# Generating 3D Optical Coherence Tomography from 2D Fundus Images via Diffusion Models

**Bowen Liu**
School of Optometry, The Hong Kong Polytechnic University

**Yue Wu**
The Hong Kong Polytechnic University

**Ruoyu Chen**
The Hong Kong Polytechnic University

**Pusheng Xu**
The Hong Kong Polytechnic University

**Peng Xiao**
State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University

**Zhen Tian**
The Hong Kong Polytechnic University

**Binwei Huang**
Shantou University Medical College

**Mingguang He**
mingguang.he@polyu.edu.hk

The Hong Kong Polytechnic University https://orcid.org/0000-0002-6912-2810

**Danli Shi**
danli.shi@polyu.edu.hk

The Hong Kong Polytechnic University https://orcid.org/0000-0001-6094-137X

**Article**

**Keywords:**

**Posted Date:** March 6th, 2025

**DOI:** https://doi.org/10.21203/rs.3.rs-6160112/v1

**Additional Declarations:** There is **NO** Competing Interest.

# Generating 3D Optical Coherence Tomography from 2D Fundus Images via Diffusion Models

## Abstract

Training machine learning models with synthetic data effectively addresses data scarcity, particularly in domains where acquiring large-scale 3D datasets is costly. We present Fundus2OCT, the first framework to synthesize high-fidelity 3D optical coherence tomography (OCT) volumes from 2D fundus photographs using diffusion models. Developed on paired fundus-OCT data from the UK Biobank, Fundus2OCT leverages a two-stage latent diffusion process to generate anatomically coherent OCT volumes (32 B-scans per volume) conditioned on fundus inputs. Quantitative evaluations demonstrate superior performance over existing methods, with Fréchet Inception Distance (FID) and Fréchet Video Distance (FVD) scores of 12.3 and 58.7, respectively. In a clinical Turing test, two ophthalmologists achieved accuracies of 49.0–57.0% (near chance-level) in distinguishing synthetic from real OCTs. To validate clinical utility, we augmented four public fundus-based disease detection tasks (AMD, glaucoma, DR, DME) with synthetic OCT data, improving multimodal classification AUC by 4.2–8.6%. By bridging 2D fundus findings with 3D structural insights, Fundus2OCT advances multimodal retinal analysis, offering a scalable solution to enhance diagnostic accuracy and accessibility in ophthalmic care.

## Introduction

Three-dimensional optical coherence tomography (3D-OCT) has revolutionized ophthalmic diagnosis by transcending the limitations of conventional two-dimensional imaging modalities. Unlike surface-level views, 3D-OCT provides three critical advancements: depth-resolved visualization spanning from the vitreous to the choroid and sclera; microscopic resolution capable of distinguishing neural layers, including photoreceptors; and comprehensive volumetric analysis of posterior eyeball curvature and lesion morphology.[1] These capabilities position 3D-OCT as an indispensable tool for managing sight-threatening conditions such as diabetic macular edema, age-related macular degeneration, and glaucoma.[2-4] However, its clinical adoption remains constrained by high costs and the need for specialized expertise. In contrast, color fundus photography (CFP) is widely accessible but lacks depth information critical for localizing pathologies. Combining these modalities could improve diagnostic accuracy, yet paired datasets remain scarce due to OCT device costs.

Generative artificial intelligence (GenAI) offers a transformative approach to addressing data scarcity through the synthesis of realistic medical images. Previous applications of GenAI include improving image quality, augmenting rare datasets, enhancing segmentation, and reducing reliance on invasive procedures.[5-15] Prior studies have explored unconditional 2D OCT slice generation using generative adversarial networks (GANs)[16,17] and variational autoencoders (VAEs)[18], though these methods fail to capture volumetric context or establish fundus-to-OCT cross-modal relationships. Emerging diffusion models, however, surpass GANs and VAEs in both fidelity and diversity across imaging domains. Their iterative denoising process preserves fine anatomical details while maintaining mode covergence, making them uniquely suited for medical imaging tasks requiring anatomical precision.[19-21]

We introduce Fundus2OCT, a two-stage latent diffusion framework that synthesizes 3D-OCT volumes from CFP inputs. Our work makes three contributions. First, Fundus2OCT represents the first volumetric OCT synthesis framework capable of generating full 3D-OCT volumes, preserving spatial context essential for evaluating lesion depth and morphology. Second, we demonstrate that synthetic OCT data enhances deep learning models, improving disease detection accuracy in scenarios where multimodal imaging is unavailable. Third, we rigorously validate synthetic OCT quality using standardized generative metrics, confirming alignment with real data distributions while preserving diversity. In a clinician-led authenticity evaluation, synthetic images were indistinguishable from real ones, underscoring their clinical relevance. By bridging the accessibility gap between CFP and 3D-OCT, Fundus2OCT paves the way for scalable, cost-effective multimodal diagnostics.

## Methods

### Dataset for developing Fundus2OCT

Paired color fundus photography (CFP) and optical coherence tomography (OCT) images were sourced from the UK Biobank, a prospective cohort of approximately 500,000 participants aged 40–69 years recruited between 2006 and 2010.[22] A subset of 67,321 participants underwent baseline ocular imaging, with an additional 17,876 imaged at follow-up, including both CFP and OCT.[23] Ethical approval for the UK Biobank study was granted by the North West Multi-Centre Research Ethics Committee (06/MRE08/65), and informed consent was obtained from all participants.

### OCT and CFP acquisition

Spectral-domain OCT scans were acquired using a Topcon 3D OCT-1000 Mark II system under standardized illumination without pupil dilation. Macular volume scans covered a 6×6 mm area, comprising 512 horizontal A-scans per B-scan and 128 B-scans per volume. CFP images were captured using the same Topcon system.

### Quality control and image preprocessing

Participants with withdrawn consent or missing CFP/OCT data were excluded. Low-quality images were removed using established criteria: CFPs classified as "Reject" by the RMHAS[24] quality assessment tool and OCTs flagged by Topcon's image control indicators[4]. After matching OCT and CFP images by eye laterality, one optimal pair per participant was retained.

For CFP, two preprocessing steps were applied: (1) exclusion of images where a circle-detection algorithm failed to identify circular contours and (2) removal of the top/bottom 0.5% of images by brightness. Images were cropped into squares based on detected contours, with padding as needed.

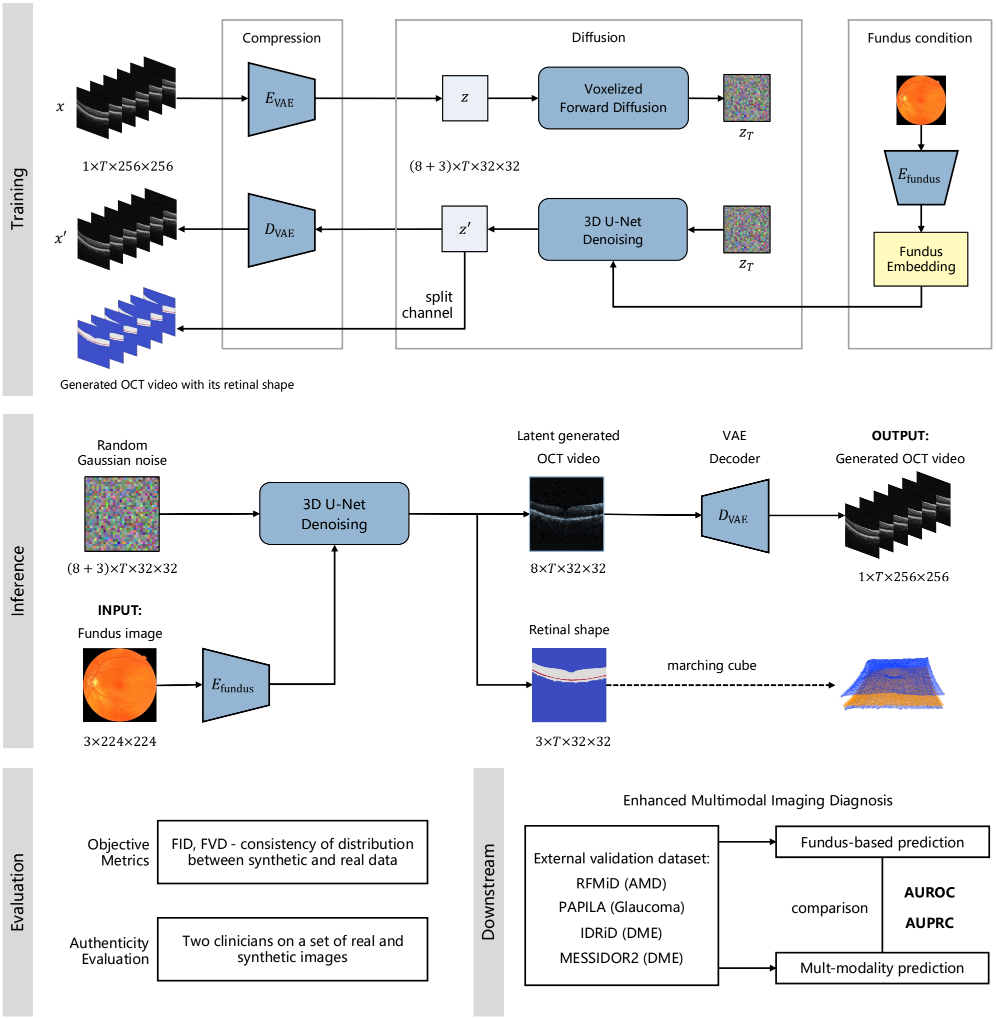
For OCT, we retained the central 32 B-scans from each volume. We used segmentation indicators, including the internal limiting membrane (ILM) edge strength metric to identify artifacts (e.g., blinks, signal fading). Retinal boundaries ($r_1$, $r_2$) derived from OCT metadata guided vertical cropping: upper and lower bounds were adjusted by h/6 (where h = original B-scan height), while width remained unchanged. Square shapes were achieved via padding, with retinal contour information preserved for downstream analysis.

The dataset was partitioned into developmental (80%) and internal test (20%) sets at the participant level to prevent data leakage. The developmental set was further divided into training (model parameter optimization) and validation (hyperparameter tuning) subsets.

### External validation datasets

Publicly available datasets were used to validate the utility of synthetic OCT data in enhancing multimodal disease classification. These included RFMiD[25] (age-related macular degeneration (AMD), reduced to binary classification by retaining only "ARMD" cases), IDRiD[26] and MESSIDOR-2[27,28] (diabetic macular edema (DME), and PAPILA[29] (glaucoma). All tasks originally relied on CFP but were extended to multimodal prediction using synthetic OCTs.

**Fig. 1: Schematic illustration of Fundus2OCT.**



**Training:** The training was conducted on paired OCT volumes and fundus images. We started by training an autoencoder to learn a compressed latent space of lower dimensionality. Our subsequent latent diffusion model was designed to work with this learned latent space. **Inference:** The generation pipeline consists of a latent video diffusion model, a VAE decoder. Specifically, we first generate a $8 \times T \times 32 \times 32$ low-resolution latent OCT video, along with its segmentation mask of retinal layers, conditioning on random Gaussian noise and the fundus photography. The latent OCT video is then upscaled to a higher resolution by the VAE decoder. Mesh representation of retinal layers can also be derived from the 3D mask. **Evaluation & Downstream:** Fundus2OCT was developed using fundus photographs to generate OCT videos. The generated OCT videos were evaluated with standard generative metrics to highlight the semantic similarly. The generated videos were evaluated by clinicians for authenticity, and also used to augment the dataset for supervised classification tasks.

### Fundus2OCT model architecture

Fundus2OCT generates a 3D OCT video with $T$ B-scans from a single CFP input (Fig. 1). Fundus-prompted generation of high-resolution 3D OCT images poses significant computational and detail-preserving challenges. To address these, we introduced a two-stage scheme. We started by training an autoencoder to learn a compressed latent space of lower dimensionality. Our subsequent diffusion model is designed to work with this learned latent space. The training scheme is illustrated in Supplementary Fig. 1.

### Model components

Specifically, Fundus2OCT consists of two trainable parts: a VAE for OCT scan compression and reconstruction, and a latent video diffusion model (LVDM) for learning the mapping between 2D fundus and 3D OCT retinal layers. Each part is trained independently. Firstly, the VAE model is trained to project each OCT B-scan $I \epsilon \mathbb{R}^{1\times H\times W}$ into two latent tensors, a mean $\mu$ and a standard deviation $\sigma$, such that $\{\mu, \sigma\} \epsilon \mathbb{R}^{8\times H/8\times W/8}$. These latent tensors are used to define an $n$-dimensional Gaussian distribution, which is sampled to produce compact embeddings of the B-scan. From this compact embedding, the VAE decoder reconstructs the input image. The VAE is trained by using a combination of $\mathcal{L}_1$ reconstruction loss, LPIPS[30] and Structural Similarity Index Measure (SSIM) loss, and a KL-divergence loss. The KL-divergence loss ensures a Gaussian distribution in the latent space, while the integration of the remaining loss functions avoids the blurring effects typically associated with relying exclusively on pixel-space losses like $\mathcal{L}_1$ objectives. After the VAE is trained, we used it to encode the 3D OCT scans into OCT latent space, and concatenate the latent with its retinal shapes derived from 4 different contours of retinal layers. These four contours divide the retinal layers into three regions, excluding the choroid-scleral interface. Therefore, we use the semantic mask to represent the retinal shape, with category 1 representing the overall shape (combining the first and the third region), category 2 for the intermediate region, and category 0 for the background.

We then train the LVDM on this joint representation $x \epsilon \mathbb{R}^{(8+3)\times T\times H/8\times W/8}$, where $T$ is the number of B-scans, $H$ and $W$ are spatial size. The LVDM is a discrete-time Denoising Diffusion Probabilistic Model (DDPM). The LVDM consists of a 3D U-Net, with four residual blocks and channel sizes of 256, 256, 512, 512. To train the LVDM we use an MSE loss for pixel supervision in conjunction with a Dice loss for shape guidance.

### Conditional mechanism

To achieve fine-grained control of the diffusion process guided by fundus images, we employ EyeFound[31] , a pretrained multimodal foundation model, to encode fundus photographs into a compact latent representation. Specifically, we selected EyeFound for its capacity to unify cross-modal ophthalmic data, compressing high-dimensional CFPs into latent embeddings—termed fundus conditional information. Trained via self-supervised reconstruction on millions of multimodal ophthalmic images spanning 11 modalities, EyeFound has demonstrated superior performance in multimodal downstream tasks. We integrate these fundus embeddings into the denoising 3D U-Net[32] using straightforward feature concatenation, enabling conditional generation of OCT volumes while preserving anatomical coherence with the input fundus image.

### Model inference

During inference, Gaussian noise and fundus embeddings were sampled to generate latent 3D OCT volumes via the LVDM (Fig. 1). The VAE decoder upscaled these OCT latents frame-by-frame into high-resolution B-scans (256×256 pixels), producing a final OCT video with 32 B-scans.

### Computational resources and training details

The VAE was trained for 900,000 steps on two NVIDIA A800 GPUs (80GB) with a batch size of 128 and a learning rate of $1\times10^{-4}$. The LVDM was trained for 95,000 steps on four A800 GPUs, using a batch size of 32 and the Adam optimizer with a learning rate of $1\times10^{-4}$. The diffusion process employed 1,000 timesteps with a linear noise schedule.

### Generation evaluation

The image reconstruction performance of the VAE was evaluated using a combination of pixel-based and feature-based metrics. Pixel-based assessments included the Mean Absolute Error (MAE), Mean Squared Error (MSE), Structural Similarity Index (SSIM), and Peak Signal-to-Noise Ratio (PSNR), while the feature-based evaluation included the Fréchet Inception Distance (FID) score[33]. Among these, lower values of MAE, MSE, and FID indicate better performance, whereas higher values are desirable for SSIM and PSNR.

For the generated OCT video analysis, objective assessment extended to the Fréchet Video Distance (FVD)[34], a video-adapted version of FID, with lower scores reflecting greater similarity to real videos. In addition to these quantitative measures, two experienced clinicians conducted a subjective evaluation to assess the authenticity of synthetic images compared to real ones.

To investigate the optimal framework configuration, we also trained two model variants: a fundus- and shape-conditioned diffusion model and a pretrained unconditional model fine-tuned for conditional generation, the architectures are demonstrated in Supplementary Fig. 4.

### External evaluation of downstream classification tasks

To assess the clinical value of the synthesized OCT images produced by Fundus2OCT, we integrated the generated data into downstream classification tasks to determine whether multimodal training enhances diagnostic performance. In the experiments, the EyeFound model served as a fixed feature extractor for both fundus and OCT images, with training limited to fine-tuning task-specific multilayer perceptron (MLP) classifiers. For 3D OCT volumes, each B-scan was processed individually through the feature extractor, and the resulting embeddings were sequentially fed into the MLP. The final prediction was derived by averaging the outputs across all B-scans. A baseline classifier trained exclusively on fundus images was compared against an augmented version trained on both fundus images and synthetic OCT scans generated by Fundus2OCT. The synthetic OCTs were created by conditioning the model on fundus images from the same dataset. Training protocols for both classifiers, detailed in Supplementary Fig. 2, were designed to isolate the impact of synthetic data augmentation. To ensure reproducibility and generalizability, only publicly available datasets were included. Official dataset splits were preserved where provided; for datasets lacking predefined splits, data were randomly divided into training, validation, and test sets at an 8:1:1 ratio. Results from five independent random splits were aggregated, with performance metrics reported as the mean and 95% confidence interval to account for variability.

## Results

### Evaluation of generation quality

The VAE model demonstrated good performance in reconstructing OCT B-scans, achieving an MSE of 0.0089 and an MAE of 0.0587 on the internal test set. Reconstructed B-scans attained a FID score of 31.5125, with comparable performance in the high myopia subgroup (MSE: 0.0091, FID: 31.5614), as detailed in Table 1. Visualizations of reconstructed B-scans confirmed high fidelity (Supplementary Fig. 3).

For 3D OCT volume synthesis, Fundus2OCT generated 12,200 volumes (32 B-scans each) and achieved an FID of 33.0827. The FVD scores, which measure temporal coherence, were 69.7257 (FVD16f, 16 uniformly sampled B-scans) and 135.9953 (FVD32f, all 32 B-scans). These results outperformed two baseline methods: a fundus- and shape-conditioned diffusion model (FVD32f: 155.7248) and a pretrained model fine-tuned for conditional generation (FVD32f: 141.0591) (Table 2). Qualitative assessments further validated the anatomical realism of generated volumes (Supplementary Fig. 5-7).

Retinal layer segmentation masks generated alongside OCT volumes achieved a Dice score of 0.9675 and an intersection over union (IoU) of 0.9374, with a Chamfer distance (surface mesh accuracy) of 0.3462. Performance remained consistent in high myopic cases (Dice: 0.9669, IoU: 0.9363), though gaps persisted compared to state-of-the-art segmentation methods (Table 3). In Table 1-3, the subgroup with high myopia is listed because this group exhibits differences in macular curvature distribution.[35]

### Clinical validation of authenticity

Two ophthalmologists evaluated 100 OCT volumes (50 real, 50 synthetic) in a blinded Turing test. Clinician 1 achieved 49% accuracy (sensitivity: 0.28, specificity: 0.70), while Clinician 2 reached 57% accuracy (sensitivity: 0.14, specificity: 1.00), indicating near-chance discrimination between real and synthetic data (Table 5). These results underscore the anatomical plausibility of Fundus2OCT-generated volumes.

### Generated OCT enhances retinal disease classification

Augmenting fundus-based classifiers with synthetic OCT data improved diagnostic performance across multiple tasks (Table 4). For age-related macular degeneration (AMD) detection on the RFMiD dataset, the augmented model achieved an AUROC of 0.963 (95% CI: 0.957–0.968) and an AUPRC of 0.763 (95% CI: 0.756–0.770), surpassing the fundus-only baseline (AUROC: 0.958, $p < 0.05$). Similarly, diabetic macular edema (DME) classification on IDRiD and MESSIDOR-2 datasets saw AUROC improvements to 0.859 and 0.925, respectively. Glaucoma detection on the PAPILA dataset also benefited, with AUROC rising from 0.840 to 0.845.

**Table 1. Performance of VAE in reconstructing OCT B-scan.**

| | | **Reconstruction metrics** | | | | **Generative metrics** |
|---|---|---|---|---|---|---|
| **Model** | **Subgroup** | **MSE ↓** | **MAE ↓** | **SSIM ↑** | **PSNR ↑** | **FID ↓** |
| VAE | - | 0.0089 | 0.0587 | 0.75 | 31.74 | 31.5125 |
| VAE | High Myopia | 0.0091 | 0.0591 | 0.74 | 32.88 | 31.5614 |

**Table 2. Performance of Fundus2OCT in generating 3D OCT video.**

| **Model** | **Subgroup** | **FVD32f ↓** | **FVD16f ↓** | **FID ↓** |
|---|---|---|---|---|
| F and S cond. | - | 155.7248 | 84.1921 | 40.8734 |
| Pretrained + F and S cond. | - | 141.0591 | 71.6402 | 37.1339 |
| Fundus2OCT (ours) | - | **139.3964** | **70.6415** | **35.4328** |
| Fundus2OCT (ours) | High myopia | 135.9953 | 69.7257 | 33.0827 |

“F and S cond.” = a diffusion model conditioned on the fundus image with retinal shape; “Pretrained + F and S cond.” = a pretrained unconditional diffusion model finetuned into a conditional generative model; Fundus2OCT= A diffusion model conditioned on the fundus image and outputting the 3DOCT image along with its retinal shape; HM = High Myopia

**Table 3. Performance of Fundus2OCT in generating retinal shape (mean ± std)**

| **Model** | **Subgroup** | **DICE ↑** | **IoU ↑** | **Chamfer distance ↓** |
|---|---|---|---|---|
| Fundus2OCT | - | 0.9675 ± 0.0056 | 0.9374 ± 0.0097 | 0.3462 ± 0.0809 |
| Fundus2OCT | High myopia | 0.9669 ± 0.0071 | 0.9363 ± 0.0118 | 0.3504 ± 0.0684 |

**Table 4. Effectiveness of synthetic datasets in enhancing downstream classification tasks**

| Dataset | Input | AUROC ↑ | AUPRC ↑ |
|---|---|---|---|
| IDRiD | Fundus | 0.846 (95% CI, 0.839-0.853) | 0.657 (95% CI, 0.634-0.680) |
| (DME) | F. + synthetic OCT | 0.859* (95% CI, 0.854-0.864) | 0.685* (95% CI, 0.661-0.710) |
| MESSIDOR2 | Fundus | 0.918 (95% CI, 0.879-0.956) | 0.591 (95% CI, 0.406-0.775) |
| (DME) | F. + synthetic OCT | 0.925 (95% CI, 0.896-0.953) | 0.609 (95% CI, 0.426-0.792) |
| PAPILA | Fundus | 0.840 (95% CI, 0.792-0.888) | 0.714 (95% CI, 0.644-0.785) |
| (Glaucoma) | F. + synthetic OCT | 0.845 (95% CI, 0.793-0.897) | 0.715 (95% CI, 0.648-0.807) |
| RFMiD | Fundus | 0.958 (95% CI, 0.952-0.963) | 0.758 (95% CI, 0.745-0.771) |
| (AMD) | F. + synthetic OCT | 0.963 (95% CI, 0.957-0.968) | 0.763 (95% CI, 0.756-0.770) |

*$P < 0.05$ (paired-sample T-test)

**Table 5. Clinician assessment by distinguishing the given OCT images as real or synthetic.**

| Metric | Clinician 1 | Clinician 2 |
|---|---|---|
| Accuracy | 0.49 | 0.57 |
| Sensitivity | 0.28 | 0.14 |
| Specificity | 0.70 | 1.00 |

## Discussion

The development of Fundus2OCT addresses a critical unmet need in global ophthalmology: enabling 3D OCT-based diagnostics in settings where OCT infrastructure is unavailable. By synthesizing high-fidelity, anatomically coherent OCT volumes directly from ubiquitous fundus photographs, our framework democratizes multimodal analysis—a capability previously limited to well-resourced clinics. Unlike prior efforts focused on multimodal prediction tasks or 2D OCT slice generation, Fundus2OCT achieves volumetric synthesis through a novel latent diffusion architecture, preserving spatial relationships across B-scans while minimizing computational costs. This advance opens avenues for large-scale "oculomics" screening using existing fundus imaging infrastructure.

Although previous studies have combined fundus and OCT data for improved diseases diagnosis[36], to the best of our knowledge, no attempt have been made to explore the direct synthesis of OCT volumes from fundus photographs (Fundus-to-3D-OCT). While multimodal prediction models can establish correlations between fundus and OCT images, they are often limited by the lack of large-scale 3D datasets for training and validation. In addition to their practical utility in data augmentation, the proposed Fundus2OCT framework leveraged semantic embeddings derived from entire fundus images, may enable the identification of detailed spatial features of multiple lesions that are not detectable on 2D-CFP.

Previous methods for generating OCT images have predominantly relied on unconditional GAN or VAE architectures, focusing primarily on generating individual B-scans (2D-OCT Generation).[16,17] To address the computational challenges of 3D-OCT volume synthesis, we introduce a latent diffusion model that generates OCT volumes conditioned on fundus latent. By compressing high-resolution OCT data into a low-dimensional latent space, our approach reduces the computational burden of direct 3D modeling, achieving significant improvements in training efficiency. Furthermore, Fundus2OCT incorporates embeddings from a pretrained fundus image foundation model—leveraging its inherent semantic richness—to generalize across diverse pathologies without requiring disease-specific training.

In clinician-led Turing tests, synthetic OCTs achieved near-chance distinguishability (49–57% accuracy; Table 5), confirming their anatomical plausibility. Notably, synthetic volumes retained diagnostic features for fluid accumulation (DME) and photoreceptor disruptions (AMD), as validated when they were used to augment classifiers. Fundus2OCT improved AUCs by 4.2–8.6% across four diseases (Table 4), demonstrating that synthetic OCTs provide biologically meaningful signals beyond mere data generation. This suggests a paradigm shift: fundus-based screening programs could now infer 3D pathological signatures without physical OCT devices, particularly impactful for detecting depth-resolved pathologies like subretinal fluid.

While Fundus2OCT is promising, certain limitations warrant consideration. First, the training data (UK Biobank) primarily represents a European demographic; expanding diversity across ethnicities, rare pathologies, and wide-field imaging devices will strengthen generalizability. Second, computational demands for 3D volume synthesis remain high, though latent diffusion reduces memory requirements compared to pixel-space approaches. Future work could optimize inference speed for real-time clinical use. Third, while synthetic OCTs improved classification, their utility in fine-grained tasks (e.g., quantifying macular thickness) and serve as "oculomics" for systemic diseases prediction[37] requires further validation against ground-truth biomarkers.

## Conclusion

Fundus2OCT establishes a new paradigm for cross-modal ophthalmic imaging, synthesizing 3D OCT volumes that are indistinguishable from real scans and clinically actionable. By augmenting fundus-based diagnostics with depth-resolved insights, this framework addresses critical challenges in data scarcity and multimodal integration. As generative AI continues to evolve, solutions like Fundus2OCT will play a pivotal role in advancing equitable, cost-effective healthcare—transforming accessible 2D imaging into a gateway for sophisticated 3D analysis.

## Data availability

UK Biobank data are available at https://www.ukbiobank.ac.uk/.

Data for ocular disease experiments are publicly available online and can be accessed through the following links: IDRiD (https://ieee-dataport.org/open-access/indian-diabetic-retinopathy-image-dataset-idrid), MESSIDOR2 (https://www.adcis.net/en/third-party/messidor2), RFMiD (https://ieee-dataport.org/open-access/retinal-fundus-multi-disease-image-dataset-rfmid), PAPILA (https://figshare.com/articles/dataset/PAPILA/14798004/1).

## Code availability

It will be placed on Code Ocean as required.

## References

1. Dahrouj, M. & Miller, J.B. Artificial intelligence (AI) and retinal optical coherence tomography (OCT). in *Seminars in Ophthalmology*, Vol. 36 341-345 (Taylor & Francis, 2021).
2. Bussel, I.I., Wollstein, G. & Schuman, J.S. OCT for glaucoma diagnosis, screening and detection of glaucoma progression. *British Journal of Ophthalmology* **98**, ii15 (2014).
3. Kim, B.Y., Smith, S.D. & Kaiser, P.K. Optical coherence tomographic patterns of diabetic macular edema. *Am J Ophthalmol* **142**, 405-412 (2006).
4. Elsharkawy, M., *et al.* Role of Optical Coherence Tomography Imaging in Predicting Progression of Age-Related Macular Disease: A Survey. *Diagnostics (Basel)* **11**(2021).
5. He, S., *et al.* Bridging the Camera Domain Gap With Image-to-Image Translation Improves Glaucoma Diagnosis. *Transl Vis Sci Technol* **12**, 20-20 (2023).
6. Shen, L., Zhao, W. & Xing, L. Patient-specific reconstruction of volumetric computed

tomography images from a single projection view via deep learning. *Nature biomedical engineering* **3**, 880-888 (2019).

7. Ktena, I.*, et al.* Generative models improve fairness of medical classifiers under distribution shifts. *Nature Medicine*, 1-8 (2024).
8. Shi, D.*, et al.* Translation of Color Fundus Photography into Fluorescein Angiography Using Deep Learning for Enhanced Diabetic Retinopathy Screening. *Ophthalmol Sci* **3**, 100401 (2023).
9. Song, F., Zhang, W., Zheng, Y., Shi, D. & He, M. A deep learning model for generating fundus autofluorescence images from color fundus photography. *Advances in ophthalmology practice and research* **3**, 192-198 (2023).
10. Shi, D.*, et al.* Fundus2Globe: Generative AI-Driven 3D Digital Twins for Personalized Myopia Management. *arXiv preprint arXiv:2502.13182* (2025).
11. Chen, R.*, et al.* EyeDiff: text-to-image diffusion model improves rare eye disease diagnosis. *arXiv preprint arXiv:2411.10004* (2024).
12. Shi, D.*, et al.* Cross-modality Labeling Enables Noninvasive Capillary Quantification as a Sensitive Biomarker for Assessing Cardiovascular Risk. *Ophthalmol Sci* **4**, 100441 (2024).
13. Chen, R.*, et al.* Translating color fundus photography to indocyanine green angiography using deep-learning for age-related macular degeneration screening. *npj Digital Medicine* **7**, 34 (2024).
14. Zhang, W.*, et al.* Fundus2Video: Cross-Modal Angiography Video Generation from Static Fundus Photography with Clinical Knowledge Guidance. in *Medical Image Computing and Computer Assisted Intervention – MICCAI* 689-699 (Springer Nature Switzerland, Morocco, 2024).
15. Wu, X.*, et al.* FFA Sora, video generation as fundus fluorescein angiography simulator. *arXiv preprint arXiv:2412.17346* (2024).
16. Tripathi, A., Kumar, P., Mayya, V. & Tulsani, A. Generating OCT B-Scan DME images using optimized Generative Adversarial Networks (GANs). *Heliyon* **9**(2023).
17. Kumar, A.J.S.*, et al.* Evaluation of generative adversarial networks for high-resolution synthetic image generation of circumpapillary optical coherence tomography images for glaucoma. *JAMA ophthalmology* **140**, 974-981 (2022).
18. Kaplan, S. & Lensu, L. Contrastive learning for generating optical coherence tomography images of the retina. in *International Workshop on Simulation and Synthesis in Medical Imaging* 112-121 (Springer, 2022).
19. Blattmann, A.*, et al.* Stable video diffusion: Scaling latent video diffusion models to large datasets. *arXiv preprint arXiv:2311.15127* (2023).
20. Rombach, R., Blattmann, A., Lorenz, D., Esser, P. & Ommer, B. High-resolution image synthesis with latent diffusion models. in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* 10684-10695 (2022).
21. Ho, J., Jain, A. & Abbeel, P. Denoising diffusion probabilistic models. *Advances in neural information processing systems* **33**, 6840-6851 (2020).
22. Sudlow, C.*, et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779 (2015).
23. Chua, S.Y.L.*, et al.* Cohort profile: design and methods in the eye and vision consortium of UK Biobank. *BMJ Open* **9**, e025077 (2019).

24. Shi, D.*, et al.* A Deep Learning System for Fully Automated Retinal Vessel Measurement in High Throughput Image Analysis. *Front Cardiovasc Med* **9**, 823436 (2022).

25. Pachade, S.*, et al.* Retinal fundus multi-disease image dataset (RFMiD): a dataset for multi-disease detection research. *Data* **6**, 14 (2021).

26. Porwal, P.*, et al.* Idrid: Diabetic retinopathy–segmentation and grading challenge. *Medical image analysis* **59**, 101561 (2020).

27. Abràmoff, M.D.*, et al.* Automated analysis of retinal images for detection of referable diabetic retinopathy. *JAMA ophthalmology* **131**, 351-357 (2013).

28. Decencière, E.*, et al.* Feedback on a publicly distributed image database: the Messidor database. *Image Analysis & Stereology*, 231-234 (2014).

29. Kovalyk, O.*, et al.* PAPILA: Dataset with fundus images and clinical data of both eyes of the same patient for glaucoma assessment. *Scientific Data* **9**, 291 (2022).

30. Zhang, R., Isola, P., Efros, A.A., Shechtman, E. & Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. in *Proceedings of the IEEE conference on computer vision and pattern recognition* 586-595 (2018).

31. Shi, D.*, et al.* Eyefound: a multimodal generalist foundation model for ophthalmic imaging. *arXiv preprint arXiv:2405.11338* (2024).

32. Ronneberger, O., Fischer, P. & Brox, T. 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* 234-241 (Springer, 2015).

33. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B. & Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems* **30**(2017).

34. Unterthiner, T.*, et al.* Towards accurate generative models of video: A new metric & challenges. *arXiv preprint arXiv:1812.01717* (2018).

35. Müller, P.L.*, et al.* Quantification and predictors of OCT-based macular curvature and dome-shaped configuration: results from the UK Biobank. *Investigative ophthalmology & visual science* **63**, 28-28 (2022).

36. Wu, J.*, et al.* Gamma challenge: glaucoma grading from multi-modality images. *Medical Image Analysis* **90**, 102938 (2023).

37. Li, C.*, et al.* Retinal oculomics and risk of incident aortic aneurysm and aortic adverse events: a population-based cohort study. *Int J Surg* (2025).

# Supplementary Files

- supplementary.pdf