Vessel-promoted OCT to OCTA image translation by heuristic contextual constraints

Shuhan Li [a], Dong Zhang [b], Xiaomeng Li [b,c,*], Chubin Ou [d], Lin An [e], Yanwu Xu [f], Weihua Yang [g], Yanchun Zhang [h], Kwang-Ting Cheng [b]

[a] *Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong, China*
[b] *Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong, China*
[c] *HKUST Shenzhen-Hong Kong Collaborative Innovation Research Institute, Futian, Shenzhen, China*
[d] *Weizhi Meditech (Foshan) Co., Ltd, China*
[e] *Guangdong Weiren Meditech Co., Ltd, China*
[f] *South China University of Technology, and Pazhou Lab, China*
[g] *Shenzhen Eye Institute, Shenzhen Eye Hospital, Jinan University, China*
[h] *Department of Ophthalmology, Shaanxi Eye Hospital, Xi'an People's Hospital (Xi'an Fourth Hospital), Affiliated People's Hospital of Northwest University, Xi'an, China*

ARTICLE INFO

ABSTRACT

Optical Coherence Tomography Angiography (OCTA) is a crucial tool in the clinical screening of retinal diseases, allowing for accurate 3D imaging of blood vessels through non-invasive scanning. However, the hardware-based approach for acquiring OCTA images presents challenges due to the need for specialized sensors and expensive devices. In this paper, we introduce a novel method called TransPro, which can translate the readily available 3D Optical Coherence Tomography (OCT) images into 3D OCTA images without requiring any additional hardware modifications. Our TransPro method is primarily driven by two novel ideas that have been overlooked by prior work. The first idea is derived from a critical observation that the OCTA projection map is generated by averaging pixel values from its corresponding B-scans along the Z-axis. Hence, we introduce a hybrid architecture incorporating a 3D adversarial generative network and a novel **H**euristic **C**ontextual **G**uidance (HCG) module, which effectively maintains the consistency of the generated OCTA images between 3D volumes and projection maps. The second idea is to improve the vessel quality in the translated OCTA projection maps. As a result, we propose a novel **V**essel **P**romoted **G**uidance (VPG) module to enhance the attention of network on retinal vessels. Experimental results on two datasets demonstrate that our TransPro outperforms state-of-the-art approaches, with relative improvements around 11.4% in MAE, 2.7% in PSNR, 2% in SSIM, 40% in VDE, and 9.1% in VDC compared to the baseline method. The code is available at: https://github.com/ustlsh/TransPro.

1. Introduction

Optical coherence tomography (OCT) is a non-invasive imaging technique that provides three-dimensional cross-sectional visualization of retinal structures (Huang et al., 1991). It plays a crucial role in the diagnosis of various eye diseases, including age-related macular degeneration, diabetic retinopathy, and glaucoma (Fujimoto and Swanson, 2016). Building upon the OCT imaging technique, an advanced modality, OCT angiography (OCTA) was developed to display blood flow in retinal microvasculature. OCTA images are obtained by repeatedly scanning the OCT sectional images at the same location of the retina, allowing for the detection of dynamic signals resulting from the movement of erythrocytes within the vascular system (Kashani et al., 2017).

The availability of OCTA images has been extremely useful in the early detection of some eye disorders, such as choroidal neovascularization, which occurs in wet age-related macular degeneration (Chen et al., 2021). Fig. 1 illustrates the 3D volumes of OCT and OCTA, which are depicted in (a) and (d), respectively. The B-scan image, shown in (b) and (e), represents a single slice within the 3D volume. On the other hand, the projection map, displayed in (c) and (f), is generated by averaging the pixel values along the *z*-axis of the 3D volume.

Acquiring OCTA images requires additional hardware and software modifications to mitigate the impact of involuntary tissue motions caused by heartbeat, respiration, and eye movement. The high cost
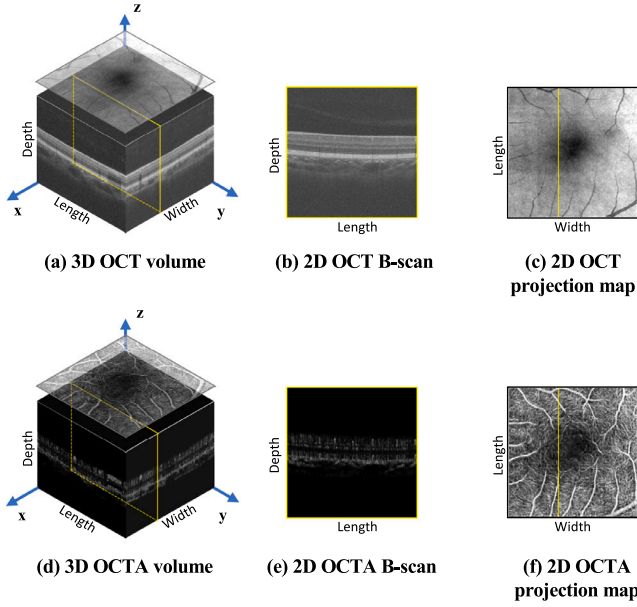

Fig. 1. Examples of OCT images (in (a)–(c)) and OCTA images (in (d)–(e)). 3D OCT volume (in (a)) or 3D OCTA volume (in (d)) consists of sequential 2D OCT or OCTA B-scan images. 2D B-scan images, which are one slice of the 3D volumes, are shown in (b) and (e). In addition, by computing the mean of pixel values, the 3D volume can be projected along Z-axis (the Depth direction) to obtain 2D projection map. The 2D OCT and OCTA projection maps are displayed in (c) and (f), respectively.

of the motion tracking module has resulted in a lower adoption rate of OCTA compared to OCT. To address this challenge, deep learning methods can serve as alternative approaches for obtaining the cumbersome and costly OCTA modality from the readily available and cost-effective OCT modality. Several studies have proposed various approaches for OCT to OCTA translation. One common approach is to use a UNet (Ronneberger et al., 2015) as the backbone and employ generative-adversarial learning frameworks for pixel-to-pixel image translation (Isola et al., 2017; Zhu et al., 2017). These existing methods can be categorized into two groups based on the type of input images they utilize. The first group of methods (Lee et al., 2019; Zhang et al., 2021b; Li et al., 2020b) focuses on translating 2D OCT B-scan images into their corresponding OCTA B-scan images. For instance, Lee et al. (2019) introduces a deep learning method for OCT to OCTA translation, utilizing an encoder–decoder structure to generate 2D OCTA B-scan images from paired 2D OCT B-scan images. More recent methods incorporate additional information, such as texture features (Zhang et al., 2021b) or adjacent B-scans (Li et al., 2020b), to enhance the quality of the translated OCTA B-scan images. The second group of methods (Pan et al., 2022) operate on 2D OCT projection maps and generate corresponding 2D OCTA projection maps.

However, the existing OCT to OCTA translation methods suffer from two limitations. Firstly, these methods focus on either B-scan images or projection maps. As illustrated in Fig. 1, the 2D B-scan image represents a single slice of 3D volume, translating OCT B-scan to OCTA B-scan disregards the continuity across images. On the other hand, the 2D projection map is obtained through dimensional reduction, which may result in the loss of complete information present in 3D volume. As a result, learning with one type of images may lead to low quality in the other type. Secondly, current OCTA translation methods do not consider vessel information, which leads to poor quality in vascular areas of translated OCTA projection maps. This problem arises from the use of pixel-wise reconstruction loss functions that do not differentiate between vessel and non-vessel pixels. As OCTA images predominantly consist of non-vessel areas, the model tends to prioritize accuracy in non-vessel regions over vascular areas.

To address the above limitations, we propose a novel TransPro method with two main ideas. The first idea is derived from a critical observation that the OCTA projection map is generated by averaging pixel values from its corresponding B-scans along the Z-axis. Hence, it is crucial to maintain this property by leveraging both B-scan and projection maps of 3D OCT. To achieve this, we propose a hybrid network architecture, consisting of a 3D generative network and a 2D generative network. ***The 3D generative network takes 3D OCT volume as input*** and generates 3D OCTA volume as output. During training, the 3D generative network is optimized by minimizing the disparities between the generated OCTA images and the ground-truth, both along the 3D volumes and the projection maps. Concurrently, ***the 2D generative network takes OCT projection maps as inputs*** and generates OCTA projection maps as outputs. To leverage the additional information acquired by the 2D network from OCT projection maps, we introduce a novel **heuristic contextual guidance (HCG)**, which minimizes the discrepancies between the outputs of the 2D network and the projection maps generated by the 3D network. By doing so, it serves as an additional mechanism to ensure the consistency of the generated OCTA projection maps between the 2D and 3D networks.

The second idea aims to enhance the vessel area during the generation process to prevent excessive attention on the background regions. To achieve this, we introduce a **Vessel Promoted Guidance (VPG)**. By utilizing a pre-trained 2D retinal vascular segmentation model to segment vessels in 2D OCTA projection maps, our proposed VPG module ensures semantic consistency between the vascular segmentation predictions of the translated OCTA images and the ground-truth images. By using these two novel approaches, our TransPro method achieves enhanced consistency between B-scans and their corresponding OCTA projection maps, as well as improved quality of blood vessels in the translated OCTA images.

We evaluate the proposed method on the two subsets of OCTA-500 public dataset (Li et al., 2020a). We compute Mean Absolute Error (MAE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM) on both translated OCTA B-scan images and projection maps. Moreover, we propose two novel metrics, named Vessel Density Error (VDE) and Vessel Density Correlation (VDC), to evaluate the quality of vessel regions in the translated OCTA projection maps. The experimental results demonstrate the superiority of the proposed approach over state-of-the-art methods. Our method achieves a remarkable 11.4% relative improvement in the Mean Absolute Error (MAE) metric on the OCTA-3M daatset. Similarly, on the larger OCTA-6M dataset, the MAE was reduced by 5.1% compared to prior art. Furthermore, our approach also yields substantial enhancements in vessel quality metrics. On the OCTA-3M dataset, there is a 40.8% relative improvement in VDE and a 14.0% boost in VDC. The OCTA-6M results are also impressive, with a 21.6% VDE increase and an 8.6% VDC increase.

The main contributions of this paper are summed up as follows:

- We identify a novel motivation in OCT-OCTA translation, which is to maintain the consistency of the translated OCTA images across both 3D volumes and projection maps.
- We introduce a novel approach called Heuristic Contextual Guidance (HCG) to provide additional guidance aimed at preserving the quality of OCTA projection maps.
- We propose a novel Vessel Promoted Guidance (VPG) module, which significantly enhances the accuracy and quality of vessel regions in translated OCTA projection maps.
- We evaluate our TransPro on translated OCTA B-scan images and projection maps. In addition to the commonly used metrics, we propose several vessel-specific metrics to quantitatively evaluate the vessel quality of translated OCTA in projection maps.
- Extensive experimental results demonstrate that our TransPro significantly outperforms state-of-the-art techniques across a range of performance criteria on both OCTA-3M and OCTA-6M datasets, including both traditional image translation metrics and clinically relevant vessel quality metrics. These results highlight the effectiveness and advantages of our proposed TransPro approach.

2. Related work

2.1. OCT and OCTA image analysis

OCT and OCTA are widely used imaging modalities for diagnosing various retinal diseases (Huang et al., 1991; Fujimoto and Swanson, 2016). OCT enables non-invasive real-time imaging of the three-dimensional cross-sectional structure of the retina, but it does not provide sufficient information for detecting choroidal neovascularization or recognizing the morphological characteristics of microvasculature, which are important for diagnosing retinal diseases (Spaide et al., 2018). As a complementary modality to OCT, OCTA is introduced to visualize the retinal blood vessels (Roisman and Goldhardt, 2017; Eladawi et al., 2018). OCTA images can be generated by detecting temporal signal changes caused by moving red blood cells in multiple OCT images obtained from the same location (Ferrara et al., 2016; Kashani et al., 2017). However, OCTA devices require additional sensors and software to alleviate the eye motion artifacts. The higher cost limits their wider usage.

2.2. Modality translation in medical images

In clinical diagnosis, the utilization of multi-modality images can provide comprehensive and complementary information that can be advantageous for treatment (Zhang et al., 2021b; Li et al., 2020b). However, the acquisition of multi-modality images poses challenges due to factors such as cost, time, and potential harm to the body (Zhang et al., 2022). To overcome this issue, a promising solution is to perform modality translation from an easily obtainable source image modality to a more challenging target image modality (McNaughton et al., 2023). Examples include translating magnetic resonance imaging (MRI) to computerized tomography (CT) (Hsu et al., 2022; Parrella et al., 2023; Zhao et al., 2023), MRI to positron emission tomography (PET) (Rajagopal et al., 2022; Hussein et al., 2022), and PET to CT (Li et al., 2022). Among these, the translation of OCT to OCTA images has been an active area of research in recent years.

The existing OCT to OCTA translation methods can be divided into two groups according to the types of input images. The first group of methods translates OCT B-scan images to their corresponding OCTA B-scan images. For example, the initial work (Lee et al., 2019) employs an encoder–decoder model structure to generate 2D OCTA B-scan images from paired 2D OCT B-scan images. Building upon this structure, Texture-UNet (Zhang et al., 2021b) extracts the texture features of OCT B-scans to guide the translation process. Recently, some methods utilize Generative Adversarial Network (GAN) to achieve improved performance in the OCT to OCTA translation task. AdjacentGAN (Li et al., 2020b) employs three adjacent OCT B-scan images as input to guide the translation of the middle OCTA B-scan image. This approach utilizes limited contextual information from neighboring OCT slices to enhance the continuity among slices. The second group of methods focuses on the translation of 2D projection maps. MultiGAN (Pan et al., 2022) proposes multiple GANs to translate OCT projection maps to OCTA projection maps obtained between different retinal layers. However, the current methods only consider either B-scan images or projection maps for OCT to OCTA translation, which results in incomplete information about 3D volumes. This omission ignores the relevance between the two types of OCT and OCTA data, leading to low-quality translations in one type when training with the other one. Moreover, the quality of vascular areas in translated OCTA projection maps is relatively poor due to the lack of specific vessel enhancement guidance. To address these issues, this paper introduces a 3D GAN with an HCG module to constrain the translation of hybrid data formats, and a VPG module to improve the quality of vessel areas.

2.3. Retinal vascular segmentation

Retinal vascular segmentation of OCTA images is a crucial and fundamental task in the diagnosis of retinal diseases. The segmentation provides information on the density and morphological structures of retinal capillary networks, which can be utilized for disease diagnosis or robotic surgery (Ronneberger et al., 2015). Various methods have been proposed for retinal vascular segmentation based on the classical UNet framework in recent years (Ma et al., 2020; Mou et al., 2019; Li et al., 2020a). For instance, CSNet adds self-attention blocks to the encoder and decoder network to emphasize curvilinear vessel structures (Mou et al., 2019). Besides, DCSSNet is designed as a semi-supervised segmentation model that uses both labeled and unlabeled images to mitigate the high cost of pixel-level annotations (Chen et al., 2022b). Existing methods consider retinal vascular segmentation as an independent medical image analysis task and treat model design and downstream tasks separately. In contrast, our work treats OCTA vascular segmentation as an auxiliary task to enhance the quality of vascular regions in the translated OCTA images, rather than solely solving the vascular segmentation problem. This approach provides a novel perspective and demonstrates the potential of OCTA vascular segmentation as an auxiliary task.

2.4. Multi-task learning

Multi-task learning is a widely used approach that aims to optimize a machine learning model for multiple tasks in a unified framework (Caruana, 1997). It employs shared feature representations to achieve balanced optimization for all the tasks (Li et al., 2021). Despite differences in the requirements for image features, combining adjacent tasks such as object detection and instance segmentation has shown remarkable empirical results (Zhang et al., 2020, 2021a). In cases where tasks have unequal importance, auxiliary learning, where auxiliary tasks are used to assist the main tasks in achieving better predictions, can be employed (Kendall et al., 2018; Liu et al., 2019). Auxiliary tasks are frequently used in generative adversarial networks, such as classification and jigsaw solving (Liebel and Körner, 2018; Odena et al., 2017). In our method, we use two auxiliary tasks, vascular segmentation, and 2D OCTA projection map translation, to help the main task, OCT to OCTA image translation, learn rich feature representations. We integrate all tasks into the same framework and train them in an end-to-end fashion. Our contribution is using two auxiliary tasks to improve the performance of the main task. The coherence between the auxiliary tasks and the main task in terms of features enhances the performance of OCT to OCTA image translation.

3. Methodology

As illustrated in Fig. 2(b), the proposed TransPro mainly contains the following three components: (1) a 3D conditional generative adversarial network in Section 3.1, which is the main body of our method, (2) a Heuristic Contextual Guidance (HCG) module in Section 3.2 to align the translation of projection maps. (3) a Vessel Promoted Guidance (VPG) module in Section 3.3 to improve the quality of vessel areas. We summarize the entire TransPro framework in Section 3.4.

3.1. 3D generative adversarial network

OCT to OCTA image translation task aims to translate an OCT volume $\mathbf{X} \in \mathbb{R}^{L \times W \times D}$ to its paired OCTA volume $\hat{\mathbf{Y}} \in \mathbb{R}^{L \times W \times D}$, where the translated OCTA volume $\hat{\mathbf{Y}}$ needs to be as close as possible to the ground-truth OCTA volume $\mathbf{Y} \in \mathbb{R}^{L \times W \times D}$ (Yang et al., 2023). In particular, each volume data is composed of a set of 2D slices with the size of $L \times D$ (i.e., B-scan images) and the number of slices for each volume is W.
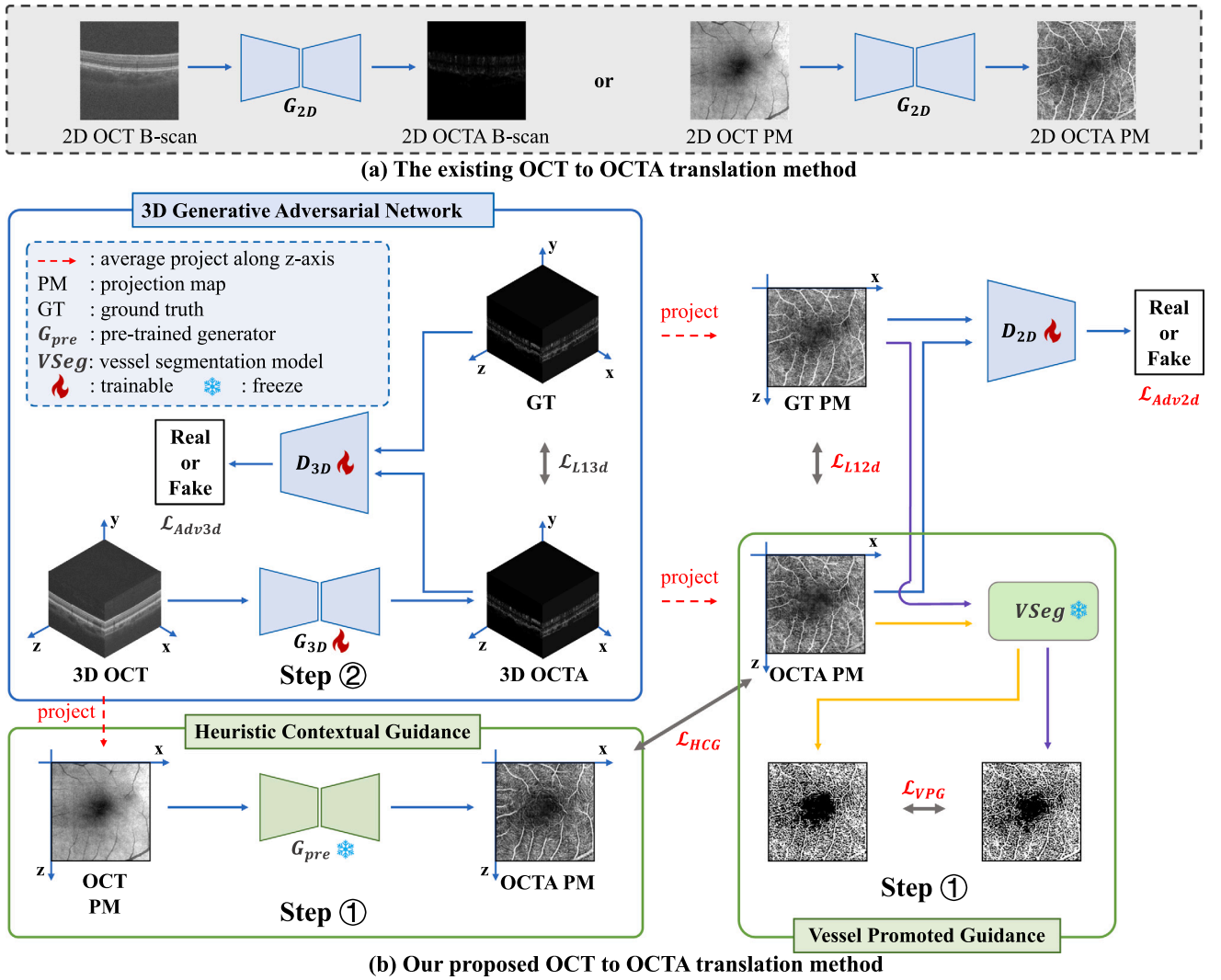
(a) The existing OCT to OCTA translation method

(b) Our proposed OCT to OCTA translation method

Fig. 2. The main ideas and contributions of this work. (a) illustrates the overall framework of the existing OCT to OCTA translation methods. (b) illustrates our proposed TransPro. Compared to existing methods, TransPro not only takes into considerations of critical vessel information via the Vessel-Promoted Guidance model but also introduces beneficial contexts via the Heuristic Contextual Guidance model.

Our proposed framework builds upon the classical image-to-image translation model, pix2pix (Isola et al., 2017). Unlike existing methods that only consider the OCT B-scan image as input to the generator to produce the corresponding OCTA B-scan image (Roisman and Goldhardt, 2017; Eladawi et al., 2018), we extend the network to three-dimensional convolutional neural networks to handle 3D volumes and enhance the global contextual learning across slices. Specifically, the input to the 3D generator is a 3D OCT volume and the output is a translated OCTA volume. Additionally, a 3D discriminator is used to differentiate between ground-truth and translated OCTA volumes. The 3D generator and discriminator are denoted as G_{3d} and D_{3d}, respectively, and the adversarial loss is formulated accordingly:

$$\mathcal{L}_{Adv3d} = \mathbb{E}_{\mathbf{Y}\sim p(OCTA)}[\log(D_{3d}(\mathbf{Y}))] + \mathbb{E}_{\mathbf{X}\sim p(OCT)}[\log(1 - (D_{3d}(G_{3d}(\mathbf{X}))))], \tag{1}$$

where $\mathbf{Y}$ denotes the ground-truth OCTA volume sampled from distribution $p(OCTA)$ and $\mathbf{X}$ denotes the input OCT volume sampled from distribution $p(OCT)$. $G_{3d}(\mathbf{X})$ which is equal to $\hat{\mathbf{Y}}$, denotes the output OCTA volume translated by G_{3d} network. In addition, there is a constraint on each voxel between the translated volume $G_{3d}(\mathbf{X})$ and the ground-truth volume $\mathbf{Y}$. To ensure less blurring, we follow Isola et al. (2017) and apply $L1$ distance instead of $L2$, which is formulated as:

$$\mathcal{L}_{L13d} = \left\| \mathbf{Y} - G_{3d}(\mathbf{X}) \right\|_1. \tag{2}$$

In addition to the 3D volume $(L \times W \times D)$, there is another type of data format, the 2D projection map $(L \times W)$ data that is capable of visualizing vascular structures. The projection maps are obtained by averaging the values of pixels along the D dimension. We denote the projection process as $Proj(\cdot)$, and the obtained OCTA projection map as $\hat{\mathbf{y}} \in \mathbb{R}^{L \times W}$, then we have:

$$\hat{\mathbf{y}} = Proj(\hat{\mathbf{Y}}). \tag{3}$$

Based on experimental observations, it has been noted that the constraints in Eqs. (1) and (2) may not be sufficient to ensure high quality in the translated OCTA projection maps. The reason behind this is the accumulation of errors during the computation of projection maps of the translated OCTA volumes with Eq. (3). Therefore, to enhance the similarity between the ground-truth and translated OCTA projection maps, we propose the use of a 2D adversarial loss (referred to as $\mathcal{L}_{Adv2d}$ in Eq. (4)) and L1 loss (referred to as $\mathcal{L}_{L12d}$ in Eq. (5)). To achieve this, we apply the projection function $Proj(\cdot)$ on the ground-truth OCTA volume $\mathbf{Y}$ and the translated OCTA volume $\hat{\mathbf{Y}}$ to obtain the corresponding ground-truth and translated projection maps, denoted as $\mathbf{y}$ and $\hat{\mathbf{y}}$, respectively. The losses are then computed as follows:

$$\mathcal{L}_{Adv2d} = \mathbb{E}_{\mathbf{Y}\sim p(OCTA)}[\log(D_{2d}(\mathbf{y}))] + \mathbb{E}_{\mathbf{X}\sim p(OCT)}[\log(1 - (D_{2d}(\hat{\mathbf{y}})))], \tag{4}$$

$$\mathcal{L}_{L12d}(G_{3d}) \;=\; \parallel \mathbf{y} - \hat{\mathbf{y}} \parallel_1, \tag{5}$$

where

$$\mathbf{y} = Proj(\mathbf{Y}),$$
$$\hat{\mathbf{y}} = Proj(\hat{\mathbf{Y}}) = Proj(G_{3d}(\mathbf{X})). \tag{6}$$

D_{2d} denotes a 2D discriminator which differentiates the ground-truth and translated OCTA projection maps.

The overall loss function for the 3D generative adversarial network is expressed as:

$$\mathcal{L}_{3DGAN} = \mathcal{L}_{Adv3d} + \mathcal{L}_{Adv2d} + \lambda_1 \mathcal{L}_{L13d} + \lambda_2 \mathcal{L}_{L12d}, \tag{7}$$

where λ_1 and λ_2 are loss-balance hyperparameters which are set to 10 in our experiments. We set them to an equal value to keep the hyperparameter search computationally feasible.

3.2. Heuristic Contextual Guidance (HCG)

The loss function described in Eq. (5) enforces the alignment between the translated and the real OCTA images in the view of projection maps. However, we observe that there exist some vessel discontinuity regions in real OCTA projection maps due to the unstable scanning by OCTA devices. Therefore, the model may suffer the overfitting problem by learning these specific patterns. To alleviate this issue, we introduce the Heuristic Contextual Guidance (HCG) module, which is a pre-trained OCT to OCTA projection map translation model. By leveraging the convolution operations applied to 2D OCTA projection maps in the HCG module, each output pixel contains information from its neighboring pixels. This spatial context plays a crucial role in improving vessel continuity in 2D OCTA projection maps.

Specifically, the pre-trained model's generator (G_{pre}) is utilized and its parameters are fixed during training. The projection map of the input OCT volume is then subjected to G_{pre}, resulting in an OCTA projection map that contains independent contextual information from the projection view. The output of the 2D translation model, G_{pre}, is denoted as $\mathbf{y}'$. The $L1$ distance is utilized to determine the dissimilarity between $\mathbf{y}'$ and the projection map of translated OCTA, $G_{3d}(\mathbf{X})$, which is produced by G_{3d}, which can be formulated as:

$$\mathcal{L}_{HCG}(G_{3d}) \;=\; \parallel \mathbf{y}' - Proj(G_{3d}(\mathbf{X})) \parallel_1, \tag{8}$$

where

$$\mathbf{y}' = G_{pre}(Proj(\mathbf{X})). \tag{9}$$

3.3. Vessel Promoted Guidance (VPG)

Given that OCTA images are primarily used to reflect vessel signals, the similarity of blood vessels between synthesized and ground-truth data is more crucial than the similarity of background tissues. However, it is challenging for the network to focus on vessel pixels in the absence of specific flow information (Ferrara et al., 2016; Kashani et al., 2017). Thus, we introduce a vessel segmentation model with fixed parameters to guide the generation accuracy of vessel regions.

Moreover, during the 3D volume translation process, the lack of vessel information leads to aimlessness, resulting in poor quality of vascular regions in translated OCTA projection maps. To this end, we propose a vascular segmentation model to focus more on the vascular regions. Initially, we pre-train a vascular segmentation model $VSeg(\cdot)$ on OCTA projection maps with annotated vessel labels. Next, we use $VSeg(\cdot)$ to extract semantic vascular segmentation logits on both ground-truth and translated OCTA projection maps, denoted as $\mathbf{l}_{seg}$ and $\hat{\mathbf{l}}_{seg}$, respectively. Finally, we employ L1 distance to minimize the discrepancy between the two logits:

$$\mathcal{L}_{VPG} \;=\; \left\| \hat{\mathbf{l}}_{seg} - \mathbf{l}_{seg} \right\|_1, \tag{10}$$

where

$$\hat{\mathbf{l}}_{seg} = VSeg(Proj(G_{3d}(\mathbf{X}))),$$
$$\mathbf{l}_{seg} = VSeg(Proj(\mathbf{Y})). \tag{11}$$

The $\mathcal{L}_{VPG}$ loss (Eq. (10)) aims to match the vascular structures in the ground-truth and translated OCTA projection maps, addressing the problem of aimlessness in vascular areas.

3.4. Overall framework

The overall framework of our proposed TransPro method consists of three parts, which are 3D GAN, the HCG module, and the VPG module. We pre-train the models for VPG and HCG modules and fix their parameters when training the 3D GAN.

During training, we jointly train the generator G_{3d} and two discriminators D_{3d} and D_{2d} following the generative-adversarial learning pattern (Isola et al., 2017; Zhu et al., 2017). Our total loss function can be formulated as:

$$\mathcal{L}_{TransPro} = \mathcal{L}_{3DGAN} + \alpha \mathcal{L}_{VPG} + \beta \mathcal{L}_{HCG}, \tag{12}$$

where α and β are task-balance hyperparameters which are set to 5 in our experiments. We set the weights to an equal value to keep the hyperparameter search computationally feasible. The final objective function is expressed as:

$$G_{3d}^* = \arg \min_{G_{3d}} \max_{D_{3d}, D_{2d}} \mathcal{L}_{TransPro}(G_{3d}, D_{3d}, D_{2d}). \tag{13}$$

In inference, only the 3D generator G_{3d} is used. Therefore, TransPro does not require any additional computational overheads in this process.

4. Experiments

4.1. Dataset (OCTA-500)

OCTA-500 (Li et al., 2020a) is a publicly accessible dataset with 500 pairs of 3D OCT and OCTA volumes. This dataset is divided into two subsets according to the field of view types, which are OCTA-3M and OCTA-6M, respectively. We implement our model on these two subsets.

- **OCTA-3M** contains 200 pairs of OCT and OCTA volumes with 3 mm × 3 mm × 2 mm field of view. Each volume is with the size of 304px × 304px × 640px, and the size of the projection map is 304px × 304px. Following Li et al. (2020a), we divide the size of the training set, validation set, and test set as 140 volumes, 10 volumes, and 50 volumes, respectively.
- **OCTA-6M** contains 300 pairs of OCT and OCTA volumes with 6 mm × 6 mm × 2 mm field of view. The volume is with the size of 400px × 400px × 640px, and the size of its projection map is 400px × 400px. The size of the training set, validation set, and test set for the OCTA-6M dataset is 180 volumes, 20 volumes, and 100 volumes, respectively.

4.2. Evaluation metrics

We evaluate the translated OCTA images from the following two aspects: OCTA volume and OCTA projection map. To evaluate the quality of OCTA volume, we use standard image quality assessment metrics, *i.e.*, Mean Absolute Error (**MAE**), Peak Signal-to-Noise Ratio (**PSNR**) (Oriani, 2022), and Structural SIMilarity (**SSIM**) (Wang et al., 2004) and average the results over each slide of the whole volume (B-scan image).

In addition to the standard evaluation metrics used to assess the quality of OCTA volume, we implement five supplementary evaluation metrics specifically designed for the OCTA projection maps. These metrics serve to quantify the similarity of vessel topology and morphology
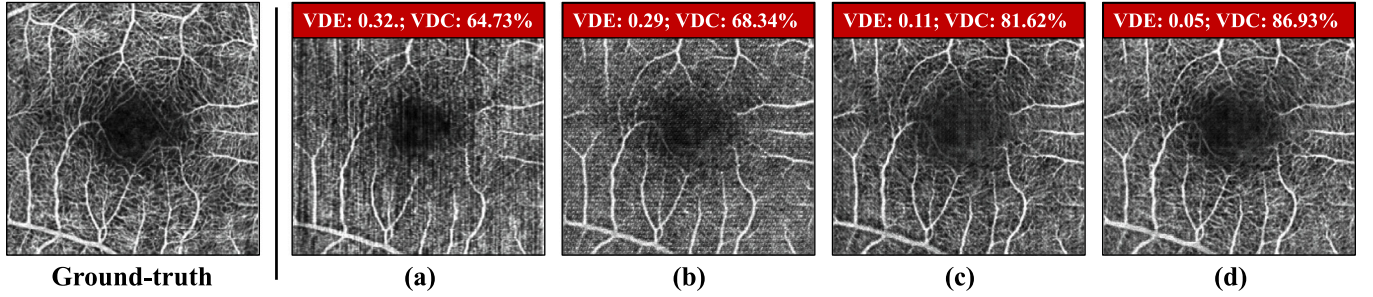
Fig. 3. The OCTA projection maps with different vessel density error (VDE) and vessel density correlation (VDC) results. The leftmost side is the ground-truth and (a)–(d) are the four translated OCTA projection maps generated by different OCT to OCTA translation methods with different qualities from low to high, namely Adjacent GAN (Li et al., 2020b), Pix2pix 3D (Isola et al., 2017), TransPro (not the best epoch), and TransPro (the best epoch), respectively. The results of VDE and VDC metrics are matched with the visualized quality.

structures in OCTA projection maps, which are routinely examined by doctors for disease diagnosis.

The first three metrics are modified from the MAE, PSNR, and SSIM metrics and are now referred to as vessel-weighted metrics. We denote these metrics as **MAE-V**, **PSNR-V**, and **SSIM-V**. We annotate binary pixel-wise vessel segmentation mask for each ground-truth OCTA projection map, denoted as $\hat{\mathbf{M}}$, that 0 represents non-vessel pixels and 1 represents vessel pixels. We preprocess the translated and the ground-truth OCTA projection maps to decrease the weights of non-vessel pixels. The values of non-vessel pixels are multiplied by a ratio $\gamma \in (0, 1.0]$, while the vessel pixels values remain unchanged. We denote this preprocessing process as $\mathbf{V}(\cdot)$ that is defined as:

$$\mathbf{V}(\mathbf{I}(i,j)) = \begin{cases} \gamma * \mathbf{I}(i,j), & if \ \hat{\mathbf{M}}(i,j) = 0 \\ \mathbf{I}(i,j), & if \ \hat{\mathbf{M}}(i,j) = 1 \end{cases}, \tag{14}$$

where (i, j) denotes the position of pixels.

Then, the vessel-weighted metrics are calculated as:

$$\mathbf{MAE\text{-}V} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{MAE}(\mathbf{V}(\mathbf{I}), \mathbf{V}(\hat{\mathbf{I}})), \tag{15}$$

$$\mathbf{PSNR\text{-}V} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{PSNR}(\mathbf{V}(\mathbf{I}), \mathbf{V}(\hat{\mathbf{I}})), \tag{16}$$

$$\mathbf{SSIM\text{-}V} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{SSIM}(\mathbf{V}(\mathbf{I}), \mathbf{V}(\hat{\mathbf{I}})), \tag{17}$$

where $\mathbf{I}$ represents the translated OCTA projection maps and $\hat{\mathbf{I}}$ denotes the ground-truth. N is the size of test set.

We set the value of γ to 0.1 for the results presented in Table 2. Additionally, we provide results for the MAE-V, PSNR-V, and SSIM-V metrics for other values of γ in Section 4.5.2.

In addition, we propose two novel metrics based on vessel density (Yao et al., 2020), *i.e.*, **V**essel **D**ensity **E**rror (**VDE**) and **V**essel **D**ensity **C**orrelation (**VDC**). Vessel density is a crucial clinical measure used to detect vessel abnormalities in various retinal diseases, particularly in their early phases (Mastropasqua et al., 2017; Al-Sheikh et al., 2018; Richter et al., 2018). When comparing two OCTA projection maps, if they possess the same vessel density, they can be considered identical. The definition of vessel density is the proportion of the vessel area over the total area (Alam et al., 2017), which can be expressed as the percentage of vessel pixels over the total pixels within a given region. We adopt a traditional image processing approach as described in Levine et al. (2020) for obtaining vessel segmentation masks. This method involves binarizing the OCTA projection maps using a global mean threshold, where pixels above the threshold are interpreted as vessels. We denote the obtained binary masks as $\mathbf{M}$ and $\hat{\mathbf{M}}$, for translated and ground-truth OCTA projection maps, respectively. Then, the vessel density is calculated as the ratio of vessel area to the entire imaged area.

- **Vessel Density Error (VDE).** The VDE metric is utilized to assess the absolute difference of vessel density between translated and ground-truth OCTA projection maps. It is computed as

$$\mathbf{VDE} = \frac{1}{N}\sum_{i=1}^{N}|\mathbf{M}_{VD} - \hat{\mathbf{M}}_{VD}|, \tag{18}$$

where N is the size of test set and the subscript VD denotes the vessel density of the image. As shown in Fig. 3, lower VDE values indicate better quality of the translation outcomes.

- **Vessel Density Correlation (VDC).** VDC is employed to assess the similarity of vessel density in a local-wise manner. Specifically, we partition each OCTA projection map into small patches $\mathbf{m}^i$ of size 16px × 16px, compute the vessel density for each patch, and derive an array of regional vessel densities, denoted as $\mathbf{M}_{VD}^{arr}$.

$$\mathbf{M}_{VD}^{arr} = [\mathbf{m}_{VD}^1, \mathbf{m}_{VD}^2, \dots, \mathbf{m}_{VD}^k], \tag{19}$$

The VDC is calculated as the Pearson correlation coefficient of the two arrays between the translated and ground-truth OCTA projection maps.

$$\mathbf{VDC} = \frac{1}{N}\sum_{i=1}^{N}\frac{cov(\mathbf{M}_{VD}^{arr}, \hat{\mathbf{M}}_{VD}^{arr})}{\sigma_{\mathbf{M}_{VD}^{arr}}\sigma_{\hat{\mathbf{M}}_{VD}^{arr}}}, \tag{20}$$

where N is the size of test set, cov is the covariance, and σ is the standard deviation. Examples in Fig. 3 indicate that a higher VDC corresponds to a better translation quality.

4.3. Implementation details

Settings. The proposed TransPro model is trained in two steps. Firstly, we pre-train two models for VPG and HCG and freeze their parameters. Secondly, we train 3D GAN following the generative adversarial learning pattern with the total loss function (Eq. (12)). The backbones and the detailed training settings for each component are introduced below.

- **VPG.** We utilize the UNet (Ronneberger et al., 2015) as the underlying model for the vascular segmentation module in the VPG framework. This module aims to generate predictions of vascular segmentation in OCTA projection maps, using vessel pixel annotations as the ground-truth labels. Apart from the large vessel annotations provided in OCTA-500 dataset, we also annotate the pixel-wise labels for capillaries in OCTA projection maps. The annotations are obtained by two experts. The splitting of training, validation, and test sets is consistent with original 3D datasets. We apply common data augmentation methods such as random cropping and flipping. The model is trained with cross-entropy loss using RMSprop optimizer with a learning rate of 10^{-5} for 100 epochs. After training, we choose the model with the best result on the validation set.

Table 1

Result comparisons with the state-of-the-arts on OCTA volumes and OCTA projection maps of OCTA-3M dataset (Li et al., 2020a). The MAE, PSNR (dB), and SSIM (%) are computed on OCTA volumes, while the MAE-V, PSNR-V (dB), SSIM-V (%), VDE, and VDC (%) are computed on OCTA projection maps. Results of other methods are produced by our re-implementation. "↓" refers to the lower the better and "↑" refers to the higher the better.

Method	MAE ↓	PSNR(dB) ↑	SSIM(%) ↑	MAE-V ↓	PSNR-V(dB) ↑	SSIM-V(%) ↑	VDE↓	VDC(%)↑
Pix2pix 2D (Isola et al., 2017)	0.0968	29.59	84.59	0.0823	17.92	85.26	0.1764	62.49
Pix2pix 3D (Isola et al., 2017)	0.0883	31.58	86.24	0.0753	20.36	89.14	0.2206	65.29
9B18CN UNet (Lee et al., 2019)	0.0918	29.34	86.04	0.0839	17.12	85.03	0.1612	63.52
Adjacent GAN (Li et al., 2020b)	0.0906	30.89	87.14	0.0833	17.56	85.88	0.2034	63.05
VQ-I2I (Chen et al., 2022a)	0.0824	31.72	87.59	0.0760	18.61	88.90	0.1469	71.07
Palette (Saharia et al., 2022)	0.0814	32.42	**88.24**	0.0793	19.17	87.97	0.1466	70.86
TransPro (ours)	**0.0782**	**32.56**	88.22	**0.0658**	**20.42**	**91.79**	**0.1304**	**74.41**

Table 2

Result comparisons with the state-of-the-arts on OCTA volumes and OCTA projection maps of OCTA-6M dataset (Li et al., 2020a). The MAE, PSNR (dB), and SSIM (%) are computed on OCTA volumes, while the MAE-V, PSNR-V (dB), SSIM-V (%), VDE, and VDC (%) are computed on OCTA projection maps. Results of other methods are produced by our re-implementation. "↓" refers to the lower the better and "↑" refers to the higher the better.

Method	MAE↓	PSNR(dB)↑	SSIM(%)↑	MAE-V↓	PSNR-V(dB)↑	SSIM-V(%)↑	VDE↓	VDC(%)↑
Pix2pix 2D (Isola et al., 2017)	0.0995	27.65	87.15	0.0845	16.72	85.47	0.1884	63.33
Pix2pix 3D (Isola et al., 2017)	0.0900	**30.66**	87.16	0.0923	17.04	84.51	0.1904	67.68
9B18CN UNet (Lee et al., 2019)	0.1135	27.91	83.69	0.0895	16.89	84.07	0.2263	62.96
Adjacent GAN (Li et al., 2020b)	0.1021	28.05	85.03	0.0879	17.19	84.44	0.2052	65.70
VQ-I2I (Chen et al., 2022a)	0.0897	29.54	86.90	0.0788	17.38	86.78	0.1523	70.50
Palette (Saharia et al., 2022)	0.0881	30.02	87.13	0.0797	17.95	85.78	0.1575	70.11
TransPro (ours)	**0.0854**	30.53	**88.35**	**0.0733**	**18.30**	**90.23**	**0.1492**	**73.47**

Table 3

Ablation studies on the effectiveness of VPG and HCG modules on OCTA-3M and OCTA-6M (Li et al., 2020a). The MAE, PSNR (dB) and SSIM (%) are computed on OCTA volumes, while the VDE and VDC (%) are computed on OCTA projection maps. "↓" refers to the lower the better and "↑" refers to the higher the better.

3D GAN	VPG	HCG	MAE ↓	PSNR(dB) ↑	SSIM(%) ↑	VDE ↓	VDC(%) ↑
			OCTA-3M Dataset				
✓			0.0883	31.5824	86.24	0.2206	65.29
✓	✓		0.0807	32.4794	87.43	0.1791	70.06
✓		✓	0.0797	31.9813	87.78	0.1682	70.31
✓	✓	✓	**0.0782**	**32.5587**	**88.22**	**0.1304**	**74.41**
			OCTA-6M Dataset				
✓			0.0900	**30.6621**	87.16	0.1904	67.68
✓	✓		0.0879	30.1424	88.15	0.1739	71.04
✓		✓	0.0860	29.8427	88.24	0.1770	69.56
✓	✓	✓	**0.0854**	30.5264	**88.35**	**0.1492**	**73.47**

- **HCG.** We apply the pix2pix (Isola et al., 2017) as the backbone of the OCT to OCTA projection map translation model for HCG module. The projection maps of OCT and OCTA volumes are computed by the projection function $Proj(\cdot)$ in Eq. (3). The splitting of training, validation, and test sets is consistent with original 3D datasets. We follow the settings in Isola et al. (2017) and train the model over 200 epochs. After training, we select the model with the best result on the validation set.

- **TransPro.** Upon completion of training the VPG and HCG modules, we proceed with training the TransPro model. The 3D pix2pix model, a modified version of the model proposed in Isola et al. (2017), is used as the backbone of TransPro. In this setup, the 3D generator G_{3d} and two discriminators D_{3d} and D_{2d} are jointly trained following the adversarial-learning manner. We use Adam optimizer and set the initial learning rate to 0.0002. The model is trained with a batch size of 1 for 200 epochs. To prevent overfitting, we evaluate the validation set using MAE and select the best model with the minimum error value.

4.4. Comparisons with the state-of-the-art methods

We re-implement several state-of-the-art methods for OCT to OCTA image translation. The results of the comparisons of our proposed TransPro and other methods on OCTA-3M and OCTA-6M datasets are summarized in Table 1 and 2, respectively. For each OCTA subject, we present the results of three commonly used metrics, MAE, PSNR,

and SSIM, applied to the OCTA volume. Additionally, we introduce five novel vessel-specific metrics, namely MAE-V, PSNR-V, SSIM-V, Vessel Density Error (VDE), and Vessel Density Correlation (VDC), calculated on OCTA projection map. These proposed metrics provide a comprehensive assessment of both overall OCTA volume quality and vessel-specific characteristics, enabling a more detailed evaluation of the translated results.

4.4.1. Experimental results on OCTA-3M

In our comparative experiments, We evaluate our TransPro method against two baseline methods, two OCT to OCTA image translation methods, and two state-of-the-art image-to-image translation methods. The re-implemented results for OCTA-3M dataset are displayed in Table 1. The two baseline methods, namely pix2pix 2D and pix2pix 3D, are implemented following the network architectures described in Isola et al. (2017). Pix2pix 2D takes paired OCT and OCTA B-scan images as inputs and outputs, while pix2pix 3D operates on paired OCT and OCTA volumes. Additionally, we compare our method with Adjacent GAN (Li et al., 2020b) and 9B18CN UNet (Lee et al., 2019), which are specifically proposed for OCT to OCTA translation. Adjacent GAN (Li et al., 2020b) incorporates three adjacent OCT B-scan images and outputs the middle OCTA B-scan image. Furthermore, we include two methods designed for paired image-to-image translation. VQ-I2I (Chen et al., 2022a) leverages vector quantization techniques within a GAN framework, achieving state-of-the-art performance in paired image-to-image translation tasks. Palette (Saharia et al., 2022) builds upon
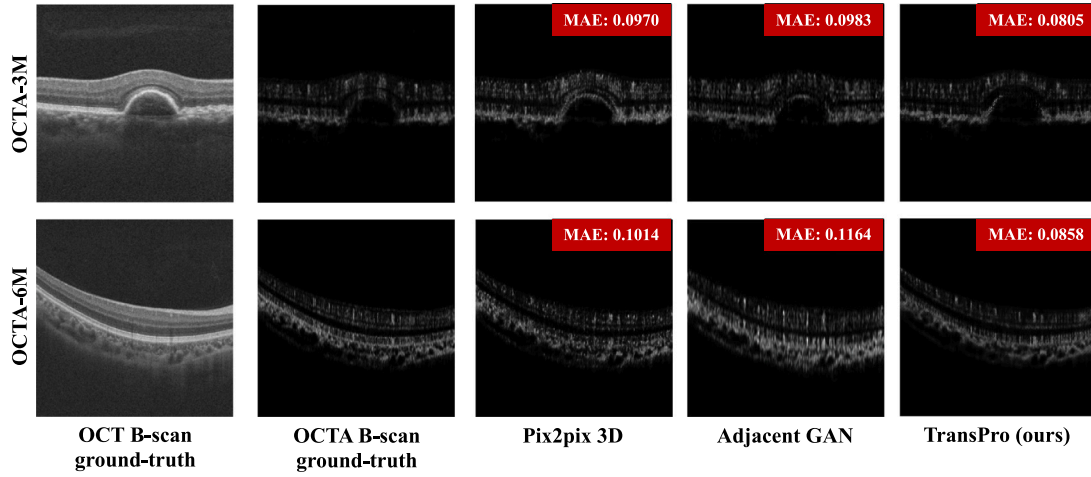
Fig. 4. The visualization examples of translated OCTA B-scans in OCTA-3M dataset and OCTA-6M dataset (Li et al., 2020a). We compared the results of different methods including pix2pix 3D, Adjacent GAN, and our proposed TransPro. The Mean Absolute Error (MAE) between the translated and the ground-truth OCTA B-scan images is computed.

conditional diffusion models, with the target domain images as conditions and noised source domain images as inputs. Among the six compared methods, pix2pix 3D is trained on 3D volumes, while the remaining methods are trained on 2D B-scan images.

From the results, we observe that pix2pix 3D generally outperformed other methods, indicating the significance of inter-slice connections in providing more global information compared to a single B-scan image. The latest approaches, VQ-I2I and Palette, achieve higher performances than previous methods. VQ-I2I benefits from its powerful backbone architectures, although it comes with a high computational cost when handling 3D input data. Palette utilizes diffusion models but had longer inference times. In contrast to these methods, our proposed TransPro method can handle 3D inputs, which leverages enhanced global contextual information. Meanwhile, it maintains negligible inference times compared to diffusion models. Moreover, TransPro incorporates vessel and contextual guidance from VPG and HCG modules, leading to significant improvements in the OCTA-3M dataset compared to other methods in both B-scan images and projection maps. Notably, in Table 1, our method demonstrates relative enhancements over the pix2pix 3D model by 40.8% in VDE and 14.0% in VDC metrics, respectively.

4.4.2. Experimental results on OCTA-6M

We further conduct the experiments of our proposed TransPro and six comparative methods on the OCTA-6M dataset, as shown in Table 2. It is observed that the quality of translated images in this dataset is generally lower compared to the OCTA-3M dataset. This discrepancy may be attributed to the lower resolution of images due to the larger field of view in the OCTA-6M dataset.

Regarding the results among the compared methods, the outcomes align with those obtained on the OCTA-3M dataset, where the performance is positively correlated with the utilization of global information by the models. Our proposed TransPro method achieves the best overall performance among all the compared methods, except for a slightly lower PSNR value compared to the Pix2pix 3D model. Notably, the remarkable improvements in the VDE and VDC metrics demonstrate the effectiveness of our method in enhancing the quality of vascular structures. These results highlight the strengths of TransPro in leveraging global contextual information and effectively enhancing the quality of vascular structures in OCTA images, even in challenging datasets such as OCTA-6M. Despite the lower resolution and potentially more complex images in this dataset, our method consistently outperforms the other compared methods, showcasing its effectiveness in improving the translation quality of OCT to OCTA images.

4.4.3. Visualization results

To evaluate the results qualitatively, we provide some visualization examples of B-scan images and projection maps translated by different methods. For the B-scan images, we select two examples from OCTA-3M and OCTA-6M datasets and compare the translated results of Pix2pix 3D, Adjacent GAN, and our TransPro method in Fig. 4. All methods are capable of learning the general outline and the texture of OCTA B-scans, albeit with variations in some details. The results obtained from Pix2pix 3D appear blurrier and exhibit discontinuities in certain areas, while the Adjacent GAN tends to predict the B-scan images with more white tissues, which may obscure the essential vessels. To further demonstrate the performance qualitatively, we compute the MAE for each image denoted in red boxes. Overall, our proposed TransPro method achieves the best translation results of OCTA B-scans in both qualitative and quantitative evaluation. For the projection maps, we choose four examples from OCTA-3M and OCTA-6M datasets respectively and compare the results of our proposed TransPro with Pix2pix 3D and Adjacent GAN methods. From the zoomed-out figures in Fig. 5, we find that the Pix2pix 3D model generates some artifacts to mimic the texture of vessels while the outputs of the Adjacent GAN model suffer from white stripes due to the limited intra-slice information of the 2D model. Compared to the ambiguous images translated from other methods, our method produces high-quality OCTA projection images with vivid vessels and background tissues.

4.5. Ablation study

4.5.1. Effectiveness of VPG and HCG

To validate the respective effectiveness of VPG and HCG components, we conduct additional experiments by separately adding these two modules to the 3D GAN baseline model. Table 3 displays the results on OCTA-3M and OCTA-6M datasets. In general, we can find that VPG and HCG modules bring improvement to the quality of B-scan images as the metrics of MAE, PSNR and SSIM evaluating OCTA B-scan images become better after adding corresponding components. For the quality of projection maps of translated OCTA images, the vessel density error (VDE) and vessel density correlation (VDC) metrics indicate the effectiveness of VPG and HCG modules on vascular areas. In particular, the VPG module shows significant improvement on VDC metric because it provides the semantic vascular segmentation map which increases the similarity of overall vascular structures. The HCG module acts as a regularizer, effectively addressing the overfitting issue caused by specific patterns found in real OCTA projection maps, such as the vessel discontinuity illustrated in Fig. 8. By employing the convolution operations on 2D OCTA projection maps in the HCG
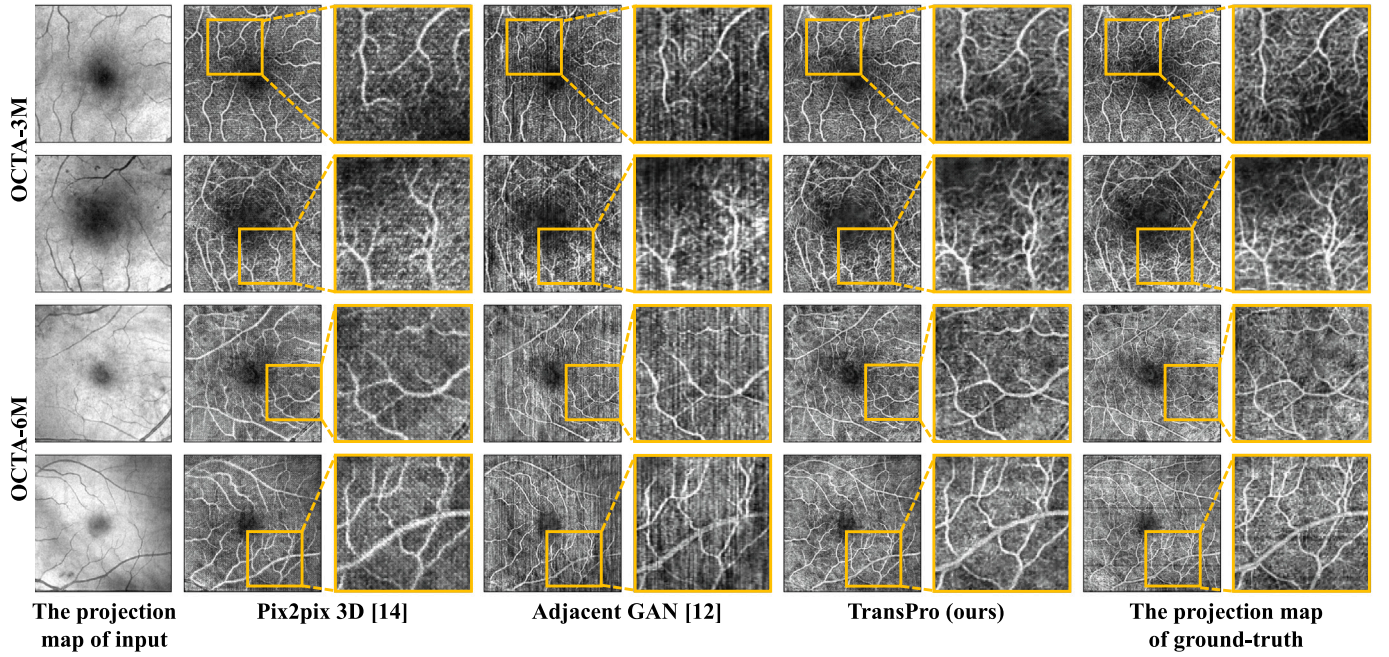
Fig. 5. Visualization result comparisons of translated OCTA projection maps in OCTA-3M dataset and OCTA-6M (Li et al., 2020a). Two examples are selected from each dataset correspondingly. We can observe that our approach can achieve better predictions.
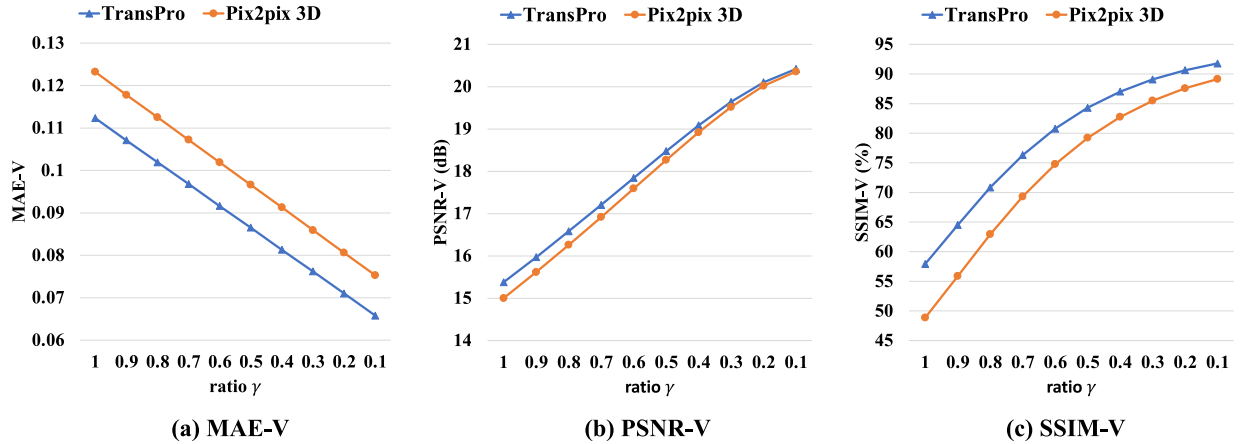

Fig. 6. Evaluation of translated vascular quality on OCTA projection in OCTA-3M (Li et al., 2020a). The non-vascular pixels are multiplied by a ratio γ from 1.0 to 0.1. The three evaluation metrics, *i.e.*, (a) MAE-V, (b) PSNR-V and (c) SSIM-V, are computed on processed OCTA projection maps. The *x*-axis refers to the coefficient multiplied to non-vascular pixels, and the *y*-axis refers to the results of corresponding metrics.

module, each output pixel incorporates information from nearby pixels. This integration of contextual information significantly enhances the model's overall performance and leads to a notable reduction in the VDE metric, particularly when applied to the OCTA-3M dataset. Moreover, when the VPG and HCG modules are applied simultaneously on the baseline model, *i.e.*, our proposed TransPro framework, the performance obtains prominent gains compared to the results with either VPG or HCG module. Therefore, we conclude that VPG and HCG modules offer their own effectiveness on the OCT to OCTA image translation task and their effects are additive when the two modules are active in TransPro method.

4.5.2. Evaluation of translated vascular quality

To evaluate the quality of translated vessels in OCTA projection maps, we further calculate vessel-weighted MAE, PSNR, and SSIM metrics by adjusting the ratio of pixel weights between vessel pixels and non-vessel pixels according to the vessel segmentation mask. We summarize the results in line charts for different ratios in OCTA-3M and

OCTA-6M datasets, as shown in Figs. 6 and 7, respectively. Specifically, we first obtain the pixel-wise vascular labels of OCTA projection maps in the test set which are annotated by ophthalmologists. Then, for the non-vessel pixels in the segmentation mask, we multiply the OCTA projection maps by a ratio $\gamma \in \{1.0, 0.9, \ldots, 0.1\}$ to gradually decrease the weights of non-vessel pixels from 1.0 to 0.1. The weights of vessel pixels remain at 1.0. We apply the weights adjusting for all translated OCTA projection maps and the ground-truth images and compute MAE-V, PSNR-V, and SSIM-V metrics between two sets of projection maps. In consequence, as the weights of vascular pixels increase, MAE-V drops linearly and the metric of PSNR-V and SSIM-V gradually improve for both OCTA-3M and OCTA-6M datasets. In particular, the PSNR-V and SSIM-V results of TransPro method can achieve over 20 dB and 90% for the OCTA-3M dataset and over 17 dB and 90% for the OCTA-6M dataset, respectively, when the weights of non-vessel pixels reduce to 0.1. In addition, by comparing our method to Pix2pix 3D method on both two datasets, the results indicate that our method can achieve
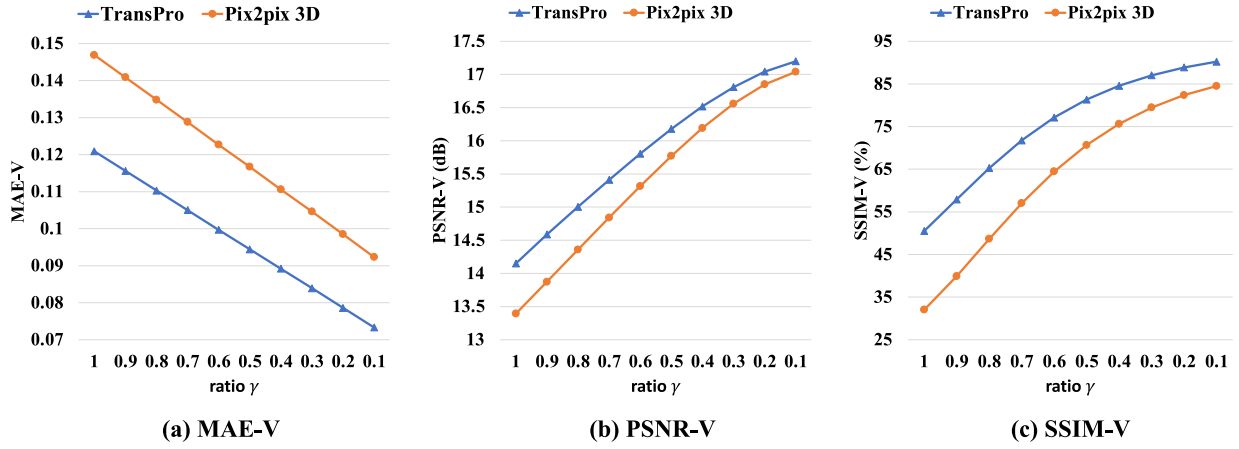
Fig. 7. Evaluation of translated vascular quality on OCTA projection in OCTA-6M (Li et al., 2020a). The non-vascular pixels are multiplied by a ratio γ from 1.0 to 0.1. The three evaluation metrics, *i.e.*, (a) MAE-V, (b) PSNR-V and (c) SSIM-V, are computed on processed OCTA projection maps. The *x*-axis refers to the coefficient multiplied to non-vascular pixels, and the *y*-axis refers to the results of corresponding metrics.

Table 4
Ablation studies on different losses of VPG and HCG modules on OCTA-3M dataset (Li et al., 2020a). The MAE, PSNR (dB), and SSIM (%) are computed on OCTA volumes, while the MAE-V, PSNR-V (dB), SSIM-V (%), VDE, and VDC (%) are computed on OCTA projection maps. "↓" refers to the lower the better and "↑" refers to the higher the better.

Module	Loss type	MAE ↓	PSNR(dB) ↑	SSIM(%) ↑	MAE-V↓	PSNR-V(dB)↑	SSIM-V(%)↑	VDE↓	VDC(%)↑
VPG	MSE	0.0819	31.78	86.99	0.0734	20.29	89.55	0.1610	69.29
	BCE	0.0794	**33.38**	87.68	0.0669	20.35	91.25	0.1673	68.84
	L1	**0.0782**	32.56	**88.22**	**0.0658**	**20.42**	**91.97**	**0.1304**	**74.41**
HCG	MSE	0.0798	32.14	87.53	0.0702	20.10	90.49	0.1452	72.12
	SSIM	0.0846	31.62	87.18	0.0797	19.31	87.61	0.1352	73.78
	L1	**0.0782**	32.56	**88.22**	**0.0658**	**20.42**	**91.97**	**0.1304**	**74.41**

Table 5
Ablation studies on task-balanced hyperparameters α and β on OCTA-3M dataset (Li et al., 2020a). The MAE, PSNR (dB), and SSIM (%) are computed on OCTA volumes, while the MAE-V, PSNR-V (dB), SSIM-V (%), VDE, and VDC (%) are computed on OCTA projection maps. "↓" refers to the lower the better and "↑" refers to the higher the better.

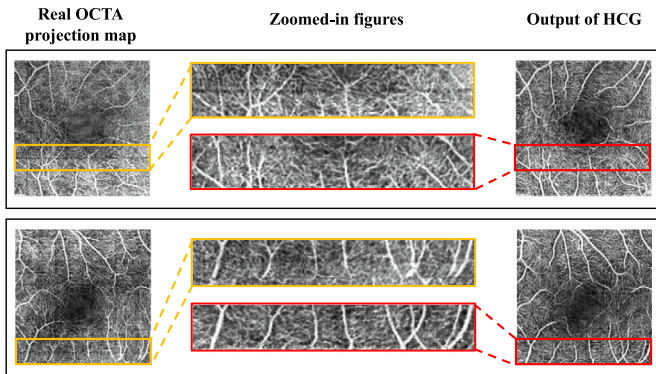| Hyperparameters	MAE ↓	PSNR(dB) ↑	SSIM(%) ↑	MAE-V↓	PSNR-V(dB)↑	SSIM-V(%)↑	VDE↓	VDC(%)↑
α, $\beta = 1$	0.0815	32.32	88.22	0.0716	20.30	90.30	0.1450	71.71
α, $\beta = 3$	0.0839	31.43	87.84	0.0762	18.05	88.85	0.1360	73.27
α, $\beta = 5$	**0.0782**	**32.56**	**88.22**	**0.0658**	**20.42**	**91.97**	**0.1304**	**74.41**
α, $\beta = 7$	0.0813	32.27	87.40	0.0692	20.20	91.10	0.1398	72.38
α, $\beta = 9$	0.0831	31.59	87.08	0.0727	19.75	89.83	0.1418	72.06

Fig. 8. Motivation of our proposed Heuristic Contextual Guidance (HCG) module. We compare the real OCTA projection maps and the outputs of the HCG model. Two examples show the vessel discontinuity in the real OCTA projection maps while the HCG output images recover the connection.

higher quality in both non-vessel regions (when the ratio is high) and vessel regions (when the ratio is low) than pix2pix 3D model.

4.5.3. Types of losses in VPG and HCG modules

To validate the suitability of L1 loss for the VPG (Eq. (10)) and HCG (Eq. (8)) modules, we conduct experiments to compare it with alternative loss functions. The translation results of L1 loss outperform other types of loss, as shown in Table 4.

For the VPG module, we employ Mean Squared Error (MSE) loss and Binary Cross Entropy (BCE) loss to enforce consistency between the segmentation predictions of the translated OCTA projection map and the ground-truth projection map. Since the output of the vessel segmentation model is normalized between 0 and 1, the MSE loss yields smaller values than the L1 loss, resulting in a weaker constraint and poorer translation results. The BCE loss demonstrates comparable performance with L1 loss, achieving a higher PSNR value for OCTA volumes while slightly reducing other metrics.

For the HPG module, we utilize MSE loss and Structure Similarity (SSIM) loss to minimize differences between two OCTA projection maps. The loss function serves as an image reconstruction metric by measuring the dissimilarities between two images. From the results, we find that MSE loss exhibits slightly inferior performance to L1 loss because it leads to blurred translated images. SSIM loss is defined to measure image similarity in terms of luminance, contrast, and structure. However, due to the noise and ambiguity inherent in OCTA projection maps, SSIM loss faces challenges in accurately reflecting image similarity.
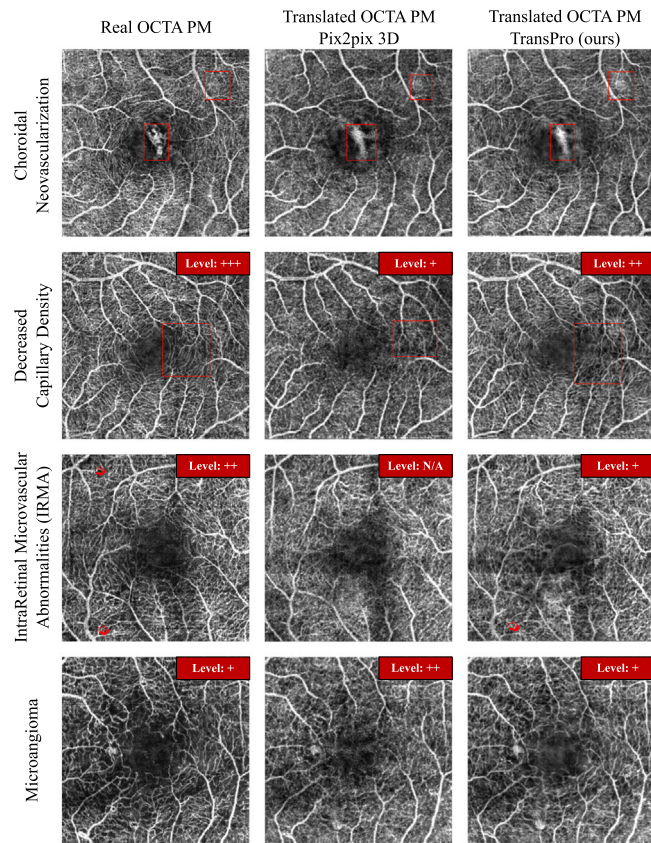
Fig. 9. The diseased patterns of translated OCTA projection maps annotated by an experienced ophthalmologist. Four different patterns are displayed, including choroidal neovascularization, decreased capillary density, intraretinal microvascular abnormalities (IRMA), and microangioma. Each pattern is presented alongside the corresponding real OCTA image, the translation results from the Pix2pix 3D method, and our TransPro method. Levels of disease patterns are annotated from "+" to "+++", where the number of "+" denotes the degree of severity.

4.5.4. Hyperparameters

We complete a series of additional experiments to evaluate the impact of the task-balance hyperparameters α and β in Eq. (12) for OCTA translation. The VPG and HCG losses exhibit comparable magnitudes during training. To maintain computational feasibility of the hyperparameter search, we assign equal weights to these losses. We evaluate the performance of our method on the OCTA-3M dataset and analyze the results, which are presented in Table 5. The results demonstrate that our proposed method consistently enhances the overall performance across a wide range of α and β values. We identify that the best performance is achieved when α and β are set to 5, indicating the importance of selecting an appropriate value to balance the main task and two auxiliary tasks.

4.5.5. Abnormal pattern identification and disease recognition

In addition to the quantitative and qualitative analysis of the translated OCTA images, we cooperate with ophthalmologists to analyze the abnormal patterns and diseases in the translated OCTA images, such that the potential practical applications of deep learning-based OCT to OCTA image translation method are evaluated. We conduct two tasks, which are abnormal pattern identification and disease recognition.

For the abnormal pattern identification task, we collaborate with an experienced ophthalmologist who examined both real and translated OCTA projection maps. We summarize four types of abnormal patterns including choroidal neovascularization, decreased capillary density, intraretinal microvascular abnormalities (IRMA), and microangioma,

as shown in Fig. 9. For each type of pattern, we compare the real OCTA images, the translation results from Pix2pix 3D method, and our TransPro method. Our findings reveal that major diseased patterns are preserved in the translated OCTA projection maps, albeit with some ambiguity in the details of fine vessel areas. Moreover, compared to the Pix2pix 3D method, our TransPro generates OCTA projection maps that are more similar to the real ones in terms of diseased levels and locations, as annotated in Fig. 9.

For the disease recognition task, we conduct an evaluation where an experienced ophthalmologist classifies diabetic retinopathy (DR) and choroidal neovascularization (CNV) images among 50 samples from both real and translated OCTA projection maps. The disease recognition results of DR and CNV on the real and translated images are identical, achieving a 100% accuracy rate. Although this experiment encompasses a limited number of disease types and samples, the high consistency and accuracy of the OCTA projection maps translated by our proposed TransPro method highlight their quality and potential for practical applications in OCTA translation. However, although the generated OCTA projection maps provide valuable insights for disease diagnosis, they do not faithfully reconstruct the finer vascular details. Therefore, our proposed TransPro method cannot completely replace the necessity for precise OCTA scanning. It is potential that future developments in deep-learning-based OCT to OCTA image translation algorithms will yield more accurate and higher fidelity results.

5. Conclusion

In this paper, a novel framework for 3D OCT to OCTA image translation, termed TransPro, is proposed. TransPro is mainly driven by two key insights: leveraging two different input views and improving vessel area quality in the translated OCTA images. The first insight is inspired by a crucial observation that the OCTA projection map is generated by averaging pixel values from its corresponding B-scans along the Z-axis. As a result, we propose a hybrid generative architecture with a novel Heuristic Contextual Guidance (HCG) module, that ensures the consistency of the translated OCTA images in both B-scan and projection views. To achieve the second insight, we propose a novel Vessel Promoted Guidance (VPG) module, which enhances the attention on retinal vessels. Experimental results on public datasets demonstrate that TransPro achieves outstanding performance. TransPro is a versatile 3D image generation model that can be applied to a wide range of computer vision tasks, such as data augmentation, semantic segmentation, and domain adaptation. Future work will explore the use of TransPro in additional image generation applications.

CRediT authorship contribution statement

Shuhan Li: Writing – review & editing, Writing – original draft, Methodology. **Dong Zhang:** Writing – review & editing, Writing – original draft. **Xiaomeng Li:** Writing – review & editing, Supervision. **Chubin Ou:** Data curation, Conceptualization. **Lin An:** Conceptualization. **Yanwu Xu:** Project administration, Investigation. **Weihua Yang:** Data curation. **Yanchun Zhang:** Validation. **Kwang-Ting Cheng:** Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Al-Sheikh, M., Iafe, N.A., Phasukkijwatana, N., Sadda, S.R., Sarraf, D., 2018. Biomarkers of neovascular activity in age-related macular degeneration using optical coherence tomography angiography. Retina 38 (2), 220–230.

Alam, M., Thapa, D., Lim, J.I., Cao, D., Yao, X., 2017. Computer-aided classification of sickle cell retinopathy using quantitative features in optical coherence tomography angiography. Biomed. Opt. Express 8 (9), 4206–4216.

Caruana, R., 1997. Multitask learning. Mach. Learn. 28 (1), 41–75.

Chen, Y.-J., Cheng, S.-I., Chiu, W.-C., Tseng, H.-Y., Lee, H.-Y., 2022a. Vector quantized image-to-image translation. In: European Conference on Computer Vision. ECCV.

Chen, C.-X., Liu, M.-L., Cao, K., Yusufu, M., Wang, J.-D., 2021. Diagnostic value of optical coherence tomography angiography for choroidal neovascularization in age-related macular degeneration: a systematic review and meta-analysis. Ophthalmic Res. 64 (5), 704–712.

Chen, Z., Xiong, Y., Wei, H., Zhao, R., Duan, X., Shen, H., 2022b. Dual-consistency semi-supervision combined with self-supervision for vessel segmentation in retinal OCTA images. Biomed. Opt. Express 13 (5), 2824–2834.

Eladawi, N., Elmogy, M., Khalifa, F., Ghazal, M., Ghazi, N., Aboelfetouh, A., Riad, A., Sandhu, H., Schaal, S., El-Baz, A., 2018. Early diabetic retinopathy diagnosis based on local retinal blood vessel analysis in optical coherence tomography angiography (OCTA) images. Med. Phys. 45 (10), 4582–4599.

Ferrara, D., Waheed, N.K., Duker, J.S., 2016. Investigating the choriocapillaris and choroidal vasculature with new optical coherence tomography technologies. Retin. Eye Res. 52, 130–155.

Fujimoto, J., Swanson, E., 2016. The development, commercialization, and impact of optical coherence tomography. Invest. Ophthalmol. Vis. Sci. 57 (9), OCT1–OCT13.

Hsu, S.-H., Han, Z., Leeman, J.E., Hu, Y.-H., Mak, R.H., Sudhyadhom, A., 2022. Synthetic CT generation for MRI-guided adaptive radiotherapy in prostate cancer. Front. Oncol. 12, 969463.

Huang, D., Swanson, E.A., Lin, C.P., Schuman, J.S., Stinson, W.G., Chang, W., Hee, M.R., Flotte, T., Gregory, K., Puliafito, C.A., Fujimoto, J.G., 1991. Optical coherence tomography. Science 254 (5035), 1178–1181.

Hussein, R., Zhao, M.Y., Shin, D., Guo, J., Chen, K.T., Armindo, R.D., Davidzon, G., Moseley, M., Zaharchuk, G., 2022. Multi-task deep learning for cerebrovascular disease classification and MRI-to-PET translation. In: International Conference on Pattern Recognition. ICPR, IEEE, pp. 4306–4312.

Isola, P., Zhu, J., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR.

Kashani, A.H., Chen, C.-L., Gahm, J.K., Zheng, F., Richter, G.M., Rosenfeld, P.J., Shi, Y., Wang, R.K., 2017. Optical coherence tomography angiography: a comprehensive review of current methods and clinical applications. Prog. Retin. Eye Res. 60, 66–100.

Kendall, A., Gal, Y., Cipolla, R., 2018. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR.

Lee, C.S., Tyring, A.J., Wu, Y., Xiao, S., Rokem, A.S., DeRuyter, N.P., Zhang, Q., Tufail, A., Wang, R.K., Lee, A.Y., 2019. Generating retinal flow maps from structural optical coherence tomography with artificial intelligence. Sci. Rep. 9 (1), 1–11.

Levine, E.S., Arya, M., Chaudhari, J., Greig, E.C., Alibhai, A.Y., Baumal, C.R., Witkin, A.J., Duker, J.S., Waheed, N.K., 2020. Repeatability and reproducibility of vessel density measurements on optical coherence tomography angiography in diabetic retinopathy. Graefe's Arch. Clin. Exp. Ophthalmol. 258, 1687–1695.

Li, R., Liu, S., Wang, G., Liu, G., Zeng, B., 2021. Jigsawgan: Auxiliary learning for solving jigsaw puzzles with generative adversarial networks. IEEE Trans. Image Process. 31, 513–524.

Li, P.L., O'Neil, C., Saberi, S., Sinder, K., Wang, K., Tan, B., Hosseinaee, Z., Bizhevat, K., Lakshminarayanan, V., 2020b. Deep learning algorithm for generating optical coherence tomography angiography (OCTA) maps of the retinal vasculature. Appl. Mach. Learn. 11511, 39–49.

Li, M., Zhang, Y., Ji, Z., Xie, K., Yuan, S., Liu, Q., Chen, Q., 2020a. Ipn-v2 and octa-500: Methodology and dataset for retinal image segmentation. In: ArXiv.

Li, Q., Zhu, X., Zou, S., Zhang, N., Liu, X., Yang, Y., Zheng, H., Liang, D., Hu, Z., 2022. Eliminating CT radiation for clinical PET examination using deep learning. Eur. J. Radiol. 154, 110422.

Liebel, L., Körner, M., 2018. Auxiliary tasks in multi-task learning. In: ArXiv.

Liu, S., Johns, E., Davison, A.J., 2019. End-to-end multi-task learning with attention. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR.

Ma, Y., Hao, H., Xie, J., Fu, H., Zhang, J., Yang, J., Wang, Z., Liu, J., Zheng, Y., Zhao, Y., 2020. ROSE: a retinal OCT-angiography vessel segmentation dataset and new model. IEEE Trans. Med. Imaging 40 (3), 928–939.

Mastropasqua, R., Toto, L., Mastropasqua, A., Aloia, R., De Nicola, C., Mattei, P.A., Di Marzio, G., Di Nicola, M., Di Antonio, L., 2017. Foveal avascular zone area and parafoveal vessel density measurements in different stages of diabetic retinopathy by optical coherence tomography angiography. Int. J. Ophthalmol. 10 (10), 1545.

McNaughton, J., Fernandez, J., Holdsworth, S., Chong, B., Shim, V., Wang, A., 2023. Machine learning for medical image translation: A systematic review. Bioengineering 10 (9), 1078.

Mou, L., Zhao, Y., Chen, L., Cheng, J., Gu, Z., Hao, H., Qi, H., Zheng, Y., Frangi, A., Liu, J., 2019. CS-Net: channel and spatial attention network for curvilinear structure segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. MICCAI.

Odena, A., Olah, C., Shlens, J., 2017. Conditional image synthesis with auxiliary classifier gans. In: International Conference on Machine Learning. ICML.

Oriani, E., 2022. Qpsnr: A quick PSNR/SSIM analyzer for linux. http://qpsnr.youlink.org. (Accessed: 07 November 2022).

Pan, B., Ji, Z., Chen, Q., 2022. MultiGAN: Multi-domain image translation from OCT to OCTA. In: Chinese Conference on Pattern Recognition and Computer Vision. PRCV, Springer, pp. 336–347.

Parrella, G., Vai, A., Nakas, A., Garau, N., Meschini, G., Camagni, F., Molinelli, S., Barcellini, A., Pella, A., Ciocca, M., et al., 2023. Synthetic CT in carbon ion radiotherapy of the abdominal site. Bioengineering 10 (2), 250.

Rajagopal, A., Natsuaki, Y., Wangerin, K., Hamdi, M., An, H., Sunderland, J.J., Laforest, R., Kinahan, P.E., Larson, P.E., Hope, T.A., 2022. Synthetic PET via domain translation of 3-D MRI. IEEE Trans. Radiat. Plasma Med. Sci. 7 (4), 333–343.

Richter, G.M., Chang, R., Situ, B., Chu, Z., Burkemper, B., Reznik, A., Bedrood, S., Kashani, A.H., Varma, R., Wang, R.K., 2018. Diagnostic performance of macular versus peripapillary vessel parameters by optical coherence tomography angiography for glaucoma. Transl. Vis. Sci. Technol. 7 (6), 21.

Roisman, L., Goldhardt, R., 2017. OCT angiography: an upcoming non-invasive tool for diagnosis of age-related macular degeneration. Curr. Ophthalmol. Rep. 5 (2), 136–140.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. MICCAI.

Saharia, C., Chan, W., Chang, H., Lee, C., Ho, J., Salimans, T., Fleet, D., Norouzi, M., 2022. Palette: Image-to-image diffusion models. In: ACM SIGGRAPH.

Spaide, R.F., Fujimoto, J.G., Waheed, N.K., Sadda, S.R., Staurenghi, G., 2018. Optical coherence tomography angiography. Prog. Retin. Eye Res. 64, 1–55.

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13 (4), 600–612.

Yang, D., Ran, A.R., Nguyen, T.X., Lin, T.P., Chen, H., Lai, T.Y., Tham, C.C., Cheung, C.Y., 2023. Deep learning in optical coherence tomography angiography: Current progress, challenges, and future directions. Diagnostics 13 (2), 326.

Yao, X., Alam, M.N., Le, D., Toslak, D., 2020. Quantitative optical coherence tomography angiography: a review. Exp. Biol. Med. 245 (4), 301–312.

Zhang, Z., Ji, Z., Chen, Q., Yuan, S., Fan, W., 2021b. Texture-guided U-Net for OCT-to-OCTA generation. In: Chinese Conference on Pattern Recognition and Computer Vision. PRCV.

Zhang, D., Lin, Y., Chen, H., Tian, Z., Yang, X., Tang, J., Cheng, K.T., 2022. Deep learning for medical image segmentation: tricks, challenges and future directions. arXiv.

Zhang, D., Zhang, H., Tang, J., Hua, X.-S., Sun, Q., 2021a. Self-regulation for semantic segmentation. In: IEEE/CVF International Conference on Computer Vision. ICCV.

Zhang, D., Zhang, H., Tang, J., Wang, M., Hua, X., Sun, Q., 2020. Feature pyramid transformer. In: European Conference on Computer Vision. ECCV.

Zhao, S., Geng, C., Guo, C., Tian, F., Tang, X., 2023. SARU: A self-attention ResUNet to generate synthetic CT images for MR-only BNCT treatment planning. Med. Phys. 50 (1), 117–127.

Zhu, J., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: IEEE/CVF International Conference on Computer Vision. ICCV.