Translation Identifiability-Guided Unsupervised Cross-Platform Super-Resolution for OCT Images

Jiahui Song†, Sagar Shrestha†, Xueshen Li⋆, Yu Gan⋆, and Xiao Fu†

†School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, USA
⋆Department of Biomedical Engineering, Stevens Insitute of Technology, Hoboken, USA

***Abstract*—*Optical Coherence Tomography* (OCT) is a non-invasive technique for obtaining detailed, cross-sectional images of coronary arteries. However, cost-effective OCT systems produce only low-resolution (LR) images. Unsupervised OCT super-resolution (OCT-SR) presents a cost-effective solution, eliminating the need for high-resolution (HR) systems or co-registered LR-HR image pairs. Existing unsupervised OCT-SR methods formulate the SR task as an image-to-image translation problem, and use CycleGAN as their backbone. However, CycleGAN is known to lack *translation identifiability* that can result in incorrect SR results. Existing methods often empirically combat this issue by using multiple regularization terms to incorporate expert-annotated side information, resulting in complicated learning losses and extensive annotations. This work proposes a translation identifiability-guided framework based on recent advances in unsupervised domain translation. Employing a *diversified distribution matching* module, our approach guarantees OCT translation identifiability under reasonable conditions using a simple and succinct learning loss. Numerical results indicate that our framework matches or surpasses the state-of-the-art (SOTA) baseline's performance while requiring considerably fewer resources, e.g., annotations, computation time, and memory.**

***Index Terms*—Optical coherence tomography, Cross-platform super-resolution, CycleGAN, Identifiability**

I. Introduction

Optical Coherence Tomography (OCT) is widely used in ophthalmology, cardiology, and dermatology. In particular, OCT plays a critical role obtaining cross-sectional images of coronary arteries (particularly for evaluating plaques and understanding blood vessel responses to intervention [1]), due to its non-invasive nature and fast data acquisition process. A preferred imaging technique is the *intravascular OCT* (IVOCT) due to portability, and ease of operation [2], [3]. However, the intrinsic hardware limitations of IVOCT systems often lead to a low optical resolution (10-20μm), hindering the precision required for detailed medical analysis. On the other hand, *high-resolution* (HR) OCT systems such as benchtop OCT can provide an optical resolution of (2μm). However, the use of HR systems is limited due to its high cost ($\sim$\$150,000) and portability challenges [2]. Hence, super-resolving IVOCT system-produced LR images is of great interest.

Many promising methods for super-resolution of LR OCT images have been proposed based on *supervised* deep learning [4]–[9]. By "supervised", it means that these methods require paired and *co-registered* LR and HR images, where "co-registration" means the LR and HR images describe exactly the same spatial area. However, it is impractical to assume availability of such pairs, since the two modalities are acquired by different imaging devices that cannot be run simultaneously. To circumvent this challenge, many existing supervised methods synthesize corresponding LR images from HR images by several downsampling and degradation procedures (see, e.g., [4]–[6]). However, such artificial downsampling approaches only reflect the true relationship between HR and LR images to a limited extent [2].

A recent work [10] considered OCT super-resolution using unpaired HR and LR images. The method there adopted the popular CycleGAN [11] framework, which was proposed in the context of unpaired image-to-image translation—as super-resolution can be considered as translating LR images to HR images. CycleGAN does not need paired or co-registered images, as they match distributions of the source and target domains. A notable challenge of CycleGAN-based frameworks like that in [10] is the lack of "translation identifiability" [12], [13]. Such a lack of translation identifiability could lead to swapping of samples between the two domains—e.g., a pathology present in image from one domain could be missing in the translated image in the other domain [14]. The existing works for unsupervised OCT super-resolution (OCT-SR), e.g., [10], [15], proposed adding multiple regularization terms on top of CycleGAN to combat this effect. These terms make use of human-annotations (e.g., labels for the presence of layers and segmentation masks) for empirical performance enhancement and showed promising super-resolution results.

In this work, we propose to employ a recently emerged unsupervised translation framework in [12] to offer an identifiability-guided design for OCT-SR. The work [12] proposed a provable remedy to CycleGAN's non-identifiability problem. The method there matches multiple distributions from the source and target domains using auxiliary information, instead of using a single distribution matching module as in CycleGAN. To apply this idea to unpaired OCT-SR, we use the presence of layered structure as a side information to define multiple conditional distributions. Empirical evaluation shows that the performance of the proposed approach is comparable to that of the state-of-the-art (SOTA) method [10]—but a number of notable advantages are obtained: First, the proposed

J. Song and S. Shrestha contributed equally. The work of X. Fu was supported by the National Science Foundation (NSF) CAREER Award under Project ECCS-2144889. The work of Y. Gan was supported by NSF CAREER Award under Project IIS-2239810.

method enjoys provable translation identifiability, reminiscent of the result in [12]. Second, the proposed approach features a simpler implementation relative to existing works in [10], [15]. This leads to substantially improved computational and memory reduction compared to the SOTA method in [10]. Third, the proposed method requires less extra information (i.e., the segmentation masks used [10]) to attain similar performance.

II. Background

A. OCT Super-resolution (OCT-SR)

Let $\mathcal{L}$ and $\mathcal{H}$ denote the space of LR and HR OCT images, respectively. We use $\boldsymbol{\ell}$ and $\boldsymbol{h}$ to denote the random vectors that are drawn from the distributions of the LR and HR images, respectively. Suppose that the paired and co-registered LR and HR images are associated by a translation function $\boldsymbol{f}^\star$ such that

$$\boldsymbol{h} = \boldsymbol{f}^\star(\boldsymbol{\ell}). \tag{1}$$

The function $\boldsymbol{f}^\star$ can be considered as the desired super-resolution operator. The goal of the *OCT super-resolution* (OCT-SR) task is to estimate $\boldsymbol{f}^\star$ so that given a test LR image, $\boldsymbol{\ell}^{\text{test}}$, one can obtain the corresponding HR image $\widehat{\boldsymbol{f}}(\boldsymbol{\ell}^{\text{test}})$, where $\widehat{\boldsymbol{f}}$ denotes the estimate of $\boldsymbol{f}^\star$.

Paired OCT-SR. In the paired setting, N pairs $\{\boldsymbol{\ell}_n, \boldsymbol{h}_n\}_{n=1}^N$ are assumed available. Then, the paired OCT-SR problem can be tackled by a regression loss:

$$\min_{\boldsymbol{f}} \ (1/N) \sum_{n=1}^{N} \ \|\boldsymbol{f}(\boldsymbol{\ell}_n) - \boldsymbol{h}_n\|_F^2,$$

which is essentially the main idea in [4]–[9] (with various loss metrics and regularization terms). However, paired LR-HR samples are not practical to obtain for OCT images because different devices are used to capture the LR and HR images. Due to the hardware limitations, simultaneously capturing the images of the same object in both HR and LR modalities is infeasible [10].

Unpaired OCT-SR. In the unpaired setting, only separately acquired samples $\{\boldsymbol{\ell}_i\}_{i=1}^I$ and $\{\boldsymbol{h}_j\}_{j=1}^J$ are available. Existing methods under this setting [10], [15] deal with this problem by adopting the CycleGAN framework [11]. The CycleGAN-based translation loss can be summarized as follows:

$$\min_{\boldsymbol{f},\boldsymbol{g}} \max_{\boldsymbol{d}_\ell,\boldsymbol{d}_h} \ \mathcal{L}_{\text{CycleGAN}} := \tag{2}$$
$$\mathcal{L}_{\text{GAN}}(\boldsymbol{f}, \boldsymbol{d}_h, \boldsymbol{\ell}, \boldsymbol{h}) + \mathcal{L}_{\text{GAN}}(\boldsymbol{g}, \boldsymbol{d}_\ell, \boldsymbol{h}, \boldsymbol{\ell}) + \lambda \mathcal{L}_{\text{cyc}}(\boldsymbol{f}, \boldsymbol{g}),$$

where $\boldsymbol{d}_\ell$ and $\boldsymbol{d}_h$ represent two discriminators in domains $\mathcal{L}$ and $\mathcal{H}$, respectively,

$$\mathcal{L}_{\text{GAN}}(\boldsymbol{f}, \boldsymbol{d}_h, \boldsymbol{\ell}, \boldsymbol{h}) = \tag{3}$$
$$\mathbb{E}_{\boldsymbol{h}}[\log \boldsymbol{d}_h(\boldsymbol{h})] + \mathbb{E}_{\boldsymbol{\ell}}[\log(1 - \boldsymbol{d}_h(\boldsymbol{f}(\boldsymbol{\ell})))],$$

$\mathcal{L}_{\text{GAN}}(\boldsymbol{g}, \boldsymbol{d}_\ell, \boldsymbol{h}, \boldsymbol{\ell})$ is defined in the same way, and the cycle-consistency term is defined as

$$\mathcal{L}_{\text{cyc}}(\boldsymbol{f}, \boldsymbol{g}) = \tag{4}$$
$$\mathbb{E}_{\boldsymbol{h}} \left[\|\boldsymbol{f}(\boldsymbol{g}(\boldsymbol{h})) - \boldsymbol{h}\|_1\right] + \mathbb{E}_{\boldsymbol{\ell}} \left[\|\boldsymbol{g}(\boldsymbol{f}(\boldsymbol{\ell})) - \boldsymbol{\ell}\|_1\right].$$

The CycleGAN loss obatins an $\widehat{\boldsymbol{f}}$ such that

$$\widehat{\boldsymbol{f}}(\boldsymbol{\ell}) \overset{(\mathsf{d})}{=\!=} \boldsymbol{h}, \tag{5}$$

i.e., an $\widehat{\boldsymbol{f}}$ that matches the distribution of $\widehat{\boldsymbol{f}}(\ell)$ and that of $\boldsymbol{h}$, and $\overset{(\mathsf{d})}{=\!=}$ is used to denote the equivalence between two distributions. This $\widehat{\boldsymbol{f}}$ is potentially of interest as $\boldsymbol{f}^\star$ also attains the same distribution matching. However, a critical issue of CycleGAN and its variants lies in the lack of *translation identifiability*. Mainly, besides $\boldsymbol{f}^\star$, there are (an infinite number of) other translation functions $\widehat{\boldsymbol{f}}$ that can attain (5) [12], [13]. When CycleGAN outputs $\widehat{\boldsymbol{f}}$ that is not $\boldsymbol{f}^\star$, then a wrong translation $\widehat{\boldsymbol{f}}(\boldsymbol{\ell}_i) \neq \boldsymbol{f}^\star(\boldsymbol{\ell}_i)$ may occur.

B. Existing Methods for Unpaired OCT-SR

The recent work *Cross-Platform Structure-Aware GAN* (CSPA-GAN) [10] leveraged side information regarding presence/absence of layered structure and the subsequent decomposition into individual layers (i.e., intima, media, and adventitia layers) to regularize the basic CycleGAN framework.

To be precise, the unpaired dataset is augmented with expert annotations to obtain $\{(\boldsymbol{\ell}_i, u_i^{(\ell)}, \boldsymbol{k}_i^{(\ell)})\}_{i=1}^I$ and $\{(\boldsymbol{h}_j, u_j^{(h)}, \boldsymbol{k}_j^{(h)})\}_{j=1}^J$, where $u_i^{(\ell)}$ and $u^{(h)}$ are binary labels representing layer/non-layer structure and $\boldsymbol{k}_i^{(\ell)}$ and $\boldsymbol{k}_j^{(h)}$ are the corresponding layer masks. The translation functions, $\boldsymbol{f}$ and $\boldsymbol{g}$, are decomposed into $(\boldsymbol{f}_{\text{E}}, \boldsymbol{g}_{\text{E}})$ and $(\boldsymbol{f}_{\text{D}}, \boldsymbol{g}_{\text{D}})$, which are the encoders and decoders such that $\boldsymbol{f} = \boldsymbol{f}_{\text{D}} \circ \boldsymbol{f}_{\text{E}}$ and $\boldsymbol{g} = \boldsymbol{g}_{\text{D}} \circ \boldsymbol{g}_{\text{E}}$, respectively. Two additional neural networks are introduced for each translation functions: the neural networks $\boldsymbol{c}_f$, $\boldsymbol{c}_g$ and $\boldsymbol{s}_f$, $\boldsymbol{s}_g$ for classification of layer/non-layer structure and segmentation of different layers, respectively. Their loss function can be written as follows:

$$\mathcal{L}_{\text{CPSA}} = \mathcal{L}_{\text{CycleGAN}} + \lambda_2 \mathcal{L}_{\text{seg}} + \lambda_3 \mathcal{L}_{\text{cls}} + \lambda_4 \mathcal{L}_{\text{emb}} + \lambda_5 \mathcal{L}_{\text{id}},$$

where, using CE to denote the cross-entropy loss, and $\mathcal{L}_{\text{seg}}$, $\mathcal{L}_{\text{cls}}$, $\mathcal{L}_{\text{emb}}$, and $\mathcal{L}_{\text{id}}$ are expressed as follows

$$\mathbb{E}_{\boldsymbol{k}^{(\ell)},\boldsymbol{\ell},\boldsymbol{k}^{(h)},\boldsymbol{h}} \left[\text{CE}(\boldsymbol{s}_{\boldsymbol{f}}(\boldsymbol{f}_{\text{E}}(\boldsymbol{\ell}), \boldsymbol{k}^{(\ell)}) + \text{CE}(\boldsymbol{s}_{\boldsymbol{g}}(\boldsymbol{g}_{\text{E}}(\boldsymbol{h}), \boldsymbol{k}^{(h)})\right],$$
$$\mathbb{E}_{u^{(\ell)},\boldsymbol{\ell},u^{(h)},\boldsymbol{h}} \left[\text{CE}(\boldsymbol{c}_{\boldsymbol{f}}(\boldsymbol{f}_{\text{E}}(\boldsymbol{\ell}), u^{(\ell)}) + \text{CE}(\boldsymbol{c}_{\boldsymbol{g}}(\boldsymbol{g}_{\text{E}}(\boldsymbol{h}), u^{(h)})\right],$$
$$\mathbb{E}_{\boldsymbol{h},\boldsymbol{\ell}} \ \|\boldsymbol{g}_{\text{E}}(\boldsymbol{f}(\boldsymbol{\ell})) - \boldsymbol{f}_{\text{E}}(\boldsymbol{\ell})\|_1 + \|\boldsymbol{f}_{\text{E}}(\boldsymbol{g}(\boldsymbol{h})) - \boldsymbol{g}_{\text{E}}(\boldsymbol{h})\|_1,$$
$$\mathbb{E}_{\boldsymbol{h},\boldsymbol{\ell}} \ \|\boldsymbol{g}(\boldsymbol{\ell}) - \boldsymbol{\ell}\|_1 + \|\boldsymbol{f}(\boldsymbol{h}) - \boldsymbol{h}\|_1,$$

respectively. The terms $\mathcal{L}_{\text{seg}}$ and $\mathcal{L}_{\text{cls}}$ encourage the latent representations to encode the class and layer information, respectively. The term $\mathcal{L}_{\text{embd}}$ makes the latent embeddings similar for the source and translated images. Finally, the term $\mathcal{L}_{\text{id}}$ is called the identity loss, and is widely used as a regularization [11] that has been observed to improve training stability [16].

The CSPA method in [10] enjoyed empirical success and achieved state-of-the-art (SOTA) results for the unpaired OCT-SR task. Nonetheless, a number of challenges remain. First, in theory, the issue of translation non-identifiability could still happen, as the regularization terms do not ensure provable identifiability of $\boldsymbol{f}^\star$. Second, $\mathcal{L}_{\text{seg}}$ relies on the availability of

the segmentation masks for layered images. Obtaining such masks is a much more challenging annotation task compared to obtaining the binary labels, $u^{(\ell)}, u^{(h)}$. The Coronary-GAN in [15] also has similar challenges. To address these challenges, in the next section, we propose a simple modification to CycleGAN loss, which ensures that $\boldsymbol{f}^\star$ is correctly estimated without using the segmentation masks.

III. Proposed Approach

A. Preliminaries

It was noticed in [12], that the reason for translation non-identifiability of CycleGAN was due to the existence of the so-called "measure preserving functions" (MPF) for the image distributions, i.e., $p(\boldsymbol{\ell})$ and $p(\boldsymbol{h})$ in our case. A non-identity function $\boldsymbol{m}_\ell : \mathcal{L} \to \mathcal{L}$ is called a MPF for $p(\boldsymbol{\ell})$ if $\boldsymbol{m}_{\boldsymbol{\ell}}(\boldsymbol{\ell})$ has the same distribution as $\boldsymbol{\ell}$, i.e., $\boldsymbol{m}_\ell(\boldsymbol{\ell}) \overset{(\mathrm{d})}{=\!=} \boldsymbol{\ell}$. Under the existence of such $\boldsymbol{m}_\ell$, a new function $\overline{\boldsymbol{f}}$ can be constructed as $\overline{\boldsymbol{f}} := \boldsymbol{f}^\star \circ \boldsymbol{m}_\ell$, where $\circ$ represents composition, which can also achieve minimum CycleGAN loss (2). Hence, optimizing CycleGAN loss may result in estimating $\widehat{\boldsymbol{f}} = \overline{\boldsymbol{f}}$ instead of the true $\boldsymbol{f}^\star$. A simple example is as follows: suppose $\boldsymbol{p}(\boldsymbol{\ell}) = \mathcal{N}(\mu, \sigma^2)$. Then the function $\boldsymbol{m}_\ell(\boldsymbol{\ell}) = -\boldsymbol{\ell} + 2\mu$ would be an MPF for $\boldsymbol{p}(\boldsymbol{\ell})$ since $\boldsymbol{m}_\ell(\boldsymbol{\ell})$ has the same distribution ($\mathcal{N}(0, \sigma^2)$) as that of $\boldsymbol{\ell}$.

The key idea in [12] is to consider multiple conditional distributions $\{p(\boldsymbol{\ell}|u = u_t)\}_{t=1}^T$ and $\{p(\boldsymbol{h}|u = u_t)\}_{t=1}^T$, defined over different assignments of the so-called auxiliary variable u. In our running example of Gaussian distributions, consider $T = 2$, and $p(\boldsymbol{\ell}|u = u_1) = \mathcal{N}(\mu_1, \sigma^2)$ and $p(\boldsymbol{\ell}|u = u_2) = \mathcal{N}(\mu_2, \sigma^2)$. Then, although $\boldsymbol{m}_\ell^{(1)}(\boldsymbol{\ell}) = -\boldsymbol{\ell} + 2\mu_1$ and $\boldsymbol{m}_\ell^{(2)}(\boldsymbol{\ell}) = -\boldsymbol{\ell} + 2\mu_2$ are individual MPFs for $p(\boldsymbol{\ell}|u = u_1)$ and $p(\boldsymbol{\ell}|u = u_2)$, respectively, we cannot find a unified function $\boldsymbol{m}_\ell$ that is the MPF for both conditional distributions. Based on this observation, the work [12] proposed to match multiple distributions of $\boldsymbol{h}|_{u=u_t}$ and $\boldsymbol{f}(\boldsymbol{\ell})|_{u=u_t}, \forall t \in \{1, \ldots, T\}$, instead of just matching the pair of $p(\boldsymbol{h})$ and $p(\boldsymbol{f}(\boldsymbol{\ell}))$ as in the original CycleGAN loss (2). The work [12] proved that if the conditional distributions are "diverse" enough, $\widehat{\boldsymbol{f}}$ is more likely to be $\boldsymbol{f}^\star$.

The resulting loss function is a simple modification of the CycleGAN loss where $\mathcal{L}_{\mathrm{GAN}}$ is replaced by $\mathcal{L}_{\mathrm{DGAN}}$ as follows:

$$\mathcal{L}_{\mathrm{DGAN}}(\boldsymbol{f}, \boldsymbol{d}_h, \boldsymbol{\ell}, \boldsymbol{h}) = \sum_{t=1}^{T} \mathsf{Pr}(u = u_t) \times \big(\mathbb{E}_{\boldsymbol{h}|u_t}[\log \boldsymbol{d}_h(\boldsymbol{h}, u_t)] + \mathbb{E}_{\boldsymbol{\ell}|u_t}[\log(1 - \boldsymbol{d}_h(\boldsymbol{f}(\boldsymbol{\ell}), u_t))]\big), \tag{6}$$

where we re-defined $\boldsymbol{d}_{\boldsymbol{h}}(\cdot, u_t)$ to denote conditional discriminators that also take in u_t as an argument. In practice, $\{u_t\}_{t=1}^T$ could correspond to different attributes or categories/labels of the samples. Hence, $\{p(\boldsymbol{\ell}|u = u_t)\}_{t=1}^T$ and $\{p(\boldsymbol{h}|u = u_t)\}_{t=1}^T$ could correspond to different (overlapped) subsets of data with corresponding attributes/labels.

B. Application to OCT Super-resolution

In order to apply the loss function (6) to unpaired OCT-SR, we need to define auxiliary variables $\{u_t\}_{t=1}^T$, such that they induce diverse conditional distributions. We also hope that $\{u_t\}_{t=1}^T$ are easy to obtain in practice. To that end, we use the presence of layered structure as u, i.e., $u_1 =$"`layered`" and $u_2 =$"`non-layered`", with $T = 2$—which is already available in the OCT dataset of [10]. With this, the overall loss function $\mathcal{L}$ can be expressed as follows:

$$\begin{aligned}\mathcal{L} = \mathcal{L}_{\mathrm{DGAN}}(\boldsymbol{f}, \boldsymbol{d}_h, \boldsymbol{\ell}, \boldsymbol{h}) \\ + \mathcal{L}_{\mathrm{DGAN}}(\boldsymbol{g}, \boldsymbol{d}_\ell, \boldsymbol{h}, \boldsymbol{\ell}) + \lambda \mathcal{L}_{\mathrm{cyc}}(\boldsymbol{f}, \boldsymbol{g}).\end{aligned} \tag{7}$$

The most attractive feature of the proposed loss function (7) is that the solution to (7) correctly identifies the $\boldsymbol{f}^\star$ under some reasonable conditions:

Theorem 1 *Assume that there exists a deterministic continuous translation function $\boldsymbol{f}^\star$ between the LR and HR OCT images as described in* (1)*. Also assume that $\boldsymbol{f}^\star$ is invertible. Let $p(\boldsymbol{\ell}|u = u_t) > 0, \forall \boldsymbol{\ell} \in \mathcal{L}, t \in \{1, 2\}$. Assume that for any disjoint and open sets $\mathcal{A}, \mathcal{B} \subseteq \mathcal{L}$, there exists $u_{(\mathcal{A},\mathcal{B})} \in \{1, 2\}$ such that*

$$\int_{\mathcal{A}} p(\boldsymbol{\ell}|u = u_{(\mathcal{A},\mathcal{B})}) d\boldsymbol{\ell} \neq \int_{\mathcal{B}} p(\boldsymbol{\ell}|u = u_{(\mathcal{A},\mathcal{B})}) d\boldsymbol{\ell}.$$

Then the solution of (7)*, $\widehat{\boldsymbol{f}}$, correctly identifies $\boldsymbol{f}^\star$, i.e., $\widehat{\boldsymbol{f}} = \boldsymbol{f}^\star$ almost everywhere. Same statements also hold for $\widehat{\boldsymbol{h}}$.*

The proof of Theorem 1 directly follows that of [12, Theorem 1]. Theorem 1 asserts that when the distributions of layered ($u = u_1$) and non-layered ($u = u_2$) OCT images are diverse, i.e., at least one of $\{p(\boldsymbol{\ell}|u = u_t)\}_{t=1}^2$ assign different probabilities to different subsets $\mathcal{A}$ and $\mathcal{B}$ of LR images, then (7) correctly recovers the true $\boldsymbol{f}^\star$. Note that for any two pairs, $(\mathcal{A}, \mathcal{B})$ and $(\mathcal{A}', \mathcal{B}')$, one could have $u_{(\mathcal{A},\mathcal{B})} \neq u_{(\mathcal{A}',\mathcal{B}')}$. This makes the condition easier to satisfy.

Remark 1 *Besides the identifiability guarantees, a notable advantage of the proposed loss lies in its relative simplicity: It does not involve multiple regularization terms as in existing works [cf. $\mathcal{L}_{\mathrm{CPSA}}$]. Therefore, training and tuning could be more efficient in practice. The proposed method also does not require the information of masks (but it was used in $\mathcal{L}_{\mathrm{seg}}$ of $\mathcal{L}_{\mathrm{CPSA}}$), which is much harder to acquire than the layer annotations.*

IV. Numerical Results

In this section, we demonstrate the efficacy of the proposed approach for super-resolving real low-resolution OCT images.

A. Dataset

We use the HR and LR OCT coronary images collected by [15] and [10]. There are a total of 104 LR images and 104 HR images that have 512×512 pixels. The HR images are collected by a Thorlabs Ganymede commercial OCT system (whose axial and lateral resolutions are 3μm and 8μm, all in

TABLE I
FID AND COMPUTE COST OF ALL METHODS.

Method	FID (↓)	Memory (GB) (↓)	Run-time / epoch (seconds) (↓)
CycleGAN	119.03 ± 8.00	14.30	13.84
Coronary-GAN	127.94 ± 2.50	23.72	21.59
CPSA-GAN	92.09 ± 6.64	37.70	37.77
Proposed	84.46 ± 3.31	25.84	21.17

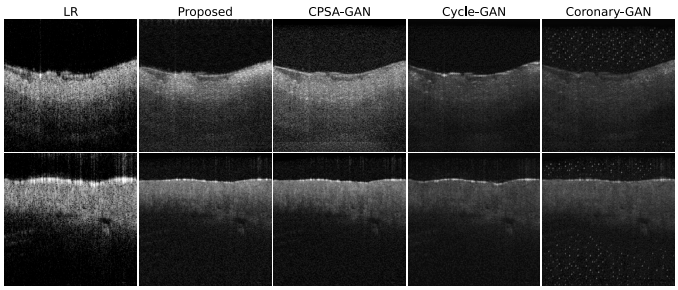

Fig. 1. A sample of super-resolution results from all methods.

air). The LR images are collected by a Lumedica OQ Labscope OCT system (where the axial and lateral resolutions are 7μm and 18μm, all in air). The numbers of layered and non-layered LR and HR images are both 52. We randomly split the dataset into 70 training images and 34 testing images. Additionally, we randomly flip the images horizontally and vertically as data augmentation strategy to mitigate potential over-fitting issue due to small the sample size.

B. Setting

We follow a similar neural architecture and the hyperparameter settings as suggested by [10]. Mainly, the generator architecture is the same as that in [10], except that we remove the classifier and segmentation mask estimating neural networks. We design our conditional discriminators by converting the discriminator in [10] into a multi-task discriminator [17] by modifying the last layer to have two output channels instead of one. Specifically, let $\boldsymbol{w}_h : \mathcal{H} \to \mathbb{R}^2$ denote the aforementioned neural network with two output channels. We have

$$\boldsymbol{d}_h(\boldsymbol{h}, u) = \begin{cases} [\boldsymbol{w}_h(\boldsymbol{h})]_0 & \text{if } u = u_0 \\ [\boldsymbol{w}_h(\boldsymbol{h})]_1 & \text{if } u = u_1. \end{cases}$$

We use Adam [18] with an initial learning rate of 0.0001 with a batch size of 2 for 3000 epochs. We also use the identity loss $\mathcal{L}_{\text{id}}$ [11] as an additional regularization for enhanced stability.

C. Evaluation Metrics and Baselines

Since paired LR-HR images are not available for testing super-resolved images with the expected high-resolution images, we use reference-free evaluation metrics to indirectly assess the translation quality and validate our theoretical claims. Mainly, we use the *Fréchet Inception Distance* (FID) [19] to evaluate the quality of the super-resolved images. FID measures the distribution divergence between the super-resolved and HR images in the representation space of a pre-trained neural network. In our context, a lower FID means that the translated LR image is better matched with the HR distribution—which is more preferred. Additionally, we make qualitative observation of the super-resolved images to assess the translation performance. Finally, we also observe the computational resources used by all methods, as computational efficiency is an important consideration for network training in low-resource settings.

We use the unpaired OCT super-resolution methods, namely, Coronary-GAN [15], and CPSA-GAN [10] as our main baselines. We also present the result of the plain-vanilla CycleGAN [11] as a reference.

D. Results

We run three-fold cross validation and observe the average result over the 3 test sets. Table I shows the quantitative results for all methods. One can see that the proposed method attains the smallest FID which suggests that the super-resolved images are very similar to HR images. The SOTA method CPSA-GAN also works well, with quite close FID scores. Nonetheless, the proposed method consumes much less GPU memory and has a smaller runtime relative to CPSA-GAN. Our method uses 25.84GB memory, which is more than 30% reduction from that of CPSA-GAN. The runtime for an epoch of our method is 44% less than that of the SOTA method. These show the advantage of our identifiability-guided design: It attains comparable translation quality with the SOTA method but uses considerably less computational resources.

Fig. 1 shows some qualitative results. One can see that the proposed method maintains structural similarity of the super-resolved images with the LR images. The results are comparable to CPSA-GAN. Both our method and CPSA-GAN perform better than CycleGAN and Coronary-GAN in terms of visual quality. Notably, our method did not use the segmentation masks as used in CPSA-GAN. This shows that translation identifiability-guided designs can potentially reduce the annotation efforts or the amount of extra information to attain SOTA performance.

V. CONCLUSION

In this work, we proposed a translation identifiability-guided method for super-resolution of OCT images. By matching multiple conditional distributions induced by the presence/absence of the layered structure in coronary images, we build a translation framework can provably identify the ground-truth translation function under reasonable conditions. The resulting method is a simple modification of the original CycleGAN objective and only requires minimal side information (i.e., layered/non-layered attribute), which does not use excessive masking annotations as in a SOTA baseline. Numerical and qualitative results on real coronary images show that the proposed method can match the performance of the SOTA baseline—but uses much less side information and computational resources.

REFERENCES

[1] W. Wijns, J. Shite, M. R. Jones, S. W.-L. Lee, M. J. Price, F. Fabbiocchi, E. Barbato, T. Akasaka, H. Bezerra, and D. Holmes, "Optical coherence tomography imaging during percutaneous coronary intervention impacts physician decision-making: Ilumien i study," *European heart journal*, vol. 36, no. 47, pp. 3346–3355, 2015.

[2] Sanghoon Kim, Michael Crose, Will J. Eldridge, Brian Cox, William J. Brown, and Adam Wax, "Design and implementation of a low-cost, portable oct system," *Biomed. Opt. Express*, vol. 9, no. 3, pp. 1232–1243, 2018.

[3] R. L. Shelton, W. Jung, S. I. Sayegh, D. T. McCormick, J. Kim, and S. A. Boppart, "Optical coherence tomography for advanced screening in the primary care office," *Journal of biophotonics*, vol. 7, no. 7, pp. 525–533, 2014.

[4] Q. Hao, K. Zhou, J. Yang, Y. Hu, Z. Chai, Y. Ma, G. Liu, Y. Zhao, S. Gao, and J. Liu, "High signal-to-noise ratio reconstruction of low bit-depth optical coherence tomography using deep learning," *Journal of Biomedical Optics*, vol. 25, no. 12, p. 123702, 2020.

[5] A. Lichtenegger, M. Salas, A. Sing, M. Duelk, R. Licandro, J. Gesperger, B. Baumann, W. Drexler, and R. A. Leitgeb, "Reconstruction of visible light optical coherence tomography images retrieved from discontinuous spectral data using a conditional generative adversarial network," *Biomedical optics express*, vol. 12, no. 11, pp. 6780–6795, 2021.

[6] B. Qiu, Y. You, Z. Huang, X. Meng, Z. Jiang, C. Zhou, G. Liu, K. Yang, Q. Ren, and Y. Lu, "N2NSR-OCT: Simultaneous denoising and super-resolution in optical coherence tomography images using semisupervised deep learning," *Journal of biophotonics*, vol. 14, no. 1, p. e202000282, 2021.

[7] Y. Zhang, T. Liu, M. Singh, E. Çetintaş, Y. Luo, Y. Rivenson, K. V. Larin, and A. Ozcan, "Neural network-based image reconstruction in swept-source optical coherence tomography using undersampled spectral data," *Light: Science & Applications*, vol. 10, no. 1, p. 155, 2021.

[8] X. Li, S. Cao, H. Liu, X. Yao, B. C. Brott, S. H. Litovsky, X. Song, Y. Ling, and Y. Gan, "Multi-scale reconstruction of undersampled spectral-spatial oct data for coronary imaging using deep learning," *IEEE transactions on bio-medical engineering*, vol. 69, no. 12, pp. 3667–3677, 2022.

[9] W. Lee, H. S. Nam, J. Y. Seok, W.-Y. Oh, J. W. Kim, and H. Yoo, "Deep learning-based image enhancement in optical coherence tomography by exploiting interference fringe," *Communications Biology*, vol. 6, no. 1, p. 464, 2023.

[10] X. Li, A. Shamouil, X. Hou, B. C. Brott, S. H. Litovsky, Y. Ling, and Y. Gan, "Cross-platform super-resolution for human coronoary oct imaging using deep learning," IEEE, 2024, accepted.

[11] J.-Y. Zhu *et al.*, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2242–2251.

[12] S. Shrestha and X. Fu, "Towards identifiable unsupervised domain translation: A diversified distribution matching approach," *arXiv preprint arXiv:2401.09671*, 2024.

[13] N. Moriakov, J. Adler, and J. Teuwen, "Kernel of cyclegan as a principle homogeneous space," *arXiv preprint arXiv:2001.09061*, 2020.

[14] J. P. Cohen, M. Luck, and S. Honari, "Distribution matching losses can hallucinate features in medical image translation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I*. Springer, 2018, pp. 529–536.

[15] X. Li, H. Liu, X. Song, B. C. Brott, S. H. Litovsky, and Y. Gan, "Structural constrained virtual histology staining for human coronary imaging using deep learning," in *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, 2023, pp. 1–5.

[16] T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu, "Contrastive learning for unpaired image-to-image translation," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16*. Springer, 2020, pp. 319–345.

[17] M.-Y. Liu, X. Huang, A. Mallya, T. Karras, T. Aila, J. Lehtinen, and J. Kautz, "Few-shot unsupervised image-to-image translation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (CVPR)*, 2019, pp. 10 551–10 560.

[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of International Conference on Learning Representations (ICLR)*, 2015.

[19] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.