbf{x}_1(:,:,z), \qquad \hat{\mathbf{x}}_{1,z} = \hat{\mathbf{x}}_1(:,:,z). \tag{13}$$

Each slice is encoded using a shared depth encoder f_{d}, producing normalized embeddings

$$\mathbf{e}_z = \frac{f_{\mathrm{d}}(\mathbf{x}_{1,z})}{\|f_{\mathrm{d}}(\mathbf{x}_{1,z})\|_2}, \qquad \hat{\mathbf{e}}_z = \frac{f_{\mathrm{d}}(\hat{\mathbf{x}}_{1,z})}{\|f_{\mathrm{d}}(\hat{\mathbf{x}}_{1,z})\|_2}. \tag{14}$$

We enforce depth-wise alignment using a contrastive objective

$$\mathcal{L}_{\mathrm{RDCA}} = -\sum_{z=1}^{Z} \log \frac{\exp(\hat{\mathbf{e}}_z^\top \mathbf{e}_z/\tau)}{\sum_{z'=1}^{Z} \exp(\hat{\mathbf{e}}_z^\top \mathbf{e}_{z'}/\tau)}. \tag{15}$$

3.5 Spatial Retinal Layer Projection Constraint

While RDCA enforces correct depth-wise localization, clinically meaningful interpretation of OCTA is performed at two complementary levels: voxel-level analysis for fine vessel details and layer-projection assessment for global vascular patterns. To ensure high-quality reconstruction that meets both clinical needs, we introduce a spatial retinal layer projection constraint that incorporates these two aspects.

First, to preserve fine vascular structures at the original spatial resolution, we introduce a voxel-wise L1 loss:

$$\mathcal{L}_{\mathrm{Voxel}} = \|\hat{\mathbf{x}}_1 - \mathbf{x}_1\|_1 \tag{16}$$

Second, to reflect clinical practice and enforce spatial consistency, we introduce a retinal layer projection constraint, following a setup similar to [Li *et al.*, 2024b; Khosravi *et al.*, 2025]. Given a volumetric OCTA image $\mathbf{x}_1$, we project it along the depth axis within predefined retinal layer ranges $\{\mathcal{Z}_k\}_{k=1}^{K}$ and enforce consistency between real and synthesized views:

$$\mathcal{L}_{\mathrm{Layer}} = \sum_{k=1}^{K} \left\| \mathcal{A}_{z\in\mathcal{Z}_k}(\hat{\mathbf{x}}_1) - \mathcal{A}_{z\in\mathcal{Z}_k}(\mathbf{x}_1) \right\|_1. \tag{17}$$

Here, $\mathcal{A} \in \{\mathrm{mean}, \max\}$ denotes the mean or maximum projection operator.

Combining these two complementary constraints, we obtain the spatial retinal layer projection constraint:

$$\mathcal{L}_{\mathrm{Proj}} = \mathcal{L}_{\mathrm{Voxel}} + \beta \cdot \mathcal{L}_{\mathrm{Layer}}, \tag{18}$$

where β balances the contribution of the layer-projection term. Both terms employ the L1 norm for its robustness to outliers and sparse signal characteristics.

3.6 Structure-Focused Objective

Since angiographic signals in OCTA volumes are highly sparse, with informative vascular structures occupying only a small fraction of the spatial domain, standard reconstruction losses are dominated by large background regions and therefore tend to underemphasize fine vascular details. To address this imbalance, we reformulate existing losses as structure-focused losses that emphasize informative vascular regions.

Specifically, we construct a 3D binary mask $\mathbf{M}$ from the ground-truth OCTA by adaptively detecting content regions with significant angiographic signals via thresholding, thereby identifying regions with significant angiographic signal. The loss is then computed between $\mathbf{M} \odot \hat{\mathbf{x}}_1$ and $\mathbf{M} \odot \mathbf{x}_1$, rather than between $\hat{\mathbf{x}}_1$ and $\mathbf{x}_1$. Similarly, for $\mathcal{L}_{\mathrm{FM}}$, we compute frequency masks and are applied to feature maps $\mathbf{w}$ instead of the input $\mathbf{x}$. We denote the structure-focused loss as $\mathcal{L}^s$.

The full training objective then is

$$\mathcal{L} = \mathcal{L}^s_{\mathrm{FM}} + \lambda_{\mathrm{Proj}}\mathcal{L}^s_{\mathrm{Proj}} + \lambda_{\mathrm{RDCA}}\mathcal{L}^s_{\mathrm{RDCA}}, \tag{19}$$

where the λ_{Proj} and λ_{RDCA} are loss weighting term.

4 Experiments

4.1 Experiments Setup

Dataset. We evaluate our method on the OCTA-500 dataset [Li *et al.*, 2024a; Li *et al.*, 2020a], which comprises two subsets with different fields of view: OCTA-3M (200 pairs, $304 \times 304 \times 640$ voxels) and OCTA-6M (300 pairs, $400 \times 400 \times 640$ voxels). Adopting the protocol from [Li *et al.*, 2024b], we partitioned OCTA-3M into 140/10/50 and OCTA-6M into 180/20/100 for training, validation, and testing, respectively. All volumes underwent standardized preprocessing, including intensity normalization and retinal region cropping. Furthermore, all data pairs are annotated with seven diagnostic categories to support the downstream clinical evaluation detailed in Section 4.3.

Implementation Details. For the proposed method, we parameterize the flow matching process using the DiffusionModelUnet framework from the Monai library. Models are trained with the Adam optimizer at a learning rate of 2×10^{-4}. Unless otherwise stated, we set $\beta = 0.2$, $\lambda_{\mathrm{Proj}} = 0.5$ and $\lambda_{\mathrm{RDCA}} = 0.005$ (calibrated to balance loss scales: RDCA ~ 4 vs. other losses ~ 0.001). Through comparative testing, we have chosen $\mathcal{A} = \max$ as the projection operator (Eq.(17)). We train the model with a batch size of 1 for 200 epochs. All experiments are conducted on a virtual GPU (vGPU) with 48 GB of memory.

Inference Configuration. For all generation tasks on the validation and test sets, we employ a fixed 3-step ODE solver to numerically solve the learned probability flow ODE [Lipman *et al.*, 2023]. This configuration balances optimal quantitative performance with high visual clarity while maintaining efficient inference. Further increasing the number of steps yields marginal or diminishing returns with substantially higher computational costs.

Method	MAE ↓ ($\times 10^{-2}$)	PSNR ↑	SSIM ↑ ($\times 10^{-2}$)	LPIPS ↓ ($\times 10^{-2}$)	FID ↓
Pix2pix 2D [[Isola *et al.*, 2017]]	1.30 / 4.03	28.41 / 19.93	86.78 / 51.78	6.15 / 12.98	<u>15.34 / 34.81</u>
Pix2pix 3D [Isola *et al.*, 2017]	1.53 / 4.67	26.82 / 18.49	83.54 / 43.07	9.44 / 20.60	77.51 / 81.60
Adjacent GAN [Li *et al.*, 2020b]	1.28 / 3.95	28.45 / 19.97	86.87 / 52.32	<u>5.99 / 12.87</u>	22.84 / 41.43
Diffusion Model [Ho *et al.*, 2020]	1.34 / 3.85	<u>28.63 / 20.39</u>	<u>86.02 / 52.84</u>	14.52 / 28.56	112.30 / 147.91
Flow Matching [Moschetto *et al.*, 2025]	1.28 / 4.05	28.52 / 20.31	86.05 / 49.85	5.87 / 12.95	32.81 / 36.89
TransPro [Li *et al.*, 2024b]	1.30 / 3.90	28.68 / 20.34	85.27 / 48.35	12.71 / 24.43	51.46 / 142.53
MuTri [Chen *et al.*, 2025b]	**1.15 / 3.34**	28.03 / 19.77	85.55 / 51.94	15.95 / 30.73	149.83 / 217.92
FlowOCT (ours)	<u>1.21 / 3.75</u>	**29.18 / 20.84**	**87.30 / 53.97**	**5.41 / 12.13**	**23.14 / 29.03**

Table 1: Quantitative comparison on the OCTA-3M dataset. Results are presented as "full volume / foreground region". The best results are highlighted in **bold**, and the second-best results are <u>underlined</u>.

Method	MAE ↓ ($\times 10^{-2}$)	PSNR ↑	SSIM ↑ ($\times 10^{-2}$)	LPIPS ↓ ($\times 10^{-2}$)	FID ↓
Pix2pix 2D [Isola *et al.*, 2017]	1.25 / 3.75	28.29 / 20.14	86.05 / 52.25	6.48 / 13.62	15.23 / 29.85
Pix2pix 3D [Isola *et al.*, 2017]	1.52 / 4.41	26.88 / 18.89	83.97 / 47.84	8.41 / 17.49	40.10 / 82.63
Adjacent GAN [Li *et al.*, 2020b]	1.21 / 3.65	28.65 / 20.49	86.26 / 52.85	5.98 / 12.32	<u>14.88 / 21.08</u>
Diffusion Model [Ho *et al.*, 2020]	1.20 / 3.58	28.58 / 20.50	86.05 / 53.58	9.00 / 16.94	31.46 / 56.86
Flow Matching [Moschetto *et al.*, 2025]	<u>1.29 / 3.52</u>	<u>29.08 / 21.18</u>	<u>85.23 / 55.83</u>	<u>6.81 / 11.98</u>	24.20 / 25.53
TransPro [Li *et al.*, 2024b]	1.36 / 4.01	27.72 / 19.69	86.38 / 54.08	7.30 / 15.75	70.90 / 82.69
MuTri [Chen *et al.*, 2025b]	1.44 / 3.64	28.87 / 21.05	85.50 / 55.55	12.07 / 19.95	180.65 / 157.68
FlowOCT (ours)	**1.19 / 3.30**	**29.46 / 21.66**	**87.12 / 58.50**	**5.49 / 10.87**	**19.38 / 19.25**

Table 2: Quantitative comparison on the OCTA-6M dataset. Results are presented as "full volume / foreground region". The best results are highlighted in **bold**, and the second-best results are <u>underlined</u>.

Evaluation Metrics. We employ a comprehensive set of metrics, including Mean Absolute Error (MAE), Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM) [Horé and Ziou, 2010], Learned Perceptual Image Patch Similarity (LPIPS) [Zhang *et al.*, 2018], and Fréchet Inception Distance (FID) [Heusel *et al.*, 2017], to thoroughly evaluate the fidelity of the reconstructed results. Specifically, LPIPS uses AlexNet as the feature extraction network, while FID is computed based on Inception-v3.

We observed that calculating these metrics on the full image volume yields artificially inflated scores due to interference from extensive background regions. This directly motivates our structure-focused objective function. To concentrate evaluation on meaningful content, we propose Foreground Metrics, which apply the evaluation only to the foreground region extracted via the structure-guided objective, in contrast to traditional Full Volume Metrics.

4.2 Comparison with Baseline Methods

We compare the proposed method with several representative baseline methods for OCT-to-OCTA synthesis. To ensure a fair comparison, all methods were trained and evaluated using the same data splits and preprocessing pipelines. The baseline methods include: Pix2pix 2D and Pix2pix 3D [Isola *et al.*, 2017], Adjacent GAN [Li *et al.*, 2020b], Diffusion Model [Ho *et al.*, 2020], Flow Matching [Moschetto *et al.*, 2025], Transpro [Li *et al.*, 2024b], and MuTri [Chen *et al.*, 2025b]. All baseline implementations followed their respective original frameworks or public repositories.

Experimental results on OCTA-500 dataset

Table 1 and Table 2 summarize the quantitative comparison results between the proposed method, FlowOCT, and various baseline methods on the OCTA-3M and OCTA-6M datasets.

On the OCTA-3M dataset, our method achieves optimal performance in perceptual quality (PSNR, SSIM) and perceptual similarity (LPIPS), outperforming state-of-the-art methods with a 1.7% improvement in Full PSNR, a 2.1% improvement in Foreground SSIM, and a 16.60% reduction in Foreground FID. On the OCTA-6M dataset, the advantages of our method in structural preservation and perceptual quality are further demonstrated. Compared to the second-best method, our approach achieves a 6.25% improvement in Foreground MAE, a 4.8% increase in Foreground SSIM, along with an 8.1% reduction in Full LPIPS and an 8.7% reduction in Foreground FID.

Overall, FlowOCT consistently outperforms all baselines across both datasets and all metrics, demonstrating its superior capability in preserving vascular structures and achieving high-fidelity OCTA synthesis.

Visualization results on OCTA-500 dataset

For qualitative evaluation of the results, we selected visual comparison examples across three projection dimensions (axial, coronal, and sagittal) generated by different methods, as illustrated in Figure 3. Additionally, we generated difference maps (see Figure 4) by performing pixel-wise subtraction between the synthesized OCTA images and the ground truth. Overall, GAN-based methods (Pix2pix 2D/3D [Isola *et al.*, 2017] and Adjacent GAN [Li *et al.*, 2020b]) tended to exhibit checkerboard artifacts or stripe-like noise, along with lower brightness. The Diffusion Model [Ho *et al.*, 2020] produced

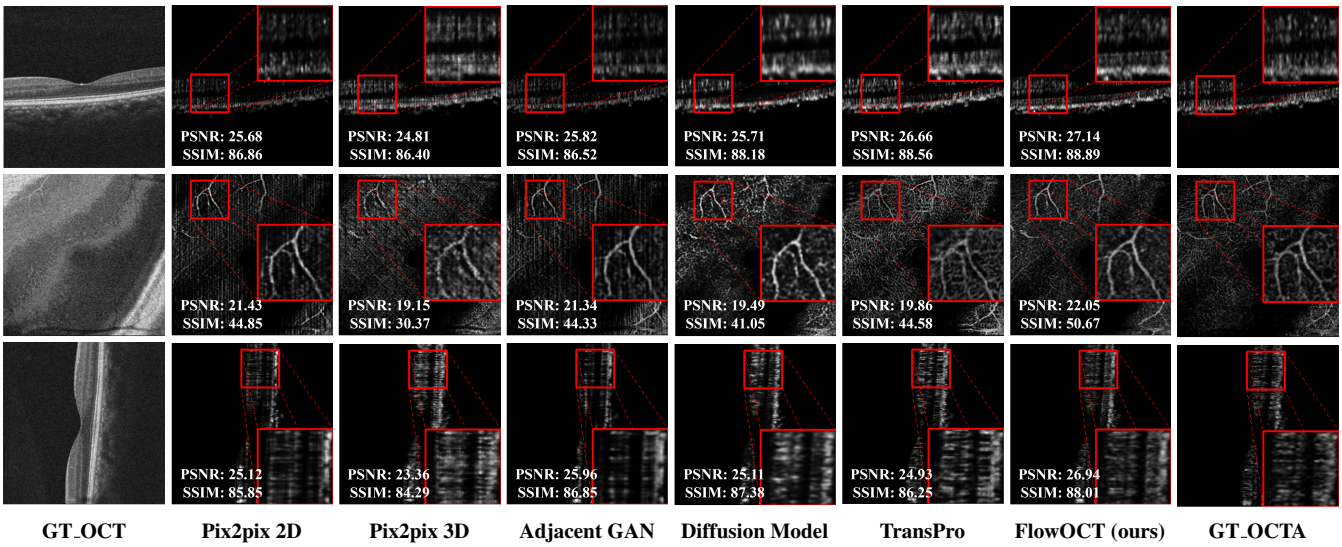

Figure 3: The visualization results of OCT-to-OCTA translation methods on the OCTA-500 dataset. Central slices of the three-dimensional volume are displayed along three projection planes. To more clearly compare differences in vascular generation, we provide zoomed-in local region comparisons for each method. Additionally, we compute the PSNR and SSIM between each synthesized slice and the corresponding ground-truth OCTA for quantitative evaluation.

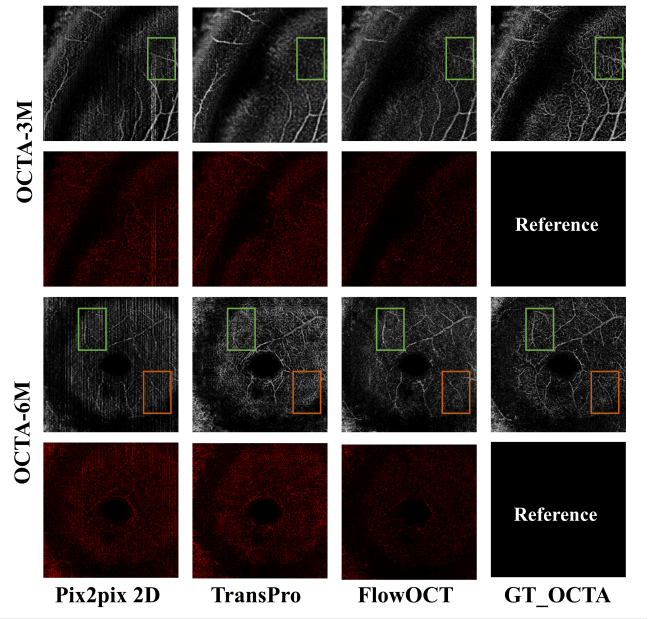

Figure 4: Qualitative comparison of synthesis errors via residual maps. Bounding boxes highlight regions with significant structural loss or artifacts, such as vessel fragmentation and spurious signals, where our method demonstrates superior fidelity.

overly blurry outputs with erroneous vessel connections and breaks. Transpro [Li *et al.*, 2024b] showed improved structural reconstruction and brightness, yet still suffered from vessel over-generation and adhesion. In contrast, our proposed FlowOCT method consistently achieved superior performance across all views: the generated OCTA images exhibited the smallest discrepancy in the difference maps, producing clearer, more continuous vessel structures with finer details, and delivering visual results closest to the ground truth.

Input OCTA Source	Accuracy (%) ↑	
	Seen Set	Unseen Set
Real OCTA	82.50	51.67
Pix2pix 2D [Isola *et al.*, 2017]	52.50$_{-30.0}$	23.33$_{-28.34}$
Pix2pix 3D [Isola *et al.*, 2017]	<u>55.00</u>$_{-27.5}$	25.00$_{-26.67}$
Adjacent GAN [Li *et al.*, 2020b]	<u>55.00</u>$_{-27.5}$	30.00$_{-21.67}$
Diffusion Model [Ho *et al.*, 2020]	47.50$_{-35.0}$	26.67$_{-25.00}$
Flow Matching [Moschetto *et al.*, 2025]	52.50$_{-30.0}$	<u>40.00</u>$_{-11.67}$
TransPro [Li *et al.*, 2024b]	47.50$_{-35.0}$	36.67$_{-15.00}$
MuTri [Chen *et al.*, 2025b]	27.50$_{-55.0}$	26.67$_{-25.00}$
FlowOCT (ours)	**65.00**$_{-17.5}$	**46.67**$_{-5.00}$

Table 3: Downstream diagnostic accuracy on synthetic data: authenticity vs. generalizability. The best results are highlighted in **bold**, and the second-best results are <u>underlined</u>. Each value for generation methods is annotated with its difference from the Real method (- indicates lower than real).

4.3 Downstream Diagnostic Evaluation

This study employs a downstream diagnostic task on the OCTA-6M dataset to evaluate generative models. A ResNet classifier under the MONAI framework is trained on a combination of 180 real training cases and the Seen Set data. The real test set is randomly split into a Seen Set (40 cases) for authenticity assessment and an Unseen Set (60 cases) for generalizability evaluation. Each generative model synthesizes corresponding data under identical conditions. Classification accuracy on the Synthetic-Seen Set measures the model's ability to imitate known real-world patterns (authenticity), while accuracy on the Synthetic-Unseen Set gauges

Configuration	MAE ↓ ($\times 10^{-2}$)	PSNR ↑	SSIM ↑ ($\times 10^{-2}$)	LPIPS ↓ ($\times 10^{-2}$)	FID ↓
w/o Wavelet-Domain	1.26 / 3.48	29.03 / 21.27	86.77 / 58.35	6.78 / 12.51	86.08 / 59.02
w/o RDCA	1.32 / 3.61	29.01 / 21.24	85.44 / 54.88	7.23 / 13.56	74.66 / 61.24
w/o Projection Constraint	1.31 / 3.82	28.28 / 20.24	86.05 / 53.66	6.36 / 12.57	20.09 / 31.05
w/o Structure-Focused Loss	1.32 / 3.74	28.50 / 20.58	85.95 / 54.58	5.61 / 11.55	73.13 / 69.23
FlowOCT (ours)	**1.19 / 3.30**	**29.46 / 21.66**	**87.12 / 58.50**	**5.49 / 10.87**	**19.38 / 19.25**

Table 4: Ablation study of key components on the OCTA-6M dataset. Results are presented as "full volume / foreground region". The best results are highlighted in **bold**, and the second-best results are underlined.

λ_{RDCA}	λ_{Proj}	MAE ↓ ($\times 10^{-2}$)	PSNR ↑	SSIM ↑ ($\times 10^{-2}$)	LPIPS ↓ ($\times 10^{-2}$)	FID ↓
0.0005	0.1	1.43 / 3.98	28.13 / 20.27	84.97 / 52.81	6.12 / 12.01	33.43 / 34.38
0.005	0.1	1.36 / 3.87	28.25 / 20.30	85.45 / 52.75	6.36 / 13.08	20.29 / 29.72
0.005 (default)	0.5 (default)	**1.19 / 3.30**	**29.46 / 21.66**	**87.12 / 58.50**	**5.49 / 10.87**	**19.38 / 19.25**
0.005	1.0	1.27 / 3.67	28.64 / 20.65	86.19 / 54.46	5.71 / 11.61	21.24 / 24.04
0.050	1.0	1.35 / 3.94	28.05 / 20.02	85.74 / 52.86	6.37 / 13.49	31.84 / 41.71
0.050	2.0	1.26 / 3.72	28.52 / 20.43	85.88 / 52.36	6.05 / 12.67	21.39 / 23.24

Table 5: Joint sensitivity analysis of loss weights on the OCTA-6M dataset. Results are presented as "full volume / foreground region". The best results are highlighted in **bold**, and the second-best results are underlined.

its generalization to unknown distributions (generalizability). This dual-dimensional strategy avoids complex classifier tuning and provides a comparable quantitative benchmark.

All models are evaluated under uniform conditions. The results are ranked primarily based on the Synthetic-Unseen Set performance, supplemented by a consistency analysis of the Synthetic-Seen Set results, with a focus on the deviation from real-data performance.

As summarized in Table 3, the evaluation reveals clear distinctions. In the authenticity dimension, where real OCTA data serves as an upper bound at 82.50%, most models achieve between 47.50% and 55.00%. Notably, our method FlowOCT reaches 65.00%, performing the best among all synthetic models and demonstrating effective imitation of known patterns. Regarding generalizability, the real data itself attains only 51.67%, underscoring the task's difficulty. Here, FlowOCT again leads at 46.67%, outperforming the second-best model (Flow Matching [Moschetto *et al.*, 2025]) by 1.67% , which confirms its strong out-of-distribution generalization. In contrast, most other models show limited generalizability (23.33%–40.00%), indicating their generated data fail to transfer effectively to unseen real distributions.

4.4 Ablation and Sensitivity Analysis

We conducted ablation studies to validate the effectiveness of each component and performed sensitivity analysis on loss-weight configurations, thereby confirming the robustness of the overall framework.

Table 4 demonstrates that FlowOCT achieves optimal performance across all metrics. Specifically, removing the wavelet representation degrades overall image quality, while omitting the RDCA module significantly impairs vascular structural integrity. Furthermore, ablating the projection constraint and structure-focused loss reduces cross-view consistency and reconstruction accuracy, respectively.

In summary, the wavelet-domain and RDCA module drive the primary performance gains by enhancing frequency-domain modeling and depth-wise structure preservation. Coupled with projection constraint and structure-focused loss, these components ensure clinical utility and detail precision, validating their collective necessity.

To further assess robustness, we perform a joint sensitivity analysis over the two loss weights, λ_{RDCA} and λ_{Proj}, which jointly govern anatomical alignment and projection consistency. As shown in Table 5, performance remains stable across a wide range of loss-weight combinations, with the default setting $\lambda_{\mathrm{RDCA}} = 0.005$ and $\lambda_{\mathrm{Proj}} = 0.5$ achieving the best overall results. This joint sensitivity analysis demonstrates that the proposed framework is not overly sensitive to precise hyperparameter tuning, supporting its robustness and practical applicability.

5 Conclusion

In this paper, we present FlowOCT, a novel framework for OCT-to-OCTA synthesis that incorporates distributional modeling and anatomical constraints. By formulating angiography generation as conditional flow matching, our method directly maps structural OCT to the OCTA distribution via deterministic paths. Operating in the wavelet domain enables the explicit modeling of vascular structures across multiple scales, while the retinal-depth contrastive alignment (RDCA) ensures anatomically consistent vessel organization. Experimental results demonstrate that FlowOCT consistently outperforms state-of-the-art methods in both reconstruction fidelity and downstream diagnostic accuracy. This work provides a principled framework for anatomy-aware angiographic synthesis and suggests a path toward leveraging routine OCT for vascular imaging.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 62541207, W2511070, 32241027 and 62472034, and in part by the Key Research and Development and Achievement Transformation Project of Inner Mongolia Autonomous Region in the Social Welfare Field under Grant 2026YFSH0177. The corresponding author is Dehui Qiu (qiudehui@cnu.edu.cn).

References

[Albergo and Vanden-Eijnden, 2023] Michael S. Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *Proceedings of the Eleventh International Conference on Learning Representations*, Kigali, Rwanda, May 2023.

[Badhon *et al.*, 2024] Rashadul Hasan Badhon, Atalie Carina Thompson, Jennifer I. Lim, Theodore Leng, and Minhaj Nur Alam. Quantitative characterization of retinal features in translated octa. *Experimental Biology and Medicine*, 249:10333, March 2024.

[Chen *et al.*, 2025a] Hao Chen, Rui Yin, Yifan Chen, Qi Chen, and Chao Li. Learning patient-specific disease dynamics with latent flow matching for longitudinal imaging generation. In *Proceedings of the International Conference on Learning Representations (ICLR) 2026*, 2025.

[Chen *et al.*, 2025b] Zhuangzhuang Chen, Hualiang Wang, Chubin Ou, and Xiaomeng Li. Mutri: Multi-view trialignment for oct to octa 3d image translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20885–20894, United States, June 2025.

[Chen *et al.*, 2026] Yifan Chen, Fei Yin, Hao Chen, Jia Wu, and Chao Li. Pmpbench: A paired multi-modal pan-cancer benchmark for medical image synthesis. *arXiv preprint arXiv:2601.15884*, 2026.

[Coscas *et al.*, 2016] Florence Coscas, Alexandre Sellam, Agnes Glacet-Bernard, Camille Jung, Mathilde Goudot, Alexandra Miere, and Eric H Souied. Normative data for vascular density in superficial and deep capillary plexuses of healthy adults assessed by optical coherence tomography angiography. *Investigative ophthalmology & visual science*, 57(9):OCT211–OCT223, 2016.

[Courtie *et al.*, 2024] Ella Courtie, James Robert Moore Kirkpatrick, Matthew Taylor, Livia Faes, Xiaoxuan Liu, Ann Logan, Tonny Veenith, Alastair K Denniston, and Richard J Blanch. Optical coherence tomography angiography analysis methods: a systematic review and meta-analysis. *Scientific Reports*, 14(1):9643, 2024.

[De Carlo *et al.*, 2015] Talisa E De Carlo, Andre Romano, Nadia K Waheed, and Jay S Duker. A review of optical coherence tomography angiography (octa). *International journal of retina and vitreous*, 1(1):5, 2015.

[Garrity *et al.*, 2017] Sean T Garrity, Nicholas A Iafe, Nopasak Phasukkijwatana, Xuejing Chen, and David Sarraf. Quantitative analysis of three distinct retinal capillary plexuses in healthy eyes using optical coherence tomography angiography. *Investigative Ophthalmology & Visual Science*, 58(12):5548–5555, 2017.

[Heusel *et al.*, 2017] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

[Ho *et al.*, 2020] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 6840–6851, 2020.

[Horé and Ziou, 2010] Alain Horé and Djemel Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th International Conference on Pattern Recognition*, pages 2366–2369, 2010.

[Huang *et al.*, 1991] David Huang, Eric A Swanson, Charles P Lin, Joel S Schuman, William G Stinson, Warren Chang, Michael R Hee, Thomas Flotte, Kenton Gregory, Carmen A Puliafito, et al. Optical coherence tomography. *science*, 254(5035):1178–1181, 1991.

[Huang *et al.*, 2024] Yi Huang, Jiancheng Huang, Jianzhuang Liu, Mingfu Yan, Yu Dong, Jiaxi Lv, Chaoqi Chen, and Shifeng Chen. Wavedm: Wavelet-based diffusion models for image restoration. *IEEE Transactions on Multimedia*, 26:7058–7073, 2024.

[Isola *et al.*, 2017] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1125–1134, Honolulu, HI, USA, July 2017.

[Jedynak, 2025] Bruno Jedynak. Optical coherence tomography harmonization with anatomy-guided latent metric schrödinger bridges. In *Wei, SThe Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[Jia *et al.*, 2012] Yali Jia, Ou Tan, Jason Tokayer, Benjamin Potsaid, Yimin Wang, Jonathan J Liu, Martin F Kraus, Hrebesh Subhash, James G Fujimoto, Joachim Hornegger, et al. Split-spectrum amplitude-decorrelation angiography with optical coherence tomography. *Optics express*, 20(4):4710–4725, 2012.

[Khosravi *et al.*, 2025] Pooya Khosravi, Kun Han, Anthony T. Wu, Arghavan Rezvani, Zexin Feng, and Xiaohui Xie. Xoct: Enhancing oct to octa translation via cross-dimensional supervised multi-scale feature learning. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Daejeon, Republic of Korea, September 2025.

[Li *et al.*, 2020a] Mingchao Li, Yuhan Zhang, Zexuan Ji, Keren Xie, Songtao Yuan, Qinghuai Liu, and Qiang Chen. Ipn-v2 and octa-500: Methodology and dataset for retinal image segmentation. *arXiv preprint arXiv:2012.07261*, 5(6):7, 2020.

[Li *et al.*, 2020b] Pei Lin Li, Céline O'Neil, Samin Saberi, Kenneth Sinder, Kathleen Wang, Bingyao Tan, Zohreh Hosseinaee, Kostadinka Bizheva, and Vasudevan Lakshminarayanan. Deep learning algorithm for generating optical coherence tomography angiography (octa) maps of the retinal vasculature. In *Proceedings of the International Conference on Applications of Machine Learning*, pages 39–49, San Diego, California, August 2020.

[Li *et al.*, 2024a] Mingchao Li, Yerui Chen, Zexuan Ji, Kifan Xie, Songtao Yuan, Qiang Chen, and Shuo Li. Octa-500: A retinal dataset for optical coherence tomography angiography study. *Medical image analysis*, 93:103092, 2024.

[Li *et al.*, 2024b] Shuhan Li, Dong Zhang, Xiaomeng Li, Chubin Ou, Lin An, Yanwu Xu, Weihua Yang, Yanchun Zhang, and Kwang-Ting Cheng. Vessel-promoted oct to octa image translation by heuristic contextual constraints. *Medical Image Analysis*, 98:103311, 2024.

[Lipman *et al.*, 2023] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *Proceedings of the Eleventh International Conference on Learning Representations*, Kigali, Rwanda, May 2023.

[Liu *et al.*, 2023] Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate with rectified flow. In *Proceedings of the Eleventh International Conference on Learning Representations*, Kigali, Rwanda, May 2023.

[Moschetto *et al.*, 2025] Andrea Moschetto, Lemuel Puglisi, Alec Sargood, Pierluigi Dell'Acqua, Francesco Guarnera, Sebastiano Battiato, and Daniele Ravì. Benchmarking gans, diffusion models, and flow matching for t1w-to-t2w mri translation, 2025.

[Müller-Franzes *et al.*, 2023] Gustav Müller-Franzes, Jan Moritz Niehues, Firas Khader, Soroosh Tayebi Arasteh, Christoph Haarburger, Christiane Kuhl, Tianci Wang, Tianyu Han, Teresa Nolte, Sven Nebelung, and Daniel Truhn. A multimodal comparison of latent denoising diffusion probabilistic models and generative adversarial networks for medical image synthesis. *Scientific Reports*, 13(1):12098, July 2023.

[Pan *et al.*, 2022] Bing Pan, Zexuan Ji, and Qiang Chen. Multigan: Multi-domain image translation from oct to octa. In *Proceedings of the Chinese Conference on Pattern Recognition and Computer Vision*, pages 336–347, Shenzhen, China, November 2022.

[Pan *et al.*, 2025] Xue Pan, Ze Xiong, Dehui Qiu, Liguo Deng, Boxuan Zhao, Xiaojie Cao, Xiaohua Wan, and Fa Zhang. A semi-supervised reinforcement learning framework incorporating the multi-scale incepmambanet network for glaucoma progression prediction. *Interdisciplinary Sciences: Computational Life Sciences*, pages 1–20, 2025.

[Phung *et al.*, 2023] Hao Phung, Quan Dao, and Anh Tran. Wavelet diffusion models are fast and scalable image generators. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10199–10208, Vancouver, Canada, June 2023.

[Ran and Cheung, 2021] Anran Ran and Carol Y Cheung. Deep learning-based optical coherence tomography and optical coherence tomography angiography image analysis: an updated summary. *Asia-Pacific Journal of Ophthalmology*, 10(3):253–260, 2021.

[Rombach *et al.*, 2022] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, New Orleans, Louisiana, June 2022.

[Song *et al.*, 2021] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations (ICLR)*, 2021.

[Tong *et al.*, 2023] Alexander Tong, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrad Rector-Brooks, Kincaid Fatemi, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, October 2023.

[Wang *et al.*, 2025] Yunxuan Wang, Ray Yin, Yumei Tan, Hao Chen, and Haiying Xia. Low-rank adaptive structural priors for generalizable diabetic retinopathy grading. *IJCNN 2025*, 2025.

[Wu *et al.*, 2023] Junde Wu, Huihui Fang, Yu Zhang, Yehui Yang, and Yanwu Xu. Medsegdiff: Medical image segmentation with diffusion probabilistic model. In *Proceedings of the International Conference on Medical Imaging with Deep Learning*, pages 1623–1639, Nashville, Tennessee, July 2023.

[Zhang *et al.*, 2018] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.

[Zhang *et al.*, 2021] Ziyue Zhang, Zexuan Ji, Qiang Chen, Songtao Yuan, and Wen Fan. Texture-guided u-net for oct-to-octa generation. In *Proceedings of the Fourth Chinese Conference on Pattern Recognition and Computer Vision*, pages 42–52, Beijing, China, October 2021.

[Zhu *et al.*, 2025a] Chunzheng Zhu, Yangfang Lin, Shen Chen, Yijun Wang, and Jianxin Lin. Medeyes: Learning dynamic visual focus for medical progressive diagnosis. *arXiv preprint arXiv:2511.22018*, 2025.

[Zhu *et al.*, 2025b] Chunzheng Zhu, Yangfang Lin, Jialin Shao, Jianxin Lin, and Yijun Wang. Pathology-aware prototype evolution via llm-driven semantic disambiguation for multicenter diabetic retinopathy diagnosis. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pages 9196–9205, 2025.