Unsupervised OCT image interpolation using deformable registration and generative models

Shuwen Wei[1](✉), Samuel W. Remedios[1,2], Zhangxing Bian[1],
Shimeng Wang[1], Junyu Chen[1,3], Yihao Liu[4], Bruno Jedynak[5],
Tin Y. A. Liu[6], Shiv Saidha[7], Peter A. Calabresi[7],
Jerry L. Prince[1], and Aaron Carass[1]

[1]Image Analysis and Communications Laboratory,
Dept. of Electrical and Computer Engineering,
Johns Hopkins University, Baltimore, MD 21218, USA
✉ `swei14@jhu.edu`
[2]Dept. of Computer Science,
Johns Hopkins University, Baltimore, MD 21218, USA
[3]Dept. of Radiology and Radiological Science, [7]Dept. of Neurology,
Johns Hopkins School of Medicine, Baltimore, MD 21287, USA
[4]Dept. of Electrical and Computer Engineering,
Vanderbilt University, Nashville, TN 37235
[5]Dept. of Mathematics and Statistics,
Portland State University, Portland, OR 97201, USA
[6]Dept. of Ophthalmology,
Wilmer Eye Institute, Baltimore, MD 21287, USA

Abstract. Optical coherence tomography (OCT) images are often acquired as highly anisotropic volumes, where the scanning step is dense along the fast axis but sparse along the slow axis. This affects image analysis, such as image registration for longitudinal alignment. To create more isotropic volumes, bicubic interpolation can be used along the slow axis, but it generally produces blurry features. Registration-based interpolation can reduce blurriness, but often fails to generate realistic OCT images. Deep generative models can sample realistic images, but lack the structural consistency constraints required for interpolation. In this paper, we propose an unsupervised image interpolation method that combines registration-based interpolation with a deep generative model to overcome their individual limitations and improve the structural accuracy and realism of interpolated OCT images. We compare the proposed method with both bicubic and registration-based interpolation on real OCT datasets, and show that it achieves the best interpolation performance.

Keywords: Unsupervised learning · Optical coherence tomography · Image interpolation · Deformable registration · Generative model.

1 Introduction

Optical coherence tomography (OCT) uses low-coherence interferometry to achieve noninvasive high-resolution retinal imaging [9]. It is widely used in the monitoring of the retina for various retinal and neurological diseases [14, 18, 19]. The optical wave coherence and natural photon fluctuations of an OCT laser cause speckle and noise in raw OCT images, affecting retinal layer visibility and thickness measurements [23]. Some commercial OCT devices, such as Heidelberg Spectralis OCT, acquire and average multiple images around the same scanning cross section to reduce speckle and noise. However, it lowers the acquisition speed, and retinal OCT images are often acquired as highly anisotropic volumes as a consequence. The scanning step is dense along the fast axis but sparse along the slow axis. Typical Spectralis retinal OCT volumes have a twenty times difference in their axes step-sizes. The highly anisotropic acquisitions affect the longitudinal monitoring of subtle retinal layer changes. This is because the alignment of retinal OCT acquisitions is difficult to reproduce at two different time points, and algorithmic alignment of the two OCT volumes is not feasible as registration does not work well with highly anisotropic volumes. Therefore, retinal OCT image interpolation along the slow axis is desirable for accurate longitudinal analysis.

Bicubic interpolation achieves a good balance between the interpolation quality and the processing speed compared with other-order B-spline methods, and it produces good reconstructions when the signal is sufficiently sampled [22]. However, the sampling along the slow axis in OCT volumes is far less than the sufficient sampling rate, and thus bicubic interpolation of OCT often produces blurry images. To obtain sharper images, registration-based interpolation algorithms have been developed [13]. They use small deformation approximation, time-stationary, or time-dependent vector fields [4] to gradually warp one slice to a neighboring slice, and thus generate interpolated slices at any position between two slices. The small deformation approximation linearly scales displacement fields warping from one slice to the other, but is not consistent under inverse warping. Both the time-stationary and time-dependent vector field methods are invertible, and can be used to warp the slice from either direction by choosing the correct integration interval. By generating smooth transitions of anatomical structures between adjacent slices, registration-based interpolation algorithms produce less blurry images than bicubic interpolation. However, they do not necessarily generate realistic images. Deep generative models achieve superior performance in sampling realistic images. These models include generative adversarial network (GAN) [5], variational autoencoder (VAE) [10], normalizing flow (NF) [16], and more recently, denoising diffusion probabilistic model (DDPM) [8], score-based generative model (SGM) [21], and flow matching [12]. However, they have not yet been used for OCT image interpolation because they lack structural consistency constraints required for interpolation.

In this paper, we propose an unsupervised image interpolation method that combines the benefits of a registration model and a deep generative model to produce realistic and anatomically consistent interpolated slices. We use a stationary vector field deformable registration model and a DDPM for the generative model. Our key idea is to find an image that aligns with the anatomical trajectory defined by the registration process while staying within the data distribution learned by the generative model. We first describe the registration and the generative model, and then describe our interpolation approach by combining these two models. We compare the proposed method with both bicubic and a registration-based interpolation on a real OCT dataset. The results show that the proposed method achieves the best interpolation performance.

2 Method

Registration model. Let retinal OCT images be in a subset of the function space $\mathbb{G} = \{g | g : \Omega \subset \mathbb{R}^2 \to \mathbb{R}\}$, where the domain of these functions Ω is a C^∞-differentiable manifold. We assume $\forall f, m \in \mathbb{G} : \exists \varphi \in \mathrm{Diff}(\Omega) : \varphi^* m = f$, where φ^* is the pullback defined as the composition $\varphi^* m = m \circ \varphi$. We call f the fixed image, and m the moving image. The diffeomorphism group $\mathrm{Diff}(\Omega)$ is a Lie group, and its Lie algebra is the tangent space $T_e\mathrm{Diff}(\Omega)$ of the Lie group $\mathrm{Diff}(\Omega)$ at the identity element $e \in \mathrm{Diff}(\Omega)$, which is isomorphic to the left-invariant vector field $\mathcal{L}(\mathrm{Diff}(\Omega))$, a subbundle of the tangent bundle $T(\mathrm{Diff}(\Omega))$. $\forall v \in \mathcal{L}(\mathrm{Diff}(\Omega))$, there is an associated flow $\phi : \mathbb{R} \times \Omega \to \Omega$ with $\forall t \in \mathbb{R}$,

$$\frac{d\phi_t}{dt} = v \circ \phi_t, \quad \phi_0 = e, \tag{1}$$

and $\{\phi_t | t \in \mathbb{R}\}$ forms a one-parameter subgroup where $\forall t, s \in \mathbb{R} : \phi_t \phi_s = \phi_{t+s}$. Then, ϕ_1 is obtained by integration or the exponential map,

$$\phi_1 = e + \int_0^1 v \circ \phi_\tau d\tau = \lim_{n \to \infty} \Big(e + \frac{v_e}{n} \Big)^n = \exp(v_e), \tag{2}$$

where $v_e \in T_e\mathrm{Diff}(\Omega)$ is the generator of the one-parameter subgroup, and it is obtained from $v \in \mathcal{L}(\mathrm{Diff}(\Omega))$ by $v_e = v(e)$. It is more efficient to compute the exponential map by scaling-and-squaring [1, 2] than integration.

The overall registration model is shown in Fig. 1(a). We use a customized U-Net [17] with parameter ψ to predict a stationary vector field v_e from f and m by minimizing the following objective,

$$\mathcal{L} = \mathrm{E}_{f,m}\Big[\gamma_1\Big(|| \exp(v_e)^* m - f||_2\Big) + \gamma_2\Big(||\nabla v_e^x||_2 + ||\nabla v_e^y||_2\Big) + \gamma_3\Big(1 - \mathrm{DSC}\Big)\Big], \tag{3}$$

where $v_e = R_\psi(f, m)$, and f and m are sampled from the training dataset. There are three terms in Eq. 3, and γ's are their corresponding weights. The first term calculates the mean squared error (MSE) between the fixed image f and the warped moving image $\exp(v_e)^* m$, where $|| \cdot ||_2$ is the L_2 norm. The second term is to regularize the smoothness of v_e, where ∇ is the gradient operator, and v_e^x and v_e^y are the horizontal and vertical components of v_e, respectively. The third term is the Dice loss between fixed segmentation masks and warped moving segmentation masks that we incorporate to improve the registration performance.

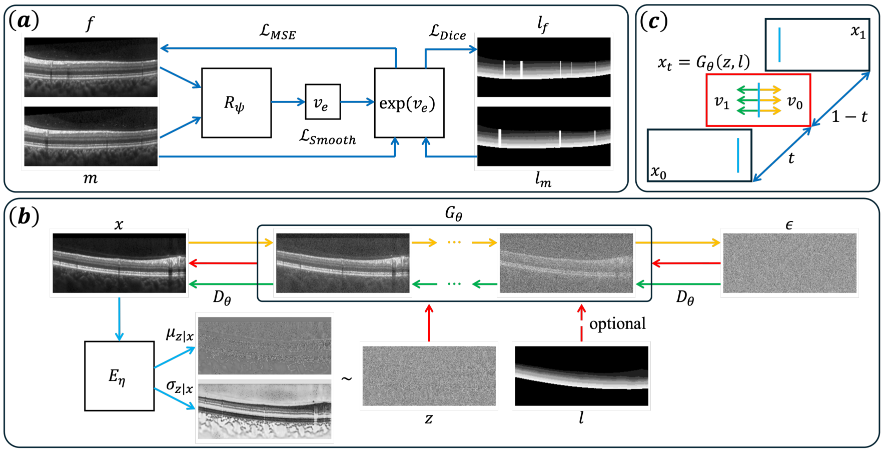


Fig. 1. **(a)** Overall architecture of the registration model. The registration model R_ψ takes in an image pair, f and m, to produce a stationary vector field v_e. The deformation field $\exp(v_e)$ warps m to f, and l_m to l_f. **(b)** Overall architecture of the generative model. The yellow arrow is the forward diffusion process. The green arrow is the reverse diffusion, which uses the diffusion model D_θ to gradually remove noise. The multi-step reverse diffusion is grouped together and denoted as G_θ, which takes in noise and conditions, and produces the final image. The blue arrow indicates the encoder network E_η. **(c)** Schematic of the interpolation model. The red rectangle is the target slice located at t between two slices x_0 and x_1. The blue vertical lines are the matching features. The yellow arrows are the vector field v_0 whose associated deformation field warps x_0 to the target slice, and the green arrows are the vector field v_1 whose associated deformation field warps x_1 to the target slice. We assume that $\forall t \in [0, 1] : (1 - t)v_0 + tv_1 = 0$.

The segmentation masks include eight retinal layers obtained using AURA [11], and a vessel plus shadow label that we find through outlier detection from the mean intensity projection within the retinal pigment epithelium complex (RPE). Two segmentation masks and corresponding B-scans are shown in Fig. 1(a).

Generative model. We use a conditional autoencoder DDPM as the deep generative model [8, 15]. We do this to make the generative model invertible, and also incorporate additional conditions during data generation. The generative model is shown in Fig. 1(b). It consists of an encoder network and a diffusion network, where both use a diffusion model U-Net [3]. The encoder network E_η with parameters η, models the posterior distribution of the latent variable z given an observed OCT image x by a Gaussian distribution $q_\eta(z|x) = \mathcal{N}(\mu_{z|x}, \sigma^2_{z|x})$ through $\mu_{z|x}, \log \sigma_{z|x} = E_\eta(x)$, where $\mu_{z|x}$ is the mean, $\sigma_{z|x}$ is the standard deviation, and we use the logarithm to make sure that $\sigma_{z|x}$ is positive.

The deep generative model uses a DDPM [8], i.e., variance preserving (VP) stochastic differential equation (SDE), and linear beta schedule β_s for $s \in \{1, \ldots, T\}$ with $T = 1000$ time steps. Let $\bar{\alpha}_\tau = \prod_{s=1}^{\tau}(1 - \beta_s)$, the diffusion network D_θ with parameters θ predicts the Gaussian noise ϵ from the time step τ, the noisy image $\sqrt{\bar{\alpha}_\tau}x + \sqrt{1-\bar{\alpha}_\tau}\epsilon$ with x being the clean image, the condition from the latent space z, and optional segmentation labels l, i.e., $\hat{\epsilon} = D_\theta\left(\tau, \sqrt{\bar{\alpha}_\tau}x + \sqrt{1-\bar{\alpha}_\tau}\epsilon, z, l\right)$. Both network parameters η and θ are jointly learned by minimizing the following objective,

$$\mathcal{L} = \mathrm{E}_{x,\epsilon,\tau,z}\left[\delta_1\Big(||\hat{\epsilon}-\epsilon||_2\Big) + \delta_2\Big(D_{KL}(q_\eta(z|x)||p(z))\Big)\right], \tag{4}$$

where x is sampled from training data, $\epsilon \sim \mathcal{N}(0, I)$, $\tau \sim U(1,\ldots,T)$, and $z \sim \mathcal{N}(\mu_{z|x}, \sigma^2_{z|x})$. There are two terms in Eq. 4, and δ's are their corresponding weights. The first term calculates the MSE between the predicted noise $\hat{\epsilon}$ by D_θ and the ground truth noise ϵ. The second term regularizes the latent space of z, where the prior $p(z)$ is a standard multivariate Gaussian $\mathcal{N}(0, I)$.

Interpolation model. After training the DDPM, we use the sampling strategy of a denoising diffusion implicit model (DDIM) [20], which allows deterministic mapping and a large jump between the Gaussian probability space and the data space. We fix the initial Gaussian noise, and the sampling process becomes a deterministic function $G_\theta(z, l)$ that maps the latent variable z and optional segmentation labels l to real data. To interpolate a slice x_t at position $t \in [0, 1]$ between two adjacent slices $x_0, x_1 \in \mathbb{G}$, we find the optimal latent variable z by,

$$\hat{z} = \arg\min_z ||(1-t)R_\psi(G_\theta(z,l), x_0) + tR_\psi(G_\theta(z,l), x_1)||_2\,, \tag{5}$$

where R_ψ and G_θ are the trained registration and generative models, respectively. The intuition behind Eq. 5 is that the vector field whose associated deformation field warps the slice x_0 to x_t should have a linear relationship based on t to the vector field whose associated deformation field warps the slice x_1 to x_t, as shown in Fig. 1(c). We use the stochastic gradient descent approach to find the optimal $\hat{z}$ for Eq. 5, and the interpolated slice is obtained by $x_t = G_\theta(\hat{z}, l)$. Note that we train two independent DDPMs, where one is conditioned on the segmentation label l and the other one is not. For the DDPM that is conditioned on the segmentation label l, the segmentation label at $t \in (0, 1)$ is obtained by bicubic interpolation of the boundary segmentation of x_0 and x_1.

3 Results

Dataset. The retinal OCT dataset contains 377 Spectralis volumes with each volume having 49 B-scans, and an additional 15 Spectralis volumes with these volumes having 97 B-scans. The 377 Spectralis volumes comes from 178 subjects, representing 296 unique eyes. While the additional 15 Spectralis volumes come from 10 subjects, representing 15 unique eyes. Between the two cohorts there are no overlapping subjects. Both cohorts have a lateral scanning range of 6 $mm \times 6$ mm. All B-scans have dimensions 496×1024, where 496 pixels are along the A-scan axis and 1024 pixels are along the B-scan axis. We crop each B-scan from 496 to 256 to remove background, and resize them to 128×512 to save memory usage during training. We split the 377 Spectralis volumes that have only 49 B-scans into a training dataset of 357 OCT volumes and an validation dataset of 20 OCT volumes. We use the remaining 15 Spectralis volumes that have 97 B-scans as a testing dataset.

Training. The registration model R_ψ is trained on samples of adjacent B-scans in the same OCT volume as the fixed and moving pair, because we do not register images with greater spacing during testing. We also enforce horizontal displacement by restricting the registration network to only predict the horizontal component v_e^x and setting the vertical component v_e^y to 0. We choose the weights empirically in Eq. 3 to be $\gamma_1 = 1$, $\gamma_2 = 0.01$, and $\gamma_3 = 0.01$, set batch size to 32, use Adam optimizer with learning rate 1×10^{-3}, and apply 6 recursive steps in the scaling-and-squaring. The diffusion model D_θ is trained on samples of each individual B-scan. We train two independent diffusion models, where one is conditioned on the segmentation labels and the other one is not. For both cases, we choose the weights empirically in Eq. 4 to be $\delta_1 = 1$ and $\delta_2 = 1 \times 10^{-7}$, set the batch size to 8, and use Adam optimizer with learning rate 2.5×10^{-5}. Models weights are selected based on the best performance on the validation dataset.

Comparison methods. We compare the performance of the proposed interpolation method with and without segmentation labels as condition for DDPM. We also compare the proposed interpolation method with bicubic interpolation [22] and two registration-based interpolation methods, including forward warp (FW) from x_0 and backward warp (BW) from x_1 to the target position $t \in [0, 1]$ by,

$$x_t^{\mathrm{FW}} = \exp\left(-tR_\psi(x_0, x_1)\right)^* x_0, \quad x_t^{\mathrm{BW}} = \exp\left((1-t)R_\psi(x_0, x_1)\right)^* x_1, \tag{6}$$

where R_ψ is the registration model, exp is the exponential map in Eq. 2, $x_0, x_1 \in \mathbb{G}$ are two adjacent slices.

Experiments. We use the testing dataset to compare the five interpolation methods. We select 49 B-scans with even indices $\{0, 2, \ldots, 96\}$ out of 97 B-scans from 15 Spectralis volumes, and we run interpolations to predict 48 B-scans with odd indices $\{1, 3, \ldots, 95\}$, and compare them with ground truth in terms of peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) and show results in Table 1. An example of results is shown in Fig. 2, where x_0 and x_1 are two even numbered B-scans, and ground truth shows the original odd numbered B-scan between them. We see that bicubic interpolation is blurry and duplicates the vessel plus shadow from the adjacent B-scans. The forward and backward warp methods produce less blurry but unrealistic images, where the vessel plus shadows are curved. The proposed methods with or without segmentation labels generate both high quality images with the correct anatomy.

We also run a deep learning based segmentation algorithm [6, 7] on both the ground truth and interpolated B-scans to obtain eight retinal layer labels. We obtain the vessel plus shadow label using the outlier detection from the RPE mean intensity projection. However, for bicubic interpolation, we approximate its vessel plus shadow label by union from adjacent B-scans, because its contrast is low and difficult to segment using the outlier detection. Moreover, for the forward and backward warp methods, we obtain their vessel plus shadow label by applying the same deformation field to the neighboring B-scan labels, because their vessel

Table 1. Mean PSNR (Std. Dev.) and mean SSIM (Std. Dev.) for the interpolation methods. Bolded numbers are the best in the row. Asterisk indicates statistical significance (i.e. Wilcoxon test comparing 1$^{\text{st}}$ and 2$^{\text{nd}}$ best results gave p-value < 0.05).

	Bicubic	**Forward Warp**	**Backward Warp**	**Proposed (No Label)**	**Proposed (Label)**
PSNR	25.291 (1.441)	24.733 (1.558)	25.025 (1.519)	25.801 (1.548)	**25.879*** (1.547)
SSIM	0.639 (0.067)	0.629 (0.069)	0.635 (0.067)	0.681 (0.071)	**0.683*** (0.070)

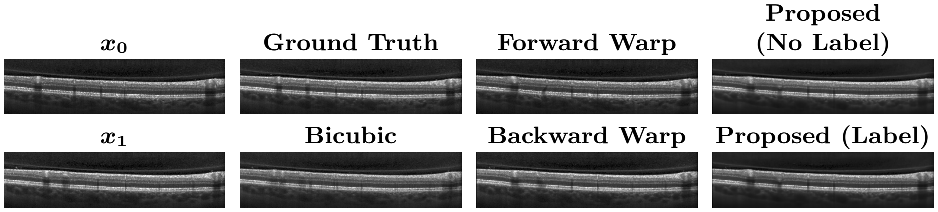


Fig. 2. Comparison of the five interpolation methods. The B-scans x_0 and x_1 are two adjacent B-scans, and the ground truth B-scan shows the original B-scan $x_{0.5}$.

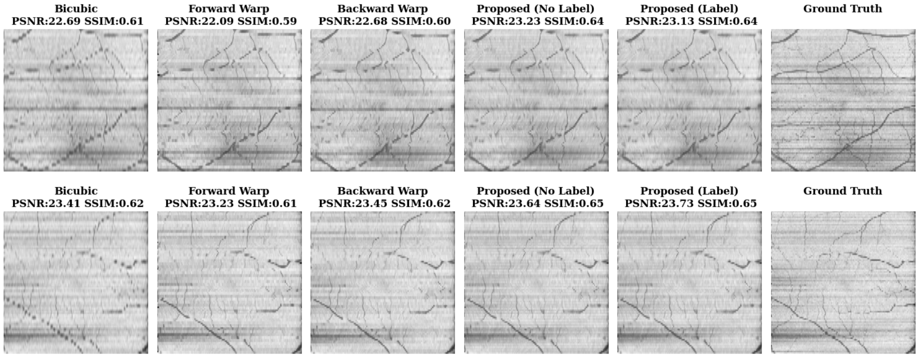


Fig. 3. Comparison of the five interpolation algorithms through en-face projection in the RPE layer of the retinal OCT volumes.

plus shadows are curved, We then calculate the Dice coefficients between labels on the ground truth and interpolated B-scans using these nine segmentation labels. The results are shown in Table 2. The results in Tables 1 and 2 show that the proposed interpolation method with or without labels achieves better performance than the bicubic, forward warp, and backward warp methods, and by adding segmentation labels as a condition, the performance improves further.

To better visualize how these interpolation methods affect the retinal OCT volumetric structure, we do a mean intensity projection in the RPE layer, and

Table 2. Mean Dice coefficients (Std. Dev.) for different retinal regions and interpolation methods. Bolded numbers are the best in the row. Asterisk indicates statistical significance (i.e. Wilcoxon test comparing 1^{st} and 2^{nd} best results gave p-value < 0.05).

	Bicubic	**Forward Warp**	**Backward Warp**	**Proposed (No Label)**	**Proposed (Label)**
RNFL	0.879 (0.048)	0.890 (0.042)	0.897 (0.039)	0.900 (0.049)	**0.910*** (0.036)
GCIPL	0.904 (0.043)	0.922 (0.033)	0.927 (0.030)	0.927 (0.033)	**0.933*** (0.029)
OPL	0.890 (0.042)	0.900 (0.037)	0.906 (0.034)	0.908 (0.035)	**0.916*** (0.031)
INL	0.861 (0.049)	0.864 (0.047)	0.871 (0.045)	0.875 (0.046)	**0.886*** (0.042)
ONL	0.921 (0.037)	0.941 (0.027)	0.945 (0.024)	0.945 (0.024)	**0.948*** (0.024)
IS	0.864 0.068)	0.877 (0.066)	0.884 (0.062)	0.885 (0.062)	**0.890*** (0.062)
OS	0.900 (0.046)	0.918 (0.040)	0.924 (0.037)	0.924 (0.037)	**0.926*** (0.038)
RPE	0.904 (0.043)	0.922 (0.037)	0.927 (0.034)	0.928 (0.034)	**0.929*** (0.035)
Vessel Shadow	0.290 (0.133)	0.313 (0.208)	0.341 (0.210)	0.425 (0.200)	**0.426** (0.200)

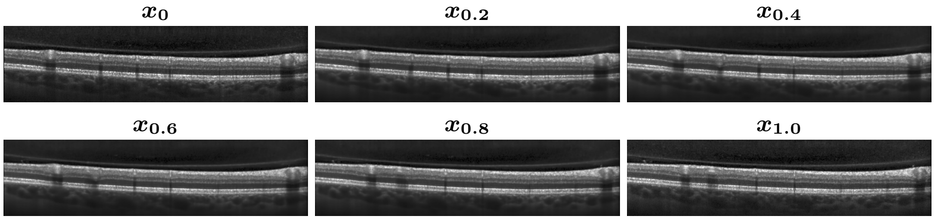
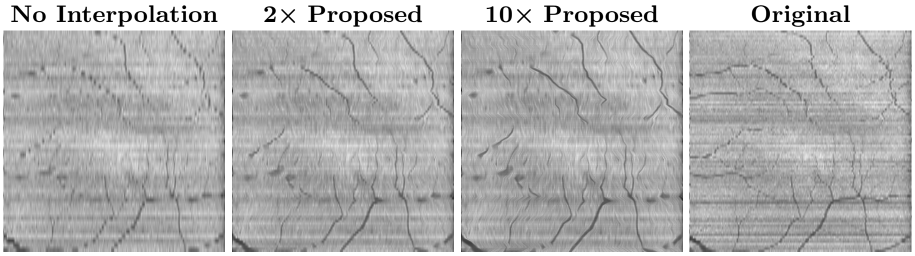


Fig. 4. 5× interpolation between x_0 and x_1 using the proposed method. Note how the blood vessel plus shadows smoothly transition between the interpolated slices.

show results in Fig. 3. Compared with the bicubic interpolation, the registration-based interpolation methods are better at connecting the discontinuous vessels. Moreover, the proposed interpolation further improves upon the registration-based interpolation both qualitatively and quantitatively. We also investigate if the proposed method can achieve an arbitrary dense interpolation. To do so, we use the proposed method with labels to perform 2×, 5× and 10× interpolation along the slow axis. We can see smooth anatomical transition between two



No Interpolation **2× Proposed** **10× Proposed** **Original**

Fig. 5. En-face images generated by 2× and 10× interpolation using the proposed method. "No Interpolation" is 49 B-scans and "Original" is the original 97 B-scans. The blood vessels in this en-face image appear dark.

adjacent slices in Fig. 4 for 5× interpolation, and denser and smoother vessels from RPE projection in Fig. 5 beyound ground truth for 10× interpolation.

4 Conclusion

In this paper, we propose an unsupervised OCT image interpolation method using registration and generative models. We compare the proposed method with both bicubic interpolation and registration-based interpolation on a real retinal OCT dataset, and show that the proposed method achieves the best interpolation performance. A limitation of the proposed interpolation method is that it does not recover the horizontal vessels as good as vertical vessels. It also remains to be demonstrated that the proposed interpolation method is beneficial for retinal OCT volumetric registration and improves longitudinal analyses.

Acknowledgments. This work is supported in part by the NIH through NEI grant R01-EY032284 (PI: J.L. Prince), NINDS grant R01-NS082347 (PI: P.A. Calabresi), as well as NIA grants RO1-AG021155 (PI: S. C. Johnson) and R01-AG027161 (PI: S. C. Johnson). This material is partially supported by the National Science Foundation grant number 2136228 (PI: J. Gopalakrishnan) and Graduate Research Fellowship grant number DGE-1746891 (S.W. Remedios). The work was made possible in part by a Johns Hopkins University Discovery Grant (PI: A. Carass).

Disclosure of Interests. J.L.P. and A.C. have received royalties from JuneBrain Inc. P.A.C. is PI on grants to JHMI from Genentech and the Myelin Repair Foundation, and has received personal consulting Honoria from Idorsia, Spolia Therapeutics, Novartis, and Lilly. The other authors have no competing interests to declare.

References

1. Arsigny, V., et al.: A log-euclidean framework for statistics on diffeomorphisms. In: 9^{th} International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2006). pp. 924–931 (2006)

2. Ashburner, J.: A fast diffeomorphic image registration algorithm. NeuroImage **38**(1), 95–113 (2007)
3. Cardoso, M.J., et al.: MONAI: An open-source framework for deep learning in healthcare. arXiv preprint arXiv:2211.02701 (2022)
4. Chen, J., et al.: A survey on deep learning in medical image registration: New technologies, uncertainty, evaluation metrics, and beyond. Medical Image Analysis p. 103385 (2024)
5. Goodfellow, I., et al.: Generative Adversarial Nets. In: Advances in Neural Information Processing Systems. vol. 27 (2014)
6. He, Y., et al.: Fully convolutional boundary regression for retina OCT segmentation. In: 22nd International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2019). pp. 120–128 (2019)
7. He, Y., et al.: Structured layer surface segmentation for retina OCT using fully convolutional regression networks. Medical Image Analysis **68**, 101856 (2021)
8. Ho, J., et al.: Denoising Diffusion Probabilistic Models. In: Advances in Neural Information Processing Systems. vol. 33, pp. 6840–6851 (2020)
9. Huang, D., et al.: Optical coherence tomography. Science **254**(5035), 1178–1181 (1991)
10. Kingma, D.P.: Auto-Encoding Variational Bayes. In: International Conference on Learning Representations (2013)
11. Lang, A., et al.: Retinal layer segmentation of macular OCT images using boundary classification. Biomed. Opt. Express **4**(7), 1133–1152 (2013)
12. Lipman, Y., et al.: Flow Matching for Generative Modeling. In: International Conference on Learning Representations (2023)
13. Penney, G., et al.: Registration-based interpolation. IEEE Transactions on Medical Imaging **23**(7), 922–926 (2004).
14. Petzold, A., et al.: Optical coherence tomography in multiple sclerosis: a systematic review and meta-analysis. The Lancet Neurology **9**(9), 921–932 (2010)
15. Preechakul, K., et al.: Diffusion Autoencoders: Toward a Meaningful and Decodable Representation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10619–10629 (2022)
16. Rezende, D., Mohamed, S.: Variational Inference with Normalizing Flows. In: International Conference on Machine Learning. pp. 1530–1538 (2015)
17. Ronneberger, O., et al.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: 18th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2015). pp. 234–241 (2015)
18. Rothman, A., et al.: Retinal measurements predict 10-year disability in multiple sclerosis. Annals of Clinical and Translational Neurology **6**(2), 222–232 (2019)
19. Saidha, S., et al.: Optical coherence tomography reflects brain atrophy in multiple sclerosis: a four-year study. Annals of Neurology **78**(5), 801–813 (2015)
20. Song, J., et al.: Denoising Diffusion Implicit Models. In: International Conference on Learning Representations (2021)
21. Song, Y., et al.: Score-Based Generative Modeling through Stochastic Differential Equations. In: International Conference on Learning Representations (2021)
22. Unser, M.: Splines: A perfect fit for signal and image processing. IEEE Signal Processing Magazine **16**(6), 22–38 (1999)
23. Wei, S., et al.: Recurrent Self Fusion: Iterative Denoising for Consistent Retinal OCT Segmentation. In: International Workshop on Ophthalmic Medical Image Analysis. pp. 42–51 (2023)