TRANSLATION AS A REFERENCE FOR ONE-STEP OCT-TO-OCTA GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Optical coherence tomography (OCT) provides volumetric retinal structure, whereas OCT angiography (OCTA) visualizes retinal microvasculature through flow-sensitive contrast. Synthesizing OCTA from routinely acquired OCT can support vascular assessment when OCTA measurements are unavailable or unreliable. Despite recent progress, deterministic translation may underrepresent angiographic signals not fully specified by OCT, whereas direct full-target generation allows the complete OCTA volume to vary, including spatial organization that should remain tightly paired with the input. We hypothesize that using translation as a reference and reserving generation for residual angiographic variation simplifies one-step OCT-to-OCTA synthesis. In this paper, we introduce reference-centered generation and instantiate it with Pixel MeanFlow by decomposing each prediction into a subject-specific OCTA reference and a generated residual. The method requires neither vessel nor retinal-layer annotations and achieves state-of-the-art performance on OCTA-3M and OCTA-6M. Mechanism analysis further supports the hypothesis that reference centering simplifies the learned finite-interval transport.

1 INTRODUCTION

Retinal and choroidal vascular abnormalities are key manifestations of several vision-threatening diseases, making reliable assessment of retinal microvasculature clinically relevant Kashani et al. (2017). Optical coherence tomography (OCT) provides volumetric measurements of retinal structure Huang et al. (1991), whereas OCT angiography (OCTA) visualizes retinal and choroidal microvasculature through flow-sensitive contrast. Although OCTA is non-invasive, its image quality is vulnerable to motion, projection, segmentation, and signal-related artifacts Hormel et al. (2021); Anvari et al. (2021). Since structural OCT is routinely acquired in retinal examinations, synthesizing OCTA from OCT offers a practical route for missing-modality analysis when a usable OCTA measurement is unavailable. The task is nevertheless underdetermined: paired OCT and OCTA volumes depict the same retinal instance Li et al. (2024a), but structural anatomy does not uniquely specify all flow-sensitive angiographic signals. OCT-to-OCTA synthesis therefore lies between deterministic translation and generic conditional generation: the output must preserve subject-specific correspondence while representing angiographic variation unresolved by OCT.

Existing OCT-to-OCTA methods predominantly formulate the task as deterministic paired translation Pan et al. (2022); Li et al. (2024b); Chen et al. (2025); Khosravi et al. (2025). Recent approaches improve vascular reconstruction through volumetric modeling, cross-view alignment, projection consistency, and vessel- or layer-aware supervision Li et al. (2024b); Chen et al. (2025); Khosravi et al. (2025). Despite these methodological differences, they map each OCT volume to a single OCTA estimate. Such point-estimation objectives recover structures that are consistently supported by OCT, but may attenuate weakly constrained angiographic signals, including fine vessels, local contrast, and vascular continuity. This limitation motivates a generative formulation for the unresolved angiographic variation, while the generated output must still correspond to the observed retinal instance.

Generative modeling offers a natural way to represent angiographic signals that remain unresolved by structural OCT. Existing generative studies rely on diffusion and bridge models, whose iterative sampling is computationally demanding for volumetric synthesis Li et al. (2023); Liu et al.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

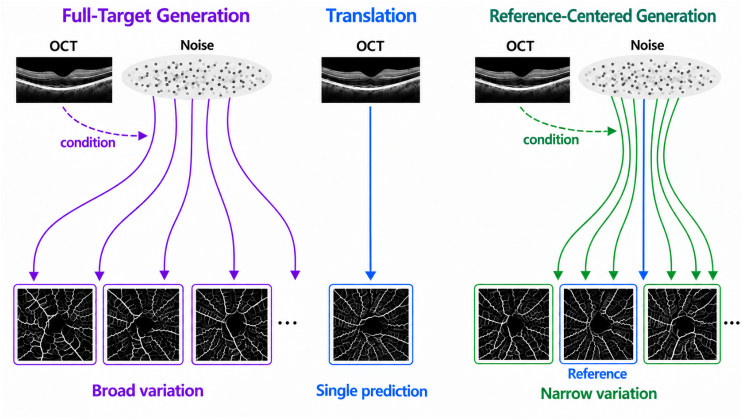


Figure 1: Motivation for reference-centered generation. Direct full-target generation allows the complete OCTA prediction to vary with the sampled state, whereas paired translation provides a single subject-specific estimate. Our formulation uses this estimate as a reference and models stochastic, reference-relative corrections. "Broad" and "narrow" depict the noise-induced predictive variation of the learned model for a fixed OCT input. The transport curves are schematic.

(2023a). Flow Matching learns continuous transport without simulating the training dynamics, while MeanFlow-style formulations further enable one-step generation without numerical ODE integration Lipman et al. (2022); Geng et al. (2026a;b); Lu et al. (2026). However, how flow-based one-step generation should exploit the dense correspondence in paired OCT-to-OCTA synthesis remains underexplored. A direct conditional formulation generates the complete OCTA target from noise without explicitly anchoring the output to the subject-specific estimate already supported by the paired input. The central question is therefore not only whether Flow Matching can be applied to OCT-to-OCTA synthesis, but how one-step transport should be organized around its paired translation structure.

We hypothesize that deterministic translation can serve as a useful subject-specific reference, allowing one-step generation to focus on the OCTA content not captured by that reference. This separates the paired content that should remain stable from the signals that are not uniquely determined by OCT, and should simplify the corresponding finite-interval transport. We therefore introduce reference-centered generation and instantiate it with Pixel MeanFlow. Rather than modifying the probability path or the MeanFlow objective, our formulation factors the image-valued prediction into a deterministic OCTA reference and a generated residual.

Experiments on the OCTA-3M and OCTA-6M subsets of OCTA-500 show that reference-centered generation achieves state-of-the-art volumetric reconstruction and projection MAE and PSNR without vessel or retinal-layer supervision. A mechanism audit further shows that reference centering reduces the relative finite-interval derivative correction by up to $41.1\times$. Together, these results support our hypothesis that reference centering provides a better formulation for densely paired OCT-to-OCTA generation.

Our contributions are threefold:

- We characterize OCT-to-OCTA synthesis as a strongly paired yet underdetermined problem and identify the limitations of both deterministic translation and direct full-target generation.
- We introduce reference-centered generation, which decomposes the one-step prediction as a deterministic subject-specific reference and a generated residual while retaining the original probability path and training objective.
- We achieve state-of-the-art reconstruction without vessel or retinal-layer annotations and provide mechanism evidence that reference centering substantially reduces the finite-interval derivative correction.

2 Related Work

2.1 Paired OCT-to-OCTA Translation

Existing OCT-to-OCTA methods predominantly formulate the task as deterministic paired image translation. MultiGAN employs multi-scale adversarial learning to improve angiographic reconstruction Pan et al. (2022), while TransPro combines volumetric context, projection consistency, and vessel-promoted guidance Li et al. (2024b). More recent methods further exploit complementary representations and supervision. MuTri aligns volumetric OCT, volumetric OCTA, and en-face views Chen et al. (2025), whereas XOCT introduces cross-dimensional multi-scale learning with retinal-layer supervision Khosravi et al. (2025). Collectively, these studies demonstrate the benefits of volumetric context, projection structure, and task-specific anatomical or vascular guidance for subject-specific OCTA reconstruction.

Despite these architectural and supervisory advances, existing methods map each OCT volume to a single OCTA prediction. Their primary goal is therefore improving the reconstruction of this estimate, rather than modeling OCTA content that is not fully determined by structural OCT. Our work instead studies how to introduce generative modeling into this paired setting while retaining the subject-specific correspondence between OCT and OCTA.

2.2 Generative Modeling for Paired Translation

Generative approaches extend paired image translation beyond pointwise reconstruction through adversarial or stochastic modeling. Conditional adversarial methods such as Pix2Pix combine a paired reconstruction loss with a conditional adversarial objective Isola et al. (2017). Diffusion- and bridge-based methods instead define iterative stochastic processes between source and target distributions. Image-to-image Schrödinger bridges treat the observed source as a structurally informative prior, whereas Brownian-bridge diffusion constructs a stochastic bridge directly between the source and target domains Liu et al. (2023a); Li et al. (2023). Although these bridge formulations incorporate the source into the generative process, their iterative sampling requires repeated network evaluations and can be particularly costly for volumetric synthesis.

This computational cost motivates generative models that learn finite-time transformations with one or few network evaluations. Flow Matching and Rectified Flow regress instantaneous velocity fields along continuous transport paths, but their standard inference generally requires numerical ODE integration Lipman et al. (2022); Liu et al. (2023b). Consistency Models learn trajectory-consistent mappings, whereas Flow Map Matching directly parameterizes two-time flow maps Song et al. (2023); Boffi et al. (2025). MeanFlow instead models finite-interval average velocity and thereby supports native one-step generation Geng et al. (2026a), while Improved MeanFlow reformulates its training target for more stable velocity regression Geng et al. (2026b). Pixel MeanFlow further represents the average velocity through an image-valued prediction, separating the network output from the velocity-space objective Lu et al. (2026). This separation makes it possible to adapt the prediction itself to the structure of a densely paired task without redefining the underlying transport process.

A direct conditional use of Pixel MeanFlow still predicts the complete OCTA target, with OCT entering through network conditioning but not explicitly through the output parameterization. Consequently, the prediction does not make the subject-specific reference supplied by the paired observation explicit. We address this gap by reparameterizing the image-valued prediction while retaining the original probability path and MeanFlow objective.

This additive parameterization is related to residual learning in image restoration Zhang et al. (2017), but differs in both the source and role of the reference. Conventional residual learning predicts a correction relative to the observed input, whereas our reference is produced by a deterministic cross-modal translator. It also differs from residual denoising diffusion models, which incorporate residual and noise components into the diffusion dynamics Liu et al. (2024). Reference-centered generation leaves the Pixel MeanFlow dynamics unchanged and learns a correction around the deterministic OCTA estimate under the original velocity-space objective.

3 METHOD

3.1 PIXEL MEANFLOW PRELIMINARIES

We adopt the data-to-noise time convention: $t = 0$ corresponds to the data distribution, and $t = 1$ to the Gaussian noise distribution. Generation follows the learned transport in reverse, from $t = 1$ to $t = 0$. Let x denote a structural OCT volume, y its paired OCTA volume, and $\epsilon \sim \mathcal{N}(0, I)$ Gaussian noise of the same shape.

Flow Matching provides a regression formulation for learning this transport from prescribed conditional probability paths Lipman et al. (2022). In our conditional synthesis setting, the path connects the paired OCTA target y to noise ϵ, while the structural OCT x is supplied to the velocity model.

$$z_t = (1 - t)y + t\epsilon, \qquad v^\star = \epsilon - y.$$

Here, z_t is the intermediate training state and $v^\star$ is its sample-wise velocity target. Regressing these targets yields the marginal instantaneous velocity field $v(z_t, x, t)$. Although each conditional path is linear, the induced marginal trajectory need not be linear because its local velocity aggregates sample pairs compatible with the current state Lipman et al. (2022); Geng et al. (2026a).

An instantaneous velocity field specifies only the local direction of transport and is ordinarily integrated over time during sampling. MeanFlow instead models the average velocity over a finite interval, so that the net displacement over that interval can be represented directly Geng et al. (2026a). Specifically, v denotes the local trajectory direction, whereas

$$u(z_t, x, r, t) = \frac{1}{t - r} \int_r^t v(z_\tau, x, \tau)\, \mathrm{d}\tau$$

denotes the average velocity from time r to time t. By continuous extension, $u(z_t, x, t, t) = v(z_t, x, t)$. The two quantities are related by the MeanFlow identity

$$v(z_t, x, t) = u(z_t, x, r, t) + (t - r)\dot{u}(z_t, x, r, t). \tag{1}$$

where $\dot{u}$ is the total derivative along the probability-flow trajectory. The derivative term measures the correction required to recover the endpoint instantaneous velocity from the finite-interval average.

Unlike MeanFlow, which directly parameterizes the average velocity, Pixel MeanFlow predicts an image-valued field g_θ and analytically converts it to u_θ (Lu et al., 2026):

$$u_\theta(z_t, x, r, t) = \frac{z_t - g_\theta(z_t, x, r, t)}{t}. \tag{2}$$

Thus, g_θ is the direct network output, while u_θ is a derived velocity-space quantity.

The supervision target $v^\star$ is instantaneous, whereas u_θ is a finite-interval average. Pixel MeanFlow obtains the local velocity by evaluating the same parameterization at the degenerate interval,

$$v_\theta(z_t, x, t) = u_\theta(z_t, x, t, t), \tag{3}$$

and uses it as the probability-flow direction in the Jacobian–vector product

$$\begin{aligned} \dot{u}_\theta &= \mathrm{JVP}_{(z_t, r, t)}[u_\theta; (v_\theta, 0, 1)]\,, \\ V_\theta &= u_\theta + (t - r)\,\mathrm{sg}(\dot{u}_\theta), \end{aligned} \tag{4}$$

where sg denotes stop-gradient.

The compound prediction V_θ is regressed against $v^\star$, so the network predicts in image space while the training objective remains in MeanFlow velocity space. At the complete interval, $g_\theta(\epsilon, x, 0, 1)$ directly gives the one-step image prediction.

3.2 REFERENCE-CENTERED PREDICTION

For a fixed OCT volume, direct conditional pMF predicts the complete OCTA target over different noisy states and sampled intervals, with the paired correspondence represented through conditioning on x. We instead make a subject-specific paired estimate explicit and model the state- and interval-dependent correction around that estimate. Let $f_\phi(x)$ be a deterministic OCT-to-OCTA predictor and $\Delta_\theta(z_t, x, r, t)$ a generative correction. We parameterize

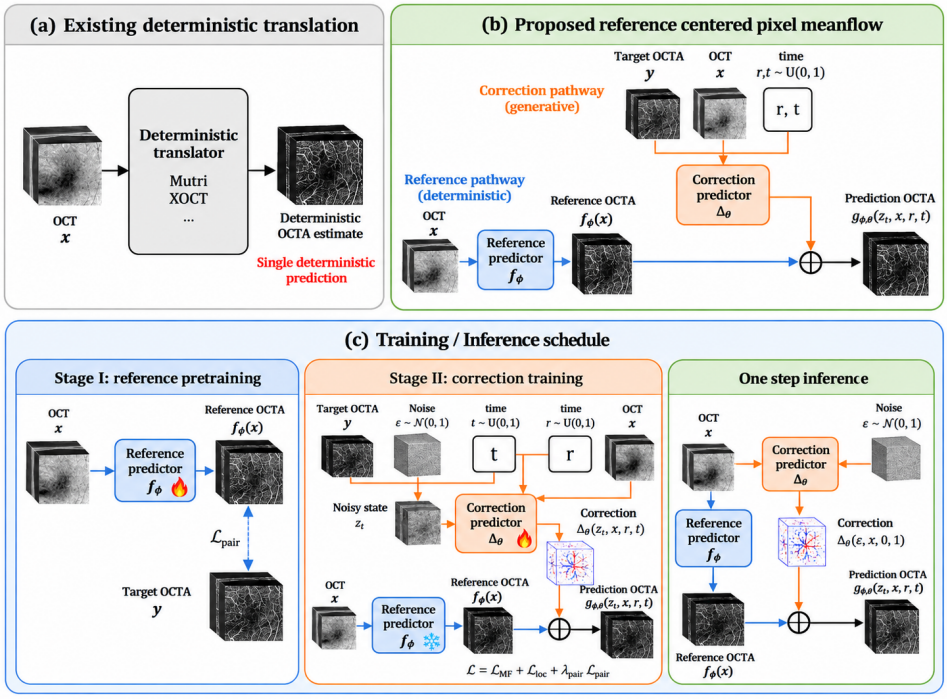


Figure 2: Overview of reference-centered pMF. Deterministic translation provides a subject-specific OCTA estimate, while the proposed formulation retains this estimate as a reference and adds a state- and interval-dependent generative correction. The lower panels show deterministic reference pretraining, subsequent correction training under the pMF objective, and one-step inference.

$$g_{\phi,\theta}(z_t, x, r, t) = f_\phi(x) + \Delta_\theta(z_t, x, r, t). \tag{5}$$

The reference $f_\phi(x)$ is shared across sampled states and intervals, while $\Delta_\theta(z_t, x, r, t)$ retains the state- and interval-dependent component of the prediction. Substituting Equation (5) into Equations (2)–(4) gives the corresponding $u_{\phi,\theta}$, $v_{\phi,\theta}$, and $V_{\phi,\theta}$ while retaining the original probability path and velocity-space objective.

3.3 Two-Stage Optimization

The decomposition in Equation (5) is trained in two stages, as illustrated in Figure 2. Stage I learns the deterministic reference. With $\Delta_\theta \equiv 0$, we optimize

$$\mathcal{L}_{\text{ref}} = \mathbb{E}_{(x,y)} \left[\|f_\phi(x) - y\|_1 \right]. \tag{6}$$

Stage II activates the zero-output-initialized correction branch Δ_θ and jointly optimizes the reference and correction parameters. Following pMF, the compound velocity $V_{\phi,\theta}$ is supervised against the instantaneous target $v^\star$, and the local velocity $v_{\phi,\theta}$ used as the JVP direction receives the corresponding local supervision. We additionally retain paired image supervision on $g_{\phi,\theta}$:

$$\mathcal{L}_{\text{flow}} = \mathbb{E}[\ell_{\text{flow}}(V_{\phi,\theta}, v^\star)], \tag{7}$$
$$\mathcal{L}_{\text{local}} = \mathbb{E}[\ell_{\text{local}}(v_{\phi,\theta}, v^\star)], \tag{8}$$
$$\mathcal{L}_{\text{pair}} = \mathbb{E}\left[\|g_{\phi,\theta}(z_t, x, r, t) - y\|_1\right]. \tag{9}$$

The Stage-II objective is

$$\mathcal{L}_{\text{pMF}} = \lambda_{\text{flow}}\mathcal{L}_{\text{flow}} + \lambda_{\text{local}}\mathcal{L}_{\text{local}} + \lambda_{\text{pair}}\mathcal{L}_{\text{pair}}. \tag{10}$$

For comparison, direct pMF optimizes the pMF objective from initialization, whereas reference-centered pMF first learns the deterministic reference before pMF optimization.

3.4 One-Step Inference

At inference, the complete interval $(r, t) = (0, 1)$ gives

$$\widehat{y} = f_\phi(x) + \Delta_\theta(\epsilon, x, 0, 1), \qquad \epsilon \sim \mathcal{N}(0, I). \tag{11}$$

The model produces a complete OCTA prediction using the deterministic reference and one complete-interval pMF correction evaluation, without iterative denoising or numerical ODE integration.

3.5 Reference Centering around a Transport-Flat Predictor

Reference centering changes how the image-valued pMF prediction is decomposed into a deterministic reference and a state- and interval-dependent correction. Let $s = (z_t, x, r, t)$ collect the noisy state, paired OCT conditioning, and interval variables entering the pMF predictor. For an arbitrary deterministic x-dependent function $h(x)$, consider

$$f'(x) = f(x) + h(x), \qquad \Delta'(s) = \Delta(s) - h(x). \tag{12}$$

The total image-valued prediction is unchanged:

$$f'(x) + \Delta'(s) = f(x) + \Delta(s) = g(s). \tag{13}$$

Thus the same pMF predictor can be represented by different reference–correction decompositions. Reference pretraining selects a particular Stage-II entry point within this parameterization.

This entry point has a useful transport structure. Define the transport-flat predictor class

$$\mathcal{M}_{\text{flat}} = \{g : \; g(z_t, x, r, t) = h(x)\}, \tag{14}$$

where $h(x)$ denotes any deterministic x-only predictor. For $g = h(x)$, substituting into Equation (2) gives

$$u_h = v_h = \frac{z_t - h(x)}{t}. \tag{15}$$

Along the probability-flow direction v_h,

$$\dot{u}_h = D_z u_h[v_h] + \partial_t u_h = \frac{v_h}{t} - \frac{z_t - h(x)}{t^2} = 0. \tag{16}$$

Therefore,

$$g \in \mathcal{M}_{\text{flat}} \quad \Longrightarrow \quad (t - r)\dot{u}_g = 0. \tag{17}$$

Equation (17) means that an x-only predictor requires no finite-interval derivative correction in the MeanFlow identity: its finite-interval average velocity coincides with the local velocity, $u_g = v_g$. We therefore refer to this predictor class as transport-flat.

Stage I explicitly searches inside this class. Let

$$\rho(x, y) = y - f(x) \tag{18}$$

denote the OCTA residual left by the learned deterministic reference. For the training path $z_t = (1 - t)y + t\epsilon$ with instantaneous target $v^\star = \epsilon - y$, the reference-only predictor satisfies

$$u_f - v^\star = v_f - v^\star = V_f - v^\star = \frac{\rho}{t}. \tag{19}$$

Thus Stage I finds a low-error predictor for which the finite-interval derivative correction vanishes exactly. Because the correction branch is zero-output initialized, Stage II begins from this transport-flat predictor.

For the frozen-reference variant, let $\theta = 0$ denote the Stage-II entry point and let τ denote Stage-II optimization time. We write the local dynamics as

$$\frac{d\theta}{d\tau} = F(\theta; \rho). \tag{20}$$

Informal Proposition 1. Residual-controlled departure from transport flatness.
Under local smoothness and finite-horizon stability, the Stage-II correction parameters, finite-interval derivative correction, and sampling sensitivity are first-order controlled by the reference residual:

$$\sup_{0\le\tau\le T}\|\theta(\tau)\| = O(\|\rho\|),$$
$$\sup_{0\le\tau\le T}\|(t-r)\dot{u}_{\theta(\tau)}\| = O(\|\rho\|),$$
$$\sup_{0\le\tau\le T}\|\nabla_\epsilon \widehat{y}_{\theta(\tau)}\| = O(\|\rho\|).$$

The formal statement and proof are given in Appendix **??**.

The fine-tuned variant starts from the same transport-flat predictor and subsequently adapts the reference together with the correction branch. These relations motivate the two quantities evaluated in Section 4.4: sampling-induced predictive variation σ^P_{pred} and the relative finite-interval correction η_{JVP}.

4 Experiments

4.1 Experimental Setup

Datasets and evaluation. We evaluate on OCTA-3M and OCTA-6M from OCTA-500 (Li et al., 2024a) and on OCTA2024 (Chen et al., 2025). For OCTA-500, we follow the XOCT split and evaluation protocol (Khosravi et al., 2025). OCTA-3M contains 140 training, 20 validation, and 40 test volumes, while OCTA-6M contains 200 training, 30 validation, and 70 test volumes. We report MAE, PSNR, and SSIM on the reconstructed 3D OCTA volumes and their mean en-face projections.

For OCTA2024, we follow the MuTri/TransPro split and evaluation convention, with 606 training, 40 validation, and 200 test volumes. Volume metrics are computed per B-scan and averaged over each volume, whereas projection metrics are evaluated on the mean en-face projection.

Training protocol. The reference-centered model is trained with 300 epochs of deterministic reference pretraining followed by 300 epochs of joint pMF optimization. The direct pMF baseline is optimized with the pMF objective for 300 epochs. Thus the main comparison matches the pMF optimization horizon rather than total training compute; reference-centered variants additionally incur Stage-I reference pretraining. Checkpoints in each stage are selected by validation projection MAE. Results are reported from one fixed training seed. Further implementation details are provided in Appendix **??**.

4.2 Comparison with Existing Methods

Table 1 reports the OCTA-500 comparison under the XOCT evaluation convention. Reference-centered pMF obtains the best reported 3D MAE, PSNR, and SSIM and the best projection MAE and PSNR on both OCTA-3M and OCTA-6M. XOCT retains the highest projection SSIM.

Table 2 reports the OCTA2024 comparison. Reference-centered pMF obtains the lowest volume MAE and projection MAE and the highest projection PSNR and SSIM among the listed methods. MuTri obtains the highest volume PSNR and SSIM.

4.3 Effect of Reference Centering

Table 3 isolates the effect of reference centering within the same pMF formulation. We compare direct full-target pMF with reference-centered variants in which the pretrained reference is either frozen or jointly updated during Stage-II optimization. Both reference-centered variants improve direct pMF in 3D MAE and projection MAE on OCTA-3M and OCTA-6M. Reference centering also consistently reduces the interval-velocity gap, with the jointly fine-tuned model giving the smallest gap on both datasets.

Table 1: OCTA-500 comparison under the XOCT protocol (Khosravi et al., 2025). P: mean en-face projection; Gen.: generation; Trans.: deterministic translation; Ref.-Gen.: reference-centered generation; "–": unavailable.

Method	Type	3D MAE ↓	3D PSNR ↑	3D SSIM ↑	P MAE ↓	P PSNR ↑	P SSIM ↑
OCTA-3M							
BBDM (Li et al., 2023)	Gen.	–	–	–	25.33	18.14	0.384
Pix2Pix (Isola et al., 2017)	Gen.	–	–	–	27.18	17.47	0.340
MultiGAN (Pan et al., 2022)	Gen.	–	–	–	24.73	18.32	0.382
BBDM3D (Choo et al., 2024)	Gen.	2.69	30.71	0.883	22.51	19.27	0.457
Pix2Pix3D (Deng et al., 2023)	Gen.	2.88	29.45	0.885	21.61	19.47	0.541
TransPro (Li et al., 2024b)	Trans.	3.30	28.46	0.866	20.93	19.83	0.573
XOCT (Khosravi et al., 2025)	Trans.	2.78	30.08	0.893	20.54	19.90	**0.596**
MuTri (Chen et al., 2025)	Trans.	3.04	29.65	0.874	25.60	18.28	0.384
Reference-centered pMF	**Ref.-Gen.**	**2.50**	**31.11**	**0.903**	**19.94**	**20.21**	0.570
OCTA-6M							
BBDM (Li et al., 2023)	Gen.	–	–	–	25.73	17.92	0.346
Pix2Pix (Isola et al., 2017)	Gen.	–	–	–	26.43	17.76	0.303
MultiGAN (Pan et al., 2022)	Gen.	–	–	–	24.69	18.43	0.332
BBDM3D (Choo et al., 2024)	Gen.	2.48	31.44	0.889	18.29	20.81	0.436
Pix2Pix3D (Deng et al., 2023)	Gen.	2.34	31.23	0.903	17.21	21.33	0.510
TransPro (Li et al., 2024b)	Trans.	2.72	29.92	0.887	17.42	21.24	0.539
XOCT (Khosravi et al., 2025)	Trans.	2.34	31.26	0.905	16.21	21.85	**0.574**
MuTri (Chen et al., 2025)	Trans.	2.68	30.21	0.885	21.91	19.28	0.397
Reference-centered pMF	**Ref.-Gen.**	**2.06**	**32.43**	**0.916**	**15.02**	**22.47**	0.558

Table 2: OCTA2024 comparison under the MuTri evaluation (Chen et al., 2025). V and P denote volume and mean en-face projection; SSIM is reported in percentage points. Prior results are reproduced from the MuTri comparison.

	Volume (V)			Projection (P)		
Method	MAE ↓	PSNR ↑	SSIM ↑	MAE ↓	PSNR ↑	SSIM ↑
Pix2Pix2D (Isola et al., 2017)	0.1301	40.65	69.78	0.01471	35.17	86.39
Pix2Pix3D (Deng et al., 2023)	0.1275	41.87	71.92	0.01324	36.88	86.74
9B18CN U-Net (Lee et al., 2019)	0.1402	40.53	70.76	0.01586	34.79	84.37
AdjacentGAN (Li et al., 2020)	0.1352	41.97	71.43	0.01459	36.18	85.97
VQ-I2I (Chen et al., 2022)	0.1187	41.25	71.31	0.01132	37.29	87.24
Palette (Saharia et al., 2022)	0.1254	41.40	71.07	0.01241	37.54	87.33
TransPro (Li et al., 2024b)	0.1098	42.69	72.52	0.01056	37.85	87.61
MuTri (Chen et al., 2025)	0.0828	**43.38**	**79.66**	0.00870	39.65	90.31
Reference-centered pMF	**0.0386**	41.85	78.42	**0.00746**	**40.42**	**91.26**

4.4 Mechanism Analysis

We examine whether reference centering changes output-level sensitivity to stochastic sampling and the learned finite-interval transport. For each OCT volume x, we draw $K = 16$ independent complete-interval predictions

$$\widehat{y}_k = g_{\phi,\theta}(\epsilon_k, x, 0, 1), \qquad \epsilon_k \sim \mathcal{N}(0, I). \tag{21}$$

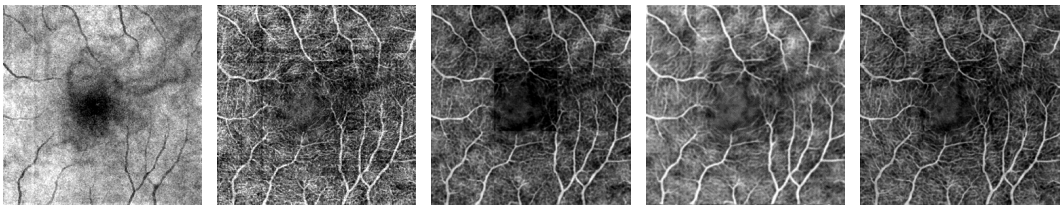
We measure the noise-induced variation of the mean en-face projection as

$$\sigma^{P}_{\mathrm{pred}}(x) = \sqrt{\frac{1}{HW}\sum_{p=1}^{HW} \mathrm{Var}_k\left[P(\widehat{y}_k)_p\right]}. \tag{22}$$

We report this quantity averaged across validation volumes.

At the transport level, the MeanFlow identity in Equation 1 gives

$$\eta_{\mathrm{JVP}} = \frac{\mathbb{E}[\|(t-r)\dot{u}\|_1]}{\mathbb{E}[\|u\|_1] + \varepsilon}. \tag{23}$$



(a) OCT (b) Ground truth (c) Direct pMF (d) MuTri (e) Ref.-pMF

Figure 3: Representative OCT-to-OCTA reconstruction with matched display settings. Ref.-pMF denotes reference-centered pMF.

Table 3: Effect of reference centering on OCTA-500. P: mean en-face projection; γ_{int}: interval-velocity gap on validation data.

Dataset	Formulation	3D MAE ↓	P MAE ↓	P SSIM ↑	γ_{int} ↓
OCTA-3M	Direct pMF	2.5674	20.5409	0.5589	7.535×10^{-3}
	Frozen reference	**2.4734**	19.9465	**0.5744**	4.012×10^{-3}
	Fine-tuned ref.	2.4957	**19.9444**	0.5698	$\mathbf{2.598 \times 10^{-3}}$
OCTA-6M	Direct pMF	2.1067	15.8061	0.5486	7.742×10^{-4}
	Frozen reference	2.0821	15.1792	**0.5668**	3.842×10^{-5}
	Fine-tuned ref.	**2.0629**	**15.0181**	0.5581	$\mathbf{1.209 \times 10^{-5}}$

which measures the finite-interval derivative correction relative to the average velocity. The jointly optimized reference-centered model reduces η_{JVP} by $3.4\times$ on OCTA-3M and $41.1\times$ on OCTA-6M. Reference-centered pMF shows concurrent reductions in output-level noise-induced variation and transport-level finite-interval correction.

Table 4: Predictive variation and finite-interval correction on validation data.

Dataset	Formulation	σ^{P}_{pred} ↓	η_{JVP} ↓
OCTA-3M	Direct pMF	4.740×10^{-3}	1.955×10^{-2}
	Reference-centered	1.255×10^{-3}	5.772×10^{-3}
OCTA-6M	Direct pMF	3.939×10^{-4}	3.575×10^{-3}
	Reference-centered	3.143×10^{-5}	8.698×10^{-5}

5 Conclusion

We introduced reference-centered one-step generation for OCT-to-OCTA synthesis. The formulation uses deterministic translation as a subject-specific OCTA reference and reserves generation for reference-relative angiographic variation. Experiments on OCTA-3M, OCTA-6M, and OCTA2024 demonstrate strong reconstruction performance, while reference-design ablations consistently improve direct full-target pMF in 3D and projection MAE. Mechanism analysis shows that reference centering reduces both noise-induced predictive variation and the finite-interval derivative correction. Together, these results support reference-centered generation as an effective formulation for strongly paired and underdetermined OCT-to-OCTA synthesis.

Reproducibility Statement

The main text specifies the pMF formulation, reference-centered parameterization, training objectives, datasets, and evaluation conventions. The supplementary appendix provides the formal theoretical statement and derivations, model architectures and initialization, training and sampling algorithms, preprocessing and optimization settings, checkpoint selection, inference, and mechanism-audit protocols.

AI Use Statement

Generative AI tools were used to assist with language editing, retrieval and discovering related works, manuscript restructuring, analysis-code drafting, and the development and critical examination of mathematical exposition. The authors remain responsible for verifying all scientific claims, derivations, experimental results, code, citations, and final manuscript content.

References

Pasha Anvari, Maryam Ashrafkhorasani, Abbas Habibi, and Khalil Ghasemi Falavarjani. Artifacts in optical coherence tomography angiography. *Journal of ophthalmic & vision research*, 16(2): 271, 2021.

Nicholas Matthew Boffi, Michael Samuel Albergo, and Eric Vanden-Eijnden. Flow map matching with stochastic interpolants: A mathematical framework for consistency models. *Transactions on Machine Learning Research*, 2025. ISSN 2835-8856. URL https://openreview.net/forum?id=cqDH0e6ak2.

Yu-Jie Chen, Shin-I Cheng, Wei-Chen Chiu, Hung-Yu Tseng, and Hsin-Ying Lee. Vector quantized image-to-image translation. In *Computer Vision – ECCV 2022*, volume 13676 of *Lecture Notes in Computer Science*, pp. 440–456. Springer, 2022. doi: 10.1007/978-3-031-19787-1_25.

Zhuangzhuang Chen, Hualiang Wang, Chubin Ou, and Xiaomeng Li. Mutri: Multi-view tri-alignment for oct to octa 3d image translation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 20885–20894, 2025.

Kyobin Choo, Youngjun Jun, Mijin Yun, and Seong Jae Hwang. Slice-consistent 3d volumetric brain ct-to-mri translation with 2d brownian bridge diffusion model. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 657–667. Springer, 2024.

Kangle Deng, Gengshan Yang, Deva Ramanan, and Jun-Yan Zhu. 3d-aware conditional image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4434–4445, 2023.

Zhengyang Geng, Mingyang Deng, Xingjian Bai, Zico Kolter, and Kaiming He. Mean flows for one-step generative modeling. *Advances in Neural Information Processing Systems*, 38:75460–75482, 2026a.

Zhengyang Geng, Yiyang Lu, Zongze Wu, Eli Shechtman, J. Zico Kolter, and Kaiming He. Improved mean flows: On the challenges of fastforward generative models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 30467–30476, June 2026b.

Tristan T Hormel, David Huang, and Yali Jia. Artifacts and artifact removal in optical coherence tomographic angiography. *Quantitative imaging in medicine and surgery*, 11(3):1120, 2021.

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.

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*, pp. 1125–1134, 2017.

Amir H Kashani, Chieh-Li Chen, Jin K Gahm, Fang Zheng, Grace M Richter, Philip J Rosenfeld, Yonggang Shi, and Ruikang K Wang. Optical coherence tomography angiography: a comprehensive review of current methods and clinical applications. *Progress in retinal and eye research*, 60: 66–100, 2017.

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 *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 695–705. Springer, 2025.

Cecilia S. Lee, Ariel J. Tyring, Yue Wu, Sa Xiao, Ariel S. Rokem, Nicolaas P. DeRuyter, Qinqin Zhang, Adnan Tufail, Ruikang K. Wang, and Aaron Y. Lee. Generating retinal flow maps from structural optical coherence tomography with artificial intelligence. *Scientific Reports*, 9(1):5694, 2019. doi: 10.1038/s41598-019-42042-y.

Bo Li, Kaitao Xue, Bin Liu, and Yu-Kun Lai. Bbdm: Image-to-image translation with brownian bridge diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern Recognition*, pp. 1952–1961, 2023.

Mingchao Li, Kun Huang, Qiuzhuo Xu, Jiadong Yang, Yuhan Zhang, Zexuan Ji, Keren Xie, Songtao Yuan, Qinghuai Liu, and Qiang Chen. Octa-500: a retinal dataset for optical coherence tomography angiography study. *Medical image analysis*, 93:103092, 2024a.

Pei Lin Li, Céline O'Neil, Samin Saberi, Kenneth Sinder, Kathleen Wang, Bingyao Tan, Zohreh Hosseinaee, Kostadinka Bizhevat, and Vasudevan Lakshminarayanan. Deep learning algorithm for generating optical coherence tomography angiography (OCTA) maps of the retinal vasculature. In *Applications of Machine Learning 2020*, volume 11511, pp. 1151109. SPIE, 2020. doi: 10.1117/12.2568629.

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, 2024b.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The eleventh international conference on learning representations*, 2022.

Guan-Horng Liu, Arash Vahdat, De-An Huang, Evangelos A Theodorou, Weili Nie, and Anima Anandkumar. I2sb: image-to-image schrödinger bridge. In *Proceedings of the 40th International Conference on Machine Learning*, pp. 22042–22062, 2023a.

Jiawei Liu, Qiang Wang, Huijie Fan, Yinong Wang, Yandong Tang, and Liangqiong Qu. Residual denoising diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2773–2783, 2024.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International conference on learning representations (ICLR)*, 2023b.

Yiyang Lu, Susie Lu, Qiao Sun, Hanhong Zhao, Zhicheng Jiang, Xianbang Wang, Tianhong Li, Zhengyang Geng, and Kaiming He. One-step latent-free image generation with pixel mean flows. In *Forty-third International Conference on Machine Learning*, 2026. URL `https://openreview.net/forum?id=WUK8JIeetF`.

Bing Pan, Zexuan Ji, and Qiang Chen. Multigan: Multi-domain image translation from oct to octa. In *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, pp. 336–347. Springer, 2022.

Chitwan Saharia, William Chan, Huiwen Chang, Chris A. Lee, Jonathan Ho, Tim Salimans, David J. Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 Conference Proceedings*, pp. 15:1–15:10. Association for Computing Machinery, 2022. doi: 10.1145/3528233.3530757.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *Proceedings of the 40th International Conference on Machine Learning*, pp. 32211–32252, 2023.

Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26 (7):3142–3155, 2017.