Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: https://www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine

Generalized degradation-based adversarial learning for unsupervised super-resolution of endomicroscopy images

Linghao Meng [a,1], Yangxi Li [b,1], Yuchao Zheng [c], Yu Feng [c], Fang Chen [b,d], Longfei Ma [c], Hongen Liao [b,c,d,*]

[a] Tanwei College, Tsinghua University, Beijing 100084, China
[b] School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[c] School of Biomedical Engineering, Tsinghua University, Beijing 100084, China
[d] Institute of Medical Robotics, Shanghai Jiao Tong University, Shanghai 200240, China

ARTICLE INFO

Keywords:
Probe-based confocal laser endomicroscopy
Degradation
Unsupervised super-resolution
Adversarial learning

ABSTRACT

Background and Objective: In recent years, probe-based confocal laser endomicroscopy (pCLE) has become an emerging *optical biopsy* method for *in situ* imaging and diagnosis, which aids in the accurate early diagnosis of diseases like inflammation and cancer. However, due to physical constraints induced by the fiber bundle used for signal acquisition, obtaining pCLE images of high resolution is challenging. Consequently, in this study, we aim to improve pCLE image quality through the implementation of advanced post-processing techniques.
Methods: Here we propose an unsupervised single image super-resolution framework, which is free of using high-resolution pCLE images as reference and improves image quality significantly. The framework consists of a degradation module, a style transformation module and a super resolution module. In the degradation module, we propose an innovative distribution assumption module to randomize the fiber optic position distribution, enabling us to simulate the imaging principles of pCLE and create synthetic pCLE images for training.
Results: With the integration of modules, both quantitative and qualitative analyses highlight the remarkable efficiency of our pipeline in super-resolving images compared to state-of-the-art methods. Our framework also demonstrates strong generalization capability, effectively mitigating the impact of pCLE system's intrinsic characteristics on image super-resolution. This feature is particularly advantageous as it allows the framework to circumvent redundant training when applied to various devices.
Conclusions: With the outstanding super-resolution and generalization capability, our proposed methodology enables clearer observation of image details and more accurate localization of micro structures, which contributes to precise identification of lesion areas and diagnostic accuracy enhancement.

1. Introduction

Gastrointestinal diseases significantly impact human health, with gastrointestinal tumors standing out for their high rates of occurrence and fatality [1,2]. Cancer statistics in 2023 shows that in the United States, the survival rate of gastric cancer is the second lowest [2]. Early detection is extremely important for reducing mortality caused by gastrointestinal diseases. Generally, X-rays are utilized to check whether a patient has gastrointestinal tumor or invasion of surrounding tissues [3]. Computed tomography (CT) can also assist in staging gastric cancer, determining the extent of the lesion, and evaluating the presence of metastasis [4]. However, X-rays and CT scans expose the human body to harmful ionizing radiation. Moreover, early-stage gastric tumor lesions often have a similar density to surrounding normal tissues [4], making it difficult to distinguish them accurately using X-rays and CT scans. Besides, magnetic resonance imaging (MRI) and ultrasound (US)imaging are not the preferred methods for detecting gastrointestinal cancer, since MRI has a long acquisition time and is susceptible to soft tissue deformation, while US images has relatively low resolution [5,6].

As an emerging optical imaging technique, probe-based confocal laser endomicroscopy (pCLE) allows *in vivo*, real-time, and harmless detection of the mucosal layer of the gastrointestinal tract at cellular and subcellular resolution for precise diagnosis [7]. Specifically, pCLE works based on the principle of a confocal microscope. Using a fiber bundle to transmit low-power laser and collect emitted fluorescence signal, the probe is capable of reaching the inside of the human body in a

[*] Corresponding author at: School of Biomedical Engineering, Tsinghua University, Beijing 100084, China.
E-mail address: liao@tsinghua.edu.cn (H. Liao).

[1] These authors contributed equally.

https://doi.org/10.1016/j.cmpb.2025.108914
Received 10 December 2024; Received in revised form 12 April 2025; Accepted 13 June 2025
Available online 28 June 2025

minimally invasive or non-invasive approach to directly observe microscopic images without tissue sampling [7,8]. Thus, pCLE is considered as a non-invasive, real-time, and near-pathological detection method for gastrointestinal and pancreatobiliary diseases [7,9], especially for early diagnosis of cancer.

However, the reliance on an optical fiber bundle fundamentally limits the image quality of pCLE [10]. On the one hand, each fiber acts as a single-pixel detector and they are irregularly arranged with intervals. Thus the original image only has discrete pixels, and the resolution is inherently limited [10]. On the other hand, the transmission of the optical signal in the fiber bundle might be affected by crosstalk, attenuation and noise, which further reduce the image quality. As hardware improvements are difficult to achieve and preprocessing methods encounter physical limitations, an alternative approach to enhance image resolution and quality is employing Single Image Super-Resolution (SISR) algorithms as postprocessing techniques [10,11]. To be specific, SISR has gained popularity for reconstructing high-resolution (HR) images from low-resolution (LR) inputs. With advancements in deep learning, convolutional neural networks (CNNs) have demonstrated significant improvements in SISR, where traditional algorithms using CNNs are trained with paired LR-HR image datasets through supervised learning [12]. However, due to the limitation of image acquisition equipment and acquisition time, it is difficult to collect ground truth to support the traditional supervised training. Hence, researchers have studied unsupervised learning frameworks and made considerable progress in SISR, where the generation of pseudo LR-HR pairs functions as a fundamental preprocessing step.

In most SISR tasks, researchers downscale HR images via the bicubic operation to simulate the corresponding LR images and then input the paired dataset into the training pipeline. Considering that SR effect could be interfered with by the limitation of original images' resolution, Chen et al. [13] presented SR frameworks consisting of novel image translation blocks, in which synthetic LR images were trained together with real LR images. Later, Wei et al. proposed a novel domain-distance aware super-resolution (DASR) approach to addresses the domain gap issue between synthetic and real images [14]. Liu et al. employed dual discriminator GANs to construct a novel unsupervised fusion-based image super-resolution framework, eliminating the distribution gap between hyperspectral and multispectral images [15]. Following the contrastive learning strategy, in the proposed LDCT SR network, Chi et al. introduced a prior degradation estimator (PDE) to estimate the degradation features in the LDCT images [16]. Mishra et al. presented a novel framework that uses contrastive training followed by a decoder to generate a style image for neural style transfer learning for image unsupervised super-resolution [17]. Lin et al. proposed two plug-and-play modules for MRI SISR with Transformer-based networks via high frequency information restoration [18]. Dong et al. introduced a novel Flow-based Truncated Denoising Diffusion Model (FTDDM) for multi-scale super-resolution of in vivo MRSI dataset [19]. Based on an orientation operator in the encoder and multi-scale feature fusion strategy, Huang et al. proposed a transformer-based SISR model for radiographic images [20]. Qiu et al. proposed the progressive feedback residual attention network (PFRN) for cardiac magnetic resonance imaging super-resolution, in which the feedback of more context information helps LR images to be better reconstructed [21]. Mishra et al. proposed a novel self-fusion-based network (Self-FuseNet) combining a self-fusion subblock and a customized-UNet, which employs no HR images and gets rid of arranging datasets of similar image distribution [22].

In the super-resolution task of pCLE images, the lack of ground truth is a challenge. Ravi et al. proposed an unsupervised adversarial SR framework for pCLE images, in which they figured out Voronoi vectorization as a degradation method specific and reasonable to pCLE, thus maintaining the same physical acquisition properties between SR and LR images [10]. Szczotka et al. extended the ZSSR framework while they obtained pseudo HR by reducing redundant pixels from a given pCLE image [23]. Later, Zhang et al. investigated a sharpness-aware SR framework, in which pseudo HR was generated by implementing augmentation on original dataset, consequently enduing different image sharpness level in LR-HR pairs after downscaling [11]. Eadie et al. combined fiber bundle rotation into a multi-frame super-resolution deep learning framework [24]. Yang et al. implemented noise prediction networks to handle degraded images with noise-based prior information in a collaborative trained manner and integrated them with super-resolution networks [25]. Zhou et al. constructed a realistic noise statistics model specific to CLE data and proposed a context-aware kernel estimation to denoise pCLE images [26]. Zhang et al. applied a deep-learning-based image super-resolution (DL-SR) method on low-resolution (LR) endomicroscopy images that were acquired by a novel end-expandable optical fiber probe [27]. Even if the methods of generating LR-HR pairs varied in previous works, the degradation process remained indispensable in all frameworks.

Although the methods above could generate LR-HR pairs with considerable degree of realism and had the ability to super-resolve some anatomical structures, they all suffered from several non-negligible shortcomings: (i) The degradation process relied on real fibers' position of pCLE devices, causing training repetition and limitations in module outcomes, which results in time consumption and possible performance reduction. (ii) Basically, although the pseudo LR-HR pairs could be formed by simulating pCLE imaging principles with synthetic degradation, it differs from real pCLE images to some extent and thus induces mismatching and error.

In this paper, we propose an unsupervised deep learning framework for pCLE image super-resolution, as shown in Fig. 1, which copes with the problems above and has considerable SISR capabilities. In this framework, there are three modules working sequentially to achieve the overall effect. To our knowledge, this paper is the first to propose a *degradation module* based on random fiber distribution assumption for unsupervised pCLE SISR, which is capable of synthesizing pseudo pCLE images from high-resolution histopathological images regardless of specific fibers' position in pCLE equipment. Thus our proposed framework has high generalization ability and is suitable for various fiber distributions. After that, to enhance the matching degree of LR-HR pairs and super-resolve images, a *style transformation module* and a *super resolution module* are proposed. Here, we also propose a novel integration of loss functions to improve the reliability and effect of super-resolution results. The outstanding performance of the framework and necessity of our degradation module are examined through qualitative and quantitative experiments. We also innovatively compare the effect of color depth on the background noise reduction in the results.

2. Methods

Details of the proposed framework are presented in Figs. 1 and 2. First, a *degradation module* is utilized to generate synthetic pCLE images based on distribution assumption module (DAM). Then in the *style transformation module*, we introduce an improved CycleGAN for transferring image styles and generating pseudo LR-HR pairs. In the *super resolution module*, we implement a modified EDSR generator as the backbone for super-resolution. The degradation process is fixed and untrainable, including random fiber point generation, downsampling, noise simulation and reconstruction steps, as described in Section 2.1. Then, to reduce the distribution gap between real pCLE images and histopathological images as well as improving image resolution, we propose an image style adjustment Algorithm 1. The Algorithm 1 is unique from other blind SISR approaches, due to the novel super-resolution generator, degradation module based on random fiber distribution assumption and the whole pseudo-supervised SISR pipeline we propose. With these implementations, the style and quality of all images with this algorithm are more appropriate, and the reliability and effect of super-resolution results can be improved.

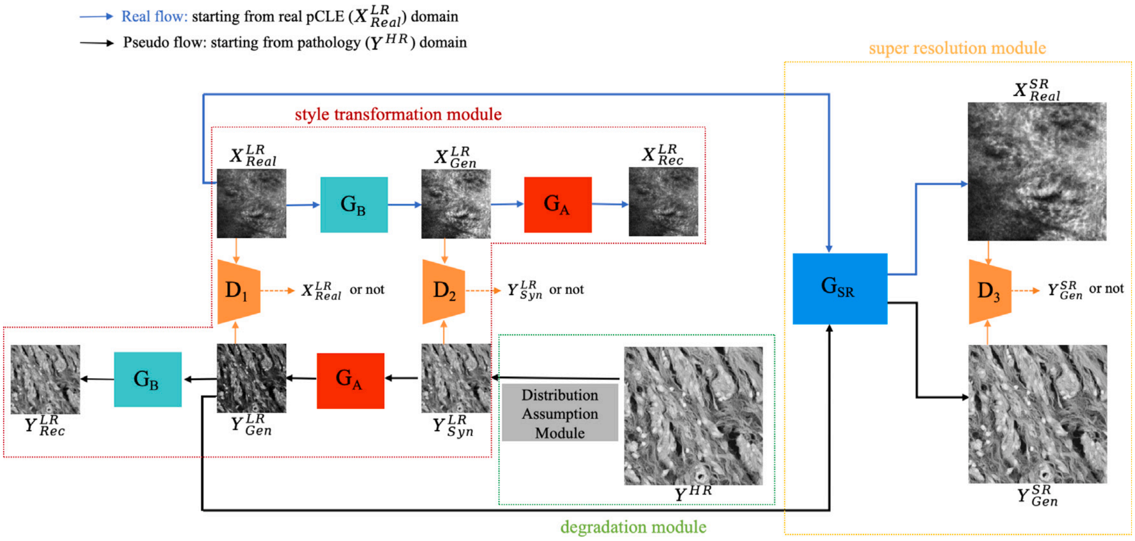

Fig. 1. Schematic illustration of the proposed super-resolution framework: after degrading HR to LR with DAM, in the style transformation module, G_A/G_B along with D_1/D_2 can learn to extract the image-level features of X^{LR}_{Real} and apply them to generate pCLE-like Y^{LR}_{Gen}. Then, in the super resolution module, G_{SR} and D_3 can learn to generate and distinguish images of higher resolution from X^{LR}_{Real} and Y^{LR}_{Gen} in a generative adversarial manner.

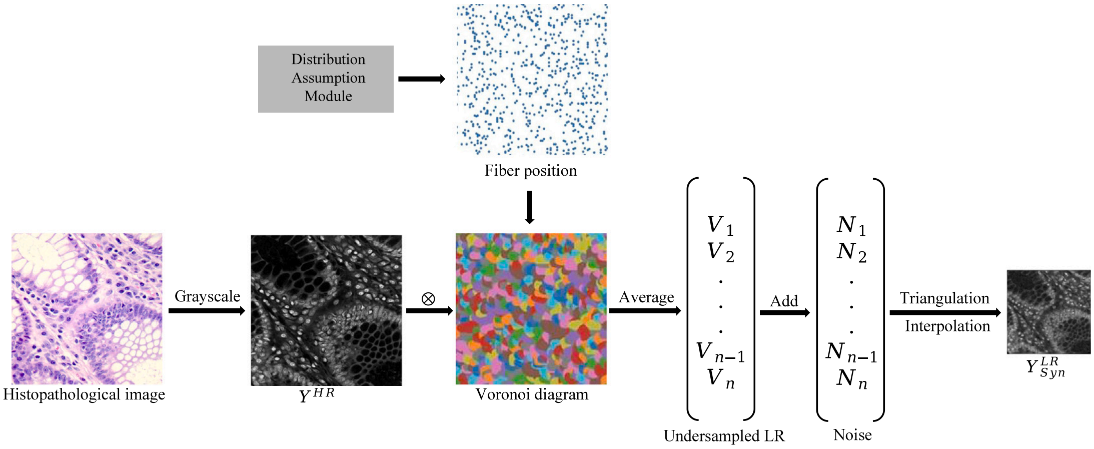

Fig. 2. Distribution Assumption Module based degradation module used in our pipeline.

2.1. Generation of synthetic pCLE images through degradation

Here, with the lack of HR pCLE images, grayscaled histopathological images are regarded as pseudo HR reference (Y^{HR}) since they are not only consistent with pCLE images in providing information of tissues and cells but also of high resolution. Histopathological images are captured using a microscope with a resolution of up to 0.25 μm, allowing for precise demonstration of the microstructural information of tissues. Due to physical constraints induced by the fiber bundle used for signal acquisition, the resolution of pCLE images normally ranges from 1 to 3 μm, leading to pCLE images being more blurred. Meanwhile, due to the differences in staining methods, pCLE images and histopathological images exhibit different contrasts when presenting tissue structures. Histopathological images use HE (Hematoxylin and Eosin) staining, which clearly displays the fine structures of tissues and the relationships between cells, providing more depth and clarity in detail. In contrast, pCLE images typically use fluorescent dyes for labeling, which bind to specific cellular components, thereby providing different signals during imaging. Due to the scattering of fluorescent signals and the presence of background noise, pCLE images generally exhibit relatively lower contrast. The previous works universally reconstruct synthetic endomicroscopy image (Y^{LR}_{Syn}) from Y^{HR} using a Delaunay-based linear interpolation that interpolates pixels from the fiber centers to a regular grid [10,11]. Thus the features and characteristics of generated Y^{LR}_{Syn} depend on the fiber arrangement to some degree, and the generalization performance of the optimized model is limited. Thus, based on the imaging principle, we propose a novel and generalized Distribution Assumption Module (DAM), through which the fiber coordinates are created randomly. Specifically, DAM includes the following two assumptions: First, there must be a certain physical distance between optical fibers in a bundle. On the generated optical fiber position maps, the pixel value is assigned 1 (positive) for positions where there is a fiber, otherwise it is assigned 0 (negative), and all positive positions should be discrete; Second, the variation in distance between fibers is limited, especially in a certain adjacent area of each fiber, where the fiber distances are nearly the same. On the basis of these assumptions, we generate the optical fiber position maps in a randomized way, so that the result can simulate the real fiber distribution as much as possible.

After determining the positions of the fibers, the next step involves computing the Voronoi diagram. This diagram divides the plane into regions, known as Voronoi cells, where each cell corresponds to a specific fiber and represents the fiber's FoV and its contribution to the image pixels [23]. Subsequently, the pixels within the patch belonging to the same Voronoi cell are averaged and make up an undersampled LR. This region-based averaging approach simulates the information

Algorithm 1 Image style adjustment algorithm

```
Require: Real pCLE image set $X_{Real}^{LR}$, synthetic endomicroscopy image set $Y_{Syn}^{LR}$, the number of iterations per epoch $n_{iteration}$, batch size $m$, the number of epoch $n_{epoch}$ and the weight of loss $\alpha, \beta, \gamma$.
Output: Generator $G_i$, total loss $L_G$ of generators and updated generator parameters $\theta_{G_i}$, i=A,B,SR. Discriminator $D_i$, discriminator loss $L_{D_i}$ and updated discriminator parameters $\theta_{D_i}$, i=1,2,3.
for $i$=1, 2, 3, ..., $n_{epoch}$ do
    for $j$=1, 2, 3, ..., $n_{iteration}$ do
        Sample $m$ real pCLE images $x^{(1)}$, ..., $x^{(m)}$ from $X_{Real}^{LR}$
        Obtain generated data $\widehat{x}^{(1)}$, ..., $\widehat{x}^{(m)}$ as $X_{Gen}^{LR}$, $\widehat{x}^{(m)} = G_B(x^{(m)})$
        Obtain generated data $\widetilde{x}^{(1)}$, ..., $\widetilde{x}^{(m)}$ as $X_{Rec}^{LR}$, $\widetilde{x}^{(m)} = G_A(\widehat{x}^{(m)})$
        Sample $m$ synthetic endomicroscopy images $y^{(1)}$, ..., $y^{(m)}$ from $Y_{Syn}^{LR}$
        Obtain generated data $\widehat{y}^{(1)}$, ..., $\widehat{y}^{(m)}$ as $Y_{Gen}^{LR}$, $\widehat{y}^{(m)} = G_A(y^{(m)})$
        Obtain generated data $\widetilde{y}^{(1)}$, ..., $\widetilde{y}^{(m)}$ as $Y_{Rec}^{LR}$, $\widetilde{y}^{(m)} = G_B(\widehat{y}^{(m)})$
        Sample $m$ real pCLE images $x^{(1)}$, ..., $x^{(m)}$ from $X_{Real}^{LR}$ and $m$ generated data $\widehat{y}^{(1)}$, ..., $\widehat{y}^{(m)}$ from $Y_{Gen}^{LR}$
        Obtain generated data $\overline{x}^{(1)}$, ..., $\overline{x}^{(m)}$ as $X_{Real}^{SR}$, $\overline{x}^{(m)} = G_{SR}(x^{(m)})$
        Obtain generated data $\overline{y}^{(1)}$, ..., $\overline{y}^{(m)}$ as $Y_{Gen}^{SR}$, $\overline{y}^{(m)} = G_{SR}(\widehat{y}^{(m)})$
        minimizing $L_G$
        Update generator parameters $\theta_{G_i}$, i=A,B,SR
    end for
    Sample $m$ real pCLE images $x^{(1)}$, ..., $x^{(m)}$ from $X_{Real}^{LR}$ and $m$ generated data $\widehat{y}^{(1)}$, ..., $\widehat{y}^{(m)}$ from $Y_{Gen}^{LR}$
    $D_1(x^{(1)}, \widehat{y}^{(1)})$, ..., $D_1(x^{(m)}, \widehat{y}^{(m)})$
    Sample $m$ generated data $\widehat{x}^{(1)}$, ..., $\widehat{x}^{(m)}$ from $X_{Gen}^{LR}$ and $m$ generated data $y^{(1)}$, ..., $y^{(m)}$ from $Y_{Syn}^{LR}$
    $D_2(\widehat{x}^{(1)}, y^{(1)})$, ..., $D_2(\widehat{x}^{(m)}, y^{(m)})$
    Sample $m$ generated data $\overline{x}^{(1)}$, ..., $\overline{x}^{(m)}$ from $X_{Real}^{SR}$ and $m$ generated data $\overline{y}^{(1)}$, ..., $\overline{y}^{(m)}$ from $Y_{Gen}^{SR}$
    $D_3(\overline{x}^{(1)}, \overline{y}^{(1)})$, ..., $D_3(\overline{x}^{(m)}, \overline{y}^{(m)})$
    minimizing $L_{D_i}$, i=1,2,3
    Update discriminator parameters $\theta_{D_i}$, i=1,2,3
end for
```

collected by each optical fiber from a surrounding area, taking into account the different collection ranges of each fiber [10]. Following that, all the elements in the resulting vector are normalized to the range of [0, 1] for further processing [10].

We also introduce a certain level of additive and multiplicative Gaussian noise to simulate the noise in the acquisition process of real pCLE. Then, consistent with the reconstruction algorithm of pCLE, the sparse image to be interpolated is reconstructed by using Delaunay triangulation and linear interpolation. Finally, we get Y_{Syn}^{LR} as the outcome of the whole degradation module.

2.2. Style transformation and LR-HR pairs construction

Due to the inherent differences between histopathological images and CLE images, as well as the imperfection of the simulated degradation process, the interior features of Y_{Syn}^{LR} still vary from real pCLE images to some extent. Therefore, we propose a style transformation module to fine-tune the degraded image and generate quasi-realistic pCLE image (Y_{Gen}^{LR}), which is more similar to real pCLE image than Y_{Syn}^{LR} in characteristics like structure and brightness. Y_{Gen}^{LR} is trained to match Y^{HR} and then LR-HR pairs are synthesized. The method is inspired by [13,28], but one key difference is that in our proposed pipeline, we attach more significance to maintaining the cycle consistency in order to make Y_{Gen}^{LR} highly realistic. It is worth mentioning that the physical constraints introduced in the degradation module are essential. If the histopathological images are processed directly with the style conversion module, the features of the simulated images obtained are quite different from those of the real pCLE, and ultimately affect the superresolution results.

The style transformation module is based on CycleGAN framework as shown in Fig. 1. Specifically, generator G_A transform the original image to a more pCLE-like one and G_B has the opposite effect, i.e. $G_B(G_A(Y_{Syn}^{LR})) = Y_{Rec}^{LR} \approx Y_{Syn}^{LR}$. Furthermore, the training of the generators requires discriminators D_1 and D_2 that detect translated samples from the real examples, i.e. $D_1(X_{Real}^{LR}, Y_{Syn}^{LR}) \in [0, 1]$.

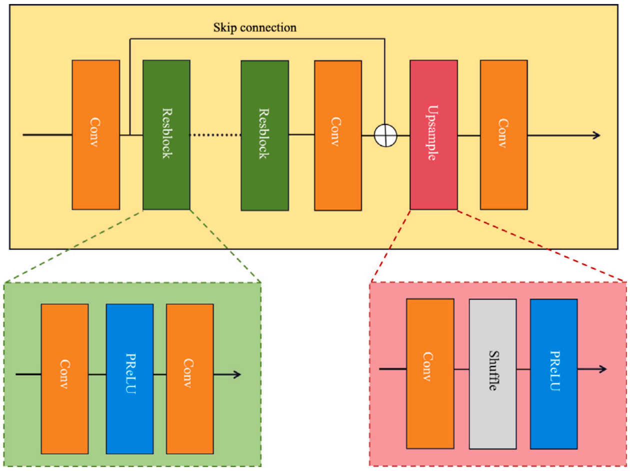

Fig. 3. The architecture of our proposed generator in super resolution module.

The generators G_A and G_B implemented in this paper are based on the 6 blocks ResnetGenerator [28], which keeps two stride-2 convolutional layers for downsampling, 6 residual blocks for 256 × 256 images and two upsampling convolutions with stride 2. The filter kernel size of each convolutional layer and residual block is 3 × 3. Each residual block includes two convolutional layer, and we choose ReLU and instance normalization as the activation function and normalization layer, respectively. Besides, we implement skip connections in the blocks to keep gradients stable and reduce the difficulty in the training stage.

As for D_1 and D_2, we construct PatchGAN classifier [29] to distinguish whether the input is real X_{Real}^{LR} and real Y_{Syn}^{LR} separately, judging the degree of realism by overlapping patches with the size of 70 × 70. The patch-level discriminator consists of 5 standard convolutional layers with a 4 × 4 kernel size, and the number of filters in the first four layers are 64, 128, 256 and 512 respectively. These layers use leakyReLU with a negative slope of 0.2 as the activation function. This architecture has fewer parameters compared to a discriminator that judges the entire image directly and can process arbitrarily-sized images in a fully convolutional manner.

2.3. Super resolution of real and synthetic pCLE images

By feeding the LR-HR pairs into our super resolution module, we get Y_{Gen}^{SR} and X_{Real}^{SR} respectively and use D_3 to identify the authenticity of images. With loss functions to restrict the difference between Y_{Gen}^{SR} and Y^{HR}, G_{SR} is trained to generate X_{Real}^{SR} resembling Y^{HR} in an essential degree. Due to the considerable resolution difference between X_{Real}^{LR} and Y^{HR}, the module can serve as an effective methodology to improve the resolution of X_{Real}^{LR}.

As shown in Fig. 3, a generator G_{SR} based on EDSR [30] is constructed. Our modified EDSR generator consists of 14 Residual blocks and an upsampling block. The Residual block is the stack of convolution layers and PReLU layers with skip connections. In each Residual block, we remove the batch normalization layers as presented in [30]. Since batch normalization layers normalize the features, they get rid of range flexibility. The upsampling block is a stack of convolution and pixel-shuffle layers, which converts the feature maps back to high-resolution image. And the filter kernel size of every convolutional layer in this generator is 3 × 3. Moreover, discriminator implemented in super resolution is completely the same as D_1 and D_2, identifying the authenticity between X_{Real}^{SR} and Y_{Gen}^{SR}.

2.4. Loss functions

We implement multiple basic loss functions for training, including generator loss, discriminator loss, cycle consistency loss, identity loss and super-resolution loss.

2.4.1. Generator loss

Here, we combine L^{G}_{GAN} and L_{SSIM} to form L_{G_i}, i=A,B,SR. With $L^{G}_{GAN}(z)$, where z is typically the generated data through G_i, it is trained to be gradually more difficult to identify z from the original data. And $L_{SSIM}(x, y)$ measures the structural similarity index between x and y, where μ and σ are their mean and standard deviation, respectively. A weight coefficient λ is utilized for adjusting the proportion of attention to structural similarity when producing images and is set as 10, keeping consistent with the parameter in CycleGAN.

$$
\begin{aligned}
&L^{G}_{GAN}(z) = E_{z\sim P_{\text{generated}}}[\log(1 - D(G(z)))],\\
&L_{SSIM}(x, y) = 1 - \frac{1}{N}\sum_{p\in P}\left(\frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}\right)\left(\frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}\right),\\
&L_{G_A} = L^{G}_{GAN}(X^{LR}_{\text{Gen}}) + \lambda L_{SSIM}(X^{LR}_{\text{Real}}, X^{LR}_{\text{Gen}}),\\
&L_{G_B} = L^{G}_{GAN}(Y^{LR}_{\text{Gen}}) + \lambda L_{SSIM}(Y^{LR}_{\text{Syn}}, Y^{LR}_{\text{Gen}}),\\
&L_{G_{SR}} = L^{G}_{GAN}(Y^{SR}_{\text{Gen}}) + \lambda L_{SSIM}(Y^{SR}_{\text{Gen}}, Y^{LR}_{\text{Gen}}).
\end{aligned}
\tag{1}
$$

2.4.2. Discriminator loss

$L_{D_i}(i = 1, 2, 3)$ is used for enhancing the discriminator's ability to discern between real and fake data. Here, x is typically the origin data and z is the prediction data obtained from G_i.

$$
\begin{aligned}
&L^{D}_{GAN}(x, z) = E_{x\sim P_{origin}}[log D(x)] + E_{z\sim P_{prediction}}[log(1 - D(G(z)))],\\
&L_{D_1} = L^{D}_{GAN}(X^{LR}_{Real}, Y^{LR}_{Gen}), L_{D_2} = L^{D}_{GAN}(Y^{LR}_{Syn}, X^{LR}_{Gen}),\\
&L_{D_3} = L^{D}_{GAN}(X^{SR}_{Real}, Y^{SR}_{Gen}).
\end{aligned}
\tag{2}
$$

2.4.3. Cycle consistency loss

The calculation of cycle consistency loss $L_{Cycle1}(x)$ and $L_{Cycle2}(x)$ involves converting the source domain data x to the target domain and then back to the source domain data y [28]. This process includes two opposite-direction cycles and aims to preserve most of the image features.

$$
\begin{aligned}
&L_{Cycle} = L_{Cycle1}(X^{LR}_{Real}) + L_{Cycle2}(Y^{LR}_{Syn})\\
&= E_{x\sim P_{source1}}\|G_A(G_B(X^{LR}_{Real})) - X^{LR}_{Real}\|_1\\
&+ E_{x\sim P_{source2}}\|G_B(G_A(Y^{LR}_{Syn})) - Y^{LR}_{Syn}\|_1.
\end{aligned}
\tag{3}
$$

2.4.4. Identity loss

$L_{Idt1}(x)$ and $L_{Idt2}(x)$ are employed to maintain consistency in image style and avoid over-mapping. L_{Idt} as the combination of two identity losses can gauge the variation in non-target areas and exercise control over them.

$$
\begin{aligned}
&L_{Idt} = L_{Idt1}(X^{LR}_{Real}) + L_{Idt2}(Y^{LR}_{Syn})\\
&= E_{x\sim P_{source1}}\|G_A(X^{LR}_{Real}) - X^{LR}_{Real}\|_1 + E_{x\sim P_{source2}}\|G_B(Y^{LR}_{Syn}) - Y^{LR}_{Syn}\|_1.
\end{aligned}
\tag{4}
$$

2.4.5. Super-resolution loss

Here, to constrain the function of G_{SR} into an appropriate one, we implement super-resolution loss L_{SR} composed of two functions: $L_{Perceptual}(x, y)$ emphasizing deep features and perceptual effects contained in the image [31]; $L_{TextureMatching}(x, y)$ enhancing the generated images to have richer and consistent texture information on a local level [32]. For both loss functions, we use a pre-trained VGG-16 network as the feature extraction function ϕ:

$$
\begin{aligned}
&L_{Perceptual}(x, y) = E_{x\sim P_{source1}, y\sim P_{source2}}\|x - y\|_2^2\\
&\qquad + 0.006E_{x\sim P_{source1}, y\sim P_{source2}}\|\phi(x) - \phi(y)\|_2^2,\\
&L_{TextureMatching}(x, y) = E_{x\sim P_{source1}, y\sim P_{source2}}\|\mathbf{G}(\phi(x)) - \mathbf{G}(\phi(y))\|_2^2,\\
&L_{SR} = L_{Perceptual}(Y^{SR}_{Gen}, Y^{HR}) + 0.006L_{TextureMatching}(Y^{SR}_{Gen}, Y^{HR}).
\end{aligned}
\tag{5}
$$

where G is the Gram matrix used to extract texture features.

2.4.6. Total loss

The loss functions relative to generators are accumulated as

$$
L_G = L_{G_A} + L_{G_B} + L_{G_{SR}} + \alpha L_{Cycle} + \beta L_{Idt} + \gamma L_{SR}, \tag{6}
$$

where α, β, γ are employed to adjust the weight of corresponding loss so that the whole framework can reach an optimal performance. Along with the L_G we also have to minimize the loss of discriminators

$$
L_D = L_{D_i}, i = 1, 2, 3. \tag{7}
$$

Compared with normal standard loss functions, first we implement an additional identity loss to strengthen the control over two generators, so that the image style can maintain consistency and avoid over-mapping in the transformation module. Another significant improvement is that in the super-resolution loss, we introduce an integration of perceptual loss and texture matching loss, which is innovative compared with other similar loss functions. Considering the resolution, contrast and texture difference between pCLE images and histopathological images, perceptual loss and texture matching loss are capable of assisting in reducing their disparity effectively. It is also examined that with the special loss functions above, the reliability and effect of super-resolution results can be improved compared with standard loss functions.

3. Results

3.1. Datasets and experiment settings

Y^{HR} **Dataset:** Since both histopathological images and pCLE images are microscopy visualizations of cells, they possess strong similarities in low-level features [11]. From the lung and colon histopathological image dataset (LC25000) [33], which consists of 25000 color images of colon adenocarcinoma, benign colonic tissue, lung adenocarcinoma, lung squamous cell carcinoma, and benign lung tissue (each 768 × 768 in size), we randomly select a total of 550 images. After randomly choosing a starting point, we crop a quarter of each image, apply color inversion, and resize them to 512 × 512, resulting in the transformed images from the LC25000 dataset labeled as Y^{HR}.

X^{LR}_{Real} **Dataset:** We collect 550 pCLE frames captured on the colon regions (each 1024 × 1024 in size), which were generated from different patients in actual clinical use with a CLE-1000 equipment (laser scanning unit LSU-1000, confocal probe U-300) from BIOPSEE Medical Technology Co., Ltd. In addition, we validate the proposed approach on an external pCLE dataset from Cellvizio (Mauna Kea Technologies, Paris, France)'s 'Screening and Diagnosis of esophageal cancer from in-vivo microscopy images' challenge data. The challenge data comprises 11161 images acquired from 61 patients (each 521 × 519 in size), among which we choose 550 images of intestinal and gastric metaplasia. Due to smaller field of view (FoV) of pCLE system, single image in this dataset contains less types but larger parts of micro structures than images in the internal dataset. To convert the images in these two datasets into X^{LR}_{Real}, we process them through cropping (into 795 × 795 and 380 × 380 in size separately), random rotations (0 to 360 degree), and resizing them to 256 × 256. Then we split the pCLE data randomly into a training set and a test set with a ratio of 9:1.

Implementation Details: Synthetic images Y^{LR}_{Syn} are first generated before training. During training, X^{LR}_{Real} and Y^{LR}_{Syn} are fed into the style transformation and super resolution modules together. All training runs were based on the Pytorch (1.13.0) framework. The parameters of the functions are randomly initialized and updated by Adam optimizer with a batch size of 8. The default scale factor of super resolution is set to 2 in our experiments, and the size of inputs is 256 × 256 while that of Y^{HR} is 512 × 512. Considering the weight assigned to each component of the loss function in Eq. (6), parameters α, β and γ were preliminarily tested, selected from the set $\{0, 5, 10\}$ during the experiment. And the optimal combination is employed, which is $\alpha = 10$, $\beta = 5$ and $\gamma = 10$. We set the learning rate to $1e-4$ empirically for training $G_i(i = A, B, SR)$ and $D_i(i = 1, 2, 3)$. In the test stage, we feed the X^{LR}_{Real} and Y^{LR}_{Syn} as test images into G_{SR} for super-resolution.

3.2. Comparison with state-of-the-art methods

Details of the compared methods and analyses: In this section, we conduct detailed comparisons and analysis of the proposed framework and other five unsupervised SISR methods (DASR [14], CycleSR [13], Ravìet al. [10], ZSSR [23] and SimUSR [34]). The former four pipelines are the same as depicted in Section "Introduction". And the only difference between SimUSR and ZSSR is that ZSSR uses a single test image for online optimization while SimUSR leverages plenty of external images for offline updates. The validation of the methods is mainly based on complementary quantitative analysis and qualitative analysis. Here, Y^{SR}_{Gen} images of different pipelines are generated from the same Y^{LR}_{Syn} dataset through our modified EDSR generator with various corresponding well-trained parameters when testing. So the performance of synthetic pCLE images domain can reflect the effect of SISR methods and keep a consistency with real pCLE images domain.

Quantitative evaluation: Focusing on synthetic pCLE images domain, Four metrics are utilized in our quantitative evaluation experiments: (i) Structural similarity index (SSIM) that evaluates the similarity between SR and HR designed by modeling image distortion [35], (ii) $\Delta\mathrm{GCF}_{\mathrm{HR}}$ that quantifies the improvement on global contrast factor (GCF) that the SR image yields with respect to the initial HR [10,36], (iii) a composite score $\mathrm{Tot}_{\mathrm{cs}}$ obtained by normalizing the value of SSIM and $\Delta\mathrm{GCF}_{\mathrm{HR}}$ to the range [0,1] and averaging the obtained results [10]. This composite score leads to a more robust evaluation of the results since, SSIM alone is not reliable when the ground truth is only estimated, while the GCF can be improved by merely adding random high frequency to the images, (iv) LPIPS (learned perceptual image patch similarity) that extracts image features and indicates perceptual distance between images, in accordance with human perception [37]. Therefore, as Eq. (11) describes, the less the figure of LPIPS, the better perceptual image patch similarity.

$$\mathrm{SSIM} = \frac{(2\mu_G\mu_I + c_1)(2\sigma_{GI} + c_2)}{(\mu_G^2 + \mu_I^2 + c_1)(\sigma_G^2 + \sigma_I^2 + c_2)}, \tag{8}$$

$$\mathrm{GCF} = \frac{1}{N}\sum_{i=1}^{N}\frac{|L_i - L_{i-1}| + |L_i - L_{i+1}| + |L_i - L_{i-w}| + |L_i - L_{i+w}|}{4}, \tag{9}$$
$$\Delta\mathrm{GCF}_{\mathrm{HR}} = GCF(I) - GCF(G),$$

$$\mathrm{Tot}_{\mathrm{cs}} = \frac{\frac{SSIM-0.6}{0.4} + \frac{\Delta GCF_{HR}+0.5}{1.8}}{2}, \tag{10}$$

$$\mathrm{LPIPS} = \sum_l \frac{1}{H_l \times W_l}\sum_{h,w}\|\omega_l \cdot (y^l_{hw} - y^l_{0hw})\|_2^2, \tag{11}$$

where μ_G and σ_G are the mean and the standard deviation of the ground truth image G, while μ_I and σ_I are the mean and the standard deviation of the SR image I, respectively. σ_{GI} is the covariance of G and I. c_1 and c_2 are set to 0.01^2 and 0.03^2, respectively. N is the total number of pixels. L is the pixel value. y^l_{hw} and y^l_{0hw} represent features of generated and real images at the lth layer output, whose weight is ω and scale is $H_l \times W_l$.

Furthermore, we choose four other metrics measuring the contrast difference between images, which are adapted to the real pCLE images domain as ground truth is not required: (i) SNR (signal-to-noise ratio) that measures ratio of useful signal to noise strength in a processed image [38]; (ii) $\Delta\mathrm{GCF}_{\mathrm{LR}}$ that quantifies the improvement on the global contrast factor that the SR image yields with respect to the initial LR [10,36]; (iii) GLCM (gray-level co-occurrence matrix) that evaluates similarity of the color position distribution and textural features in images by calculating maximal correlation coefficient of the co-occurrence matrix of grayscale images [39]; and (iv) ENL (equivalent numbers of looks) that is commonly used to measure the image smoothness especially after denoising. A larger ENL value indicates that the image is smoothed better [40].

$$\mathrm{SNR} = 10\log_{10}(\max(I^2)/\sigma_n^2) \tag{12}$$

$$\Delta\mathrm{GCF}_{\mathrm{LR}} = GCF(I) - GCF(I_0) \tag{13}$$

$$\mathrm{GLCM} = \frac{\sum_{I,I_0}(M_1(I) - \frac{1}{N}\sum_i M_1(i))\cdot(M_2(I_0) - \frac{1}{N}\sum_i M_2(i))}{\sqrt{\sum_I(M_1(I) - \frac{1}{N}\sum_i M_1(i))^2 \cdot \sum_{I_0}(M_2(I_0) - \frac{1}{N}\sum_i M_2(i))^2}} \tag{14}$$

$$\mathrm{ENL} = \frac{1}{L}\sum_{l=1}^{R}\frac{\mu_l^2}{\sigma_n^2} \tag{15}$$

where I is the super-resolution image concerning its original image I_0. N is the total number of pixels and M is the gray-level co-occurrence matrix of an image. μ_l is the mean of the lth homogeneous regions of interest (ROI) and σ_n is the variance of the background regions.

Based on those quantitative metrics, we compare the performance of the proposed framework with state-of-the-art blind SISR methods on our datasets and present the results in Table 1. As we can see, the proposed framework outperforms the other methods in two domains (synthetic pCLE images and real pCLE images), showing the best results on almost all metrics. It is worth mentioning that our pipeline not only attains significant results in the non-sensory evaluations methods but also works well in perceptual evaluation (LPIPS), which means the outcome is more conducive to adapting to human observation in clinical practice.

Qualitative analysis: As exhibited in Figs. 4 and 5, our proposed framework can restore edge and texture details of images more realistically, clearly and smoothly compared with other pipelines, both on synthetic pCLE and real pCLE images. In particular, we could observe visual distortions such as edge blurring and internal stripe missing in some cases constructed by other methods. However, in our results, the edges are sharpened with better contrast, and those small details are enhanced, thus the image resolution is promoted while the consistency of the structures compared to the original image is maintained. The yellow arrows in Fig. 4 indicate the super-resolution effects of different methods on the edges. These results demonstrate the effectiveness of our proposed pipeline in pCLE image quality enhancement.

3.3. Comparison of loss weights

To determine the optimal weight coefficients for the loss functions, we empirically select α, β and γ from the set $\{0, 5, 10\}$ to compare their corresponding results. Focusing on synthetic pCLE images domain, the increase of α and β enhances the structural metric value and perceptual image patch similarity. Conversely, an increase in γ improves ΔGCF_{HR} score but leads to a substantial reduction in $SSIM$. According to the results presented in Table 2, the combination of $\alpha = 10$, $\beta = 5$ and $\gamma = 10$ exhibits the best balanced performance in reinforcing the texture and edge details of images, simultaneously minimizing the sacrifice of pixel information.

3.4. Comparison of different model settings

Table 3 demonstrates the quantitative performance of our proposed framework under various generator settings in synthetic pCLE images domain. The experiments compare the impact of generators type and number of residual blocks on the super-resolution outcomes. Basically, the increase of residual block numbers leads to deeper information excavation of images and better performance, which however has different effects when combined with various networks. Results show that the combination of 6 ResnetGenerator blocks as G_A/G_B and modified EDSR generator as G_{SR} results in the most outstanding enhancement across evaluation metrics in our experiments.

Table 1
Quantitative comparison of super-resolution effect with other unsupervised SISR methods. Note that in this paper the best results are displayed in bold.

Domain		Proposed	DASR	CycleSR	SimUSR	Ravìet al	ZSSR
Synthetic pCLE images	SSIM ↑	**0.7287 ± 0.0849**	0.7143 ± 0.0628	0.7140 ± 0.0661	0.6658 ± 0.0830	0.7006 ± 0.0995	0.6911 ± 0.1194
	$\Delta\text{GCF}_{\text{HR}}$ ↑	**1.5911 ± 0.5090**	1.5478 ± 0.4532	1.5325 ± 0.3902	0.6123 ± 0.5571	0.4413 ± 0.2478	0.4879 ± 0.4623
	Tot_{cs} ↑	**0.7417**	0.7117	0.7071	0.3912	0.3872	0.3883
	LPIPS ↓	0.1844 ± 0.0693	0.1524 ± 0.0104	**0.1142 ± 0.0273**	0.4530 ± 0.0446	0.3866 ± 0.0538	0.3712 ± 0.0237
Real pCLE images	SNR ↑	**28.0562 ± 3.8291**	24.3832 ± 3.9123	24.2390 ± 3.3298	21.4711 ± 2.1450	23.2593 ± 5.0063	25.4957 ± 2.9547
	$\Delta\text{GCF}_{\text{LR}}$ ↑	1.0036 ± 0.4820	**2.7293 ± 0.3125**	0.7943 ± 0.3137	2.6011 ± 0.4698	1.3110 ± 0.1378	1.3490 ± 0.2319
	GLCM ↑	**0.7859 ± 0.0810**	0.7414 ± 0.0347	0.6547 ± 0.0544	0.4342 ± 0.0238	0.6191 ± 0.0325	0.3125 ± 0.0423
	ENL ↑	**49.3946 ± 1.6309**	43.4182 ± 3.4289	44.5372 ± 1.3982	37.0702 ± 6.7210	21.2989 ± 1.0826	25.7816 ± 3.6301

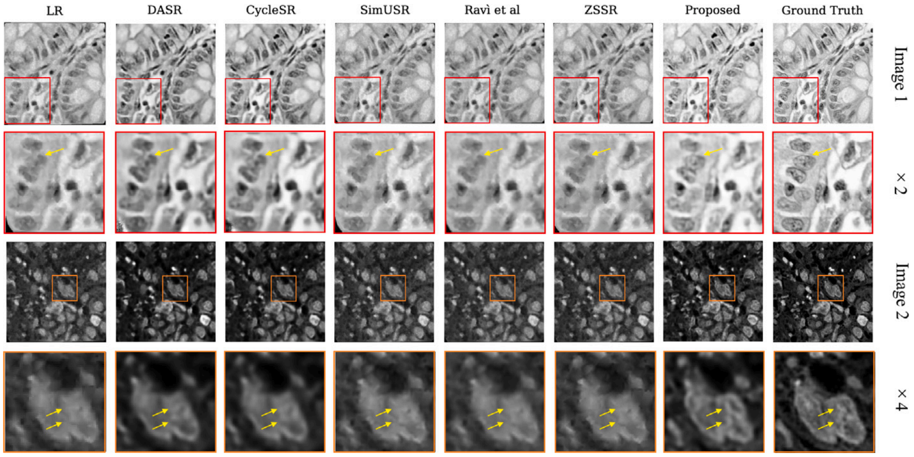

Fig. 4. Example of visual results obtained by our approach in comparison with other state-of-the-art approaches in synthetic pCLE images domain.

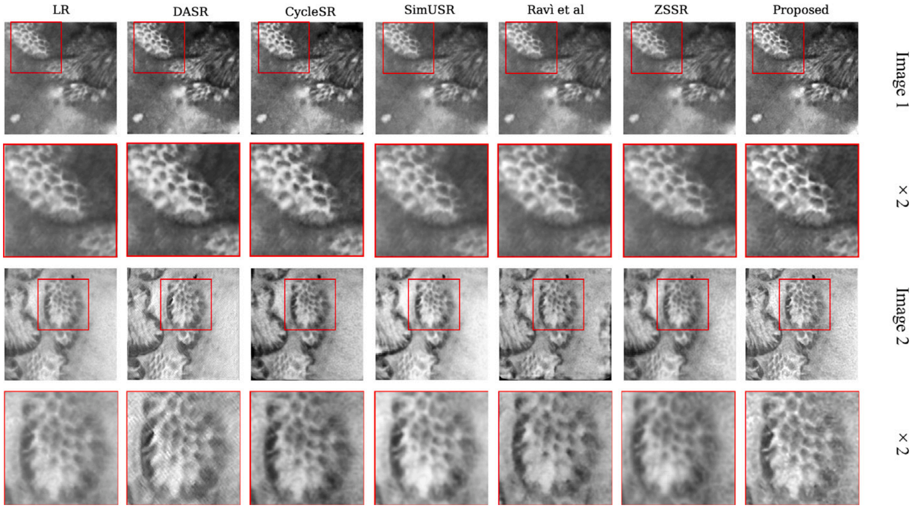

Fig. 5. Example of visual results obtained by our approach in comparison with other state-of-the-art approaches in real pCLE images domain (no ground truth is available).

Table 2
Quantitative comparison of various loss weights combinations.

α	β	γ	SSIM	$\Delta\text{GCF}_{\text{HR}}$	Tot_{cs}	LPIPS
0	0	10	0.6704 ± 0.1364	1.4270 ± 0.4102	0.6233	0.4318 ± 0.0533
0	5	10	0.7252 ± 0.1183	0.9663 ± 0.3830	0.5638	0.3546 ± 0.0435
5	0	10	0.7122 ± 0.0817	0.8233 ± 0.1072	0.5078	0.4149 ± 0.0334
5	5	10	**0.7394 ± 0.0929**	1.5337 ± 0.4314	0.7392	0.3674 ± 0.0368
5	10	10	0.7348 ± 0.0920	1.4908 ± 0.5585	0.7215	0.3533 ± 0.0418
10	10	10	0.7375 ± 0.0969	1.4207 ± 0.3864	0.7054	0.3517 ± 0.0407
10	10	5	0.7275 ± 0.1123	1.5861 ± 0.3129	0.7388	0.3542 ± 0.0425
10	5	5	0.7324 ± 0.1010	1.4094 ± 0.4098	0.6959	0.2625 ± 0.0426
10	5	10	0.7287 ± 0.0849	**1.5911 ± 0.5090**	**0.7417**	**0.1844 ± 0.0693**

Table 3
Quantitative comparison of different model settings.

G_A/G_B	G_{SR}	SSIM	ΔGCF$_{\text{HR}}$	Tot$_{\text{cs}}$	LPIPS
6 blocks Resnet	6 blocks Resnet	0.7380 ± 0.1095	0.7990 ± 0.3774	0.5333	0.3332 ± 0.0506
6 blocks Resnet	9 blocks Resnet	0.7151 ± 0.0959	1.4593 ± 0.5336	0.6881	0.3563 ± 0.0544
9 blocks Resnet	9 blocks Resnet	**0.7391 ± 0.1162**	1.3631 ± 0.4516	0.6914	0.3392 ± 0.0497
9 blocks Resnet	modified EDSR	0.7306 ± 0.1091	1.4019 ± 0.3271	0.6916	0.2436 ± 0.0408
6 blocks Resnet	modified EDSR	0.7287 ± 0.0849	**1.5911 ± 0.5090**	**0.7417**	**0.1844 ± 0.0693**

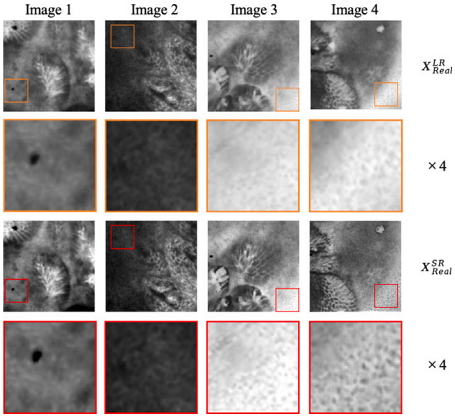

Fig. 6. Examples of real pCLE images domain visual results obtained by the proposed approach. Focusing on background noise, image 1 and 2 possess the dark backgrounds like most situations while image 3 and 4 are typical of predominantly white backgrounds.

3.5. The effect of color depth on background noise reduction

For pCLE, each fiber functions as a pixel detector, capturing and transmitting pixel information within the image. These fibers are then combined into a unified bundle. During the imaging period, noise from pCLE equipment is unavoidable. So to generate more realistic Y^{LR}_{Syn} from Y^{HR}, the noise intensity we set approximates the noise level in X^{LR}_{Real}. However, through our previous experiments (not shown here), the output images indicate that the background noise might be magnified. To tackle this problem, as [32] suggests, we add perceptual loss and texture matching loss simultaneously to restrict SR generator, and then background noise is controlled to a large extent generally. But part of noise within a predominantly white background is not easily reducible compared to normally dark backgrounds, as depicted in Fig. 6. The reason might be that noise in these areas is considered as the structural information to be enhanced.

3.6. Ablation study

In this section, at first we construct real fiber position-based degradation module (RFP) [10] and downscale-based degradation module (DS) [11], as shown in Fig. 7. RFP keeps consistent with our proposed module except for utilizing real fiber distribution to form Voronoi diagram. And DS consists of grayscale and downscale steps, in which there is no specific process simulating pCLE acquisition and reconstruction. To verify the performance and necessity of the proposed degradation approach DAM, an ablation study is carried out, in which aforementioned three degradation modules are implemented and followed by the same style transformation and super resolution modules.

Here we evaluate the processing results of real pCLE images in order to better reflect the practical effects brought by the degradation module. According to the results shown in Fig. 8 and Table 4, although real fiber position-based module outperforms in the global contrast improvement, the intense sharpening effect with this module is part of excessive. To be specific, in X^{SR}_{RFP}, background noise regions outside the main structures are also enhanced, leading to artifacts that might affect tissue identification and diagnosis. Additionally, due to an excessive degree of sharpening and contrast-enhancing, the edge of interior stripes and the pixel value distribution are changed, which results in the loss of structural authenticity and continuity. The reason might be the overfitting caused by the lack of real fiber position data. As for the downscale-based module which preserves edge information well, the blurry parts in the original image have not been sufficiently clarified. The performance of X^{SR}_{DS} indicates that pipeline with downscale-based degradation module is limited to a large extent, which acts far from being as good as X^{SR}_{DAM} and X^{SR}_{RFP}. It is examined that simply downscaling and adding noise cannot provide the enough deep-level information and characteristics of pCLE image, causing less-matched LR-HR pairs and limited training process.

Compared with the two commonly used degradation modules, our proposed DAM-based method showcases the best performance to keep the balance between image super-resolution effect and reliability maintenance. The value of SNR and ΔGCF$_{\text{LR}}$ indicates that our method assists in improving resolution in the target area and constraining noise generated simultaneously. Besides, the focused structures have not undergone excessive deformation due to relatively high GLCM and ENL values. The illustration above is also consistent with the details in Fig. 8.

3.7. External validation

Quantitative analysis: Based on those quantitative metrics, we compare the performance of the proposed framework with state-of-the-art blind SISR methods on external dataset and present the results in Table 5. It is examined that with the external data, our proposed framework still outperforms the other methods in two domains, showing the best results on almost all metrics. For real pCLE images, the rankings of our proposed method across four evaluation metrics are consistent on both internal and external datasets, which verifies the versatility and effect of the network as well.

Qualitative analysis: As exhibited in Fig. 9, in the real pCLE image results of our proposed method, the edges are sharpened with better contrast, and the small internal structural stripes are more enhanced compared with other state-of-the-art methods. Consequently, our proposed framework can restore details of images more realistically and clearly. Meanwhile, the visual super-resolution features and effects across internal and external datasets stay the same, validating our network's adaptability and flexibility.

4. Discussion

4.1. Generalization capability of proposed framework

Our proposed framework is generally applicable to pCLE images from different sampling areas. Specifically, in the style transformation module, we can generate pseudo LR-HR pairs of histopathological images. And what LR differs from HR in pairs is merely the resolution decreasement caused by pCLE imaging simulation. So any real pCLE images can be super-resolved regardless of the location of pCLE images sampling. Furthermore, our framework does not rely on real fibers' position of pCLE devices. Consequently, the optimized parameters obtained through training can be applied to pCLE images from distinct equipment, greatly facilitating the completion of super-resolution tasks.

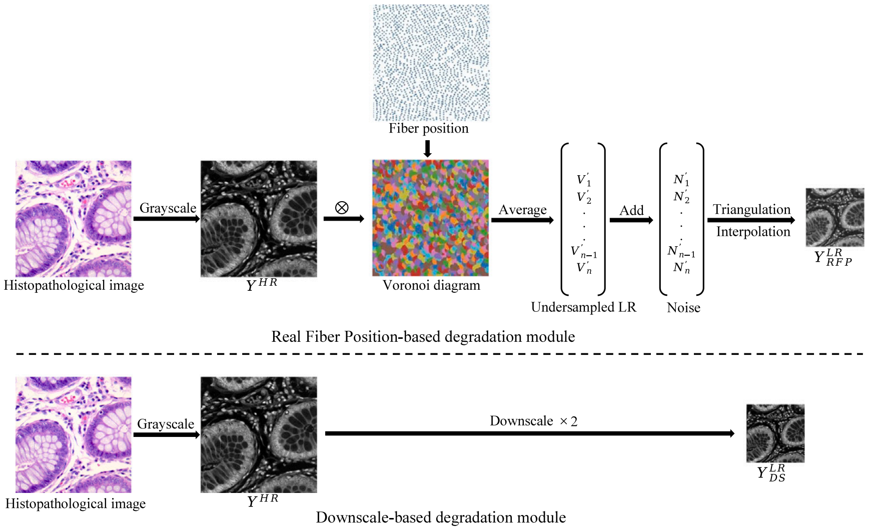

Fig. 7. Real Fiber Position-based degradation module and Downscale-based degradation module used in our ablation study.

Table 4
Quantitative comparison of super-resolution effect with other frameworks in real pCLE images domain, which possess different degradation modules.

Method	SNR	ΔGCF_{LR}	GLCM	ENL
DAM	**28.0562 ± 3.8291**	1.0036 ± 0.4820	**0.7859 ± 0.0810**	**49.3946 ± 1.6309**
RFP	23.4446 ± 3.5113	**1.3425 ± 0.9307**	0.6458 ± 0.0578	35.3273 ± 1.8693
DS	24.9485 ± 3.7904	0.6196 ± 0.5682	0.6283 ± 0.0114	40.5500 ± 1.9451

Table 5
Quantitative comparison of super-resolution effect with other unsupervised SISR methods on external dataset.

Domain		Proposed	DASR	CycleSR	SimUSR	Ravìet al	ZSSR
Synthetic pCLE images	SSIM ↑	**0.7222 ± 0.0183**	0.7198 ± 0.0516	0.7107 ± 0.0302	0.6515 ± 0.0417	0.6810 ± 0.0169	0.6968 ± 0.0540
	ΔGCF_{HR} ↑	1.4536 ± 0.4704	**1.4781 ± 0.5178**	1.4307 ± 0.5305	0.7392 ± 0.5362	0.5064 ± 0.1605	1.2515 ± 0.1956
	Tot_{cs} ↑	0.6954	**0.6992**	0.6747	0.4086	0.3808	0.6075
	LPIPS ↓	**0.3625 ± 0.0519**	0.3729 ± 0.0295	0.3938 ± 0.0326	0.4761 ± 0.0170	0.5159 ± 0.0197	0.4913 ± 0.0593
Real pCLE images	SNR ↑	**30.6370 ± 5.8481**	27.2696 ± 2.3349	23.8345 ± 4.3458	18.7648 ± 5.1956	22.5683 ± 7.9621	24.1526 ± 6.1925
	ΔGCF_{LR} ↑	2.3389 ± 0.1245	2.7560 ± 0.1343	**2.9357 ± 0.3550**	1.5237 ± 0.5438	1.9124 ± 0.2350	2.2938 ± 0.2459
	GLCM ↑	**0.7252 ± 0.0255**	0.6891 ± 0.0213	0.6658 ± 0.0569	0.4358 ± 0.0561	0.5346 ± 0.0264	0.6759 ± 0.0583
	ENL ↑	**36.3020 ± 1.4052**	33.5903 ± 1.4679	26.6648 ± 2.3685	25.5793 ± 5.6924	22.9347 ± 3.1265	24.4070 ± 2.5048

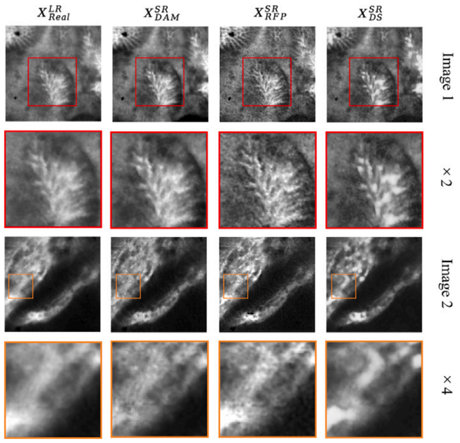

Fig. 8. Examples of real pCLE images domain visual results.

4.2. Limitation and prospect

In the style transformation module, an improved CycleGAN is employed for transferring image styles and generating pseudo LR-HR pairs. This approach demonstrates high efficiency and flexibility during training, making it a commendable choice for image style transfer tasks. While GANs can sometimes present challenges in terms of stability and output control, our proposed algorithm significantly enhances the overall reliability and desirability of the generated images. In the network training process, we use a total loss function which enables the parameters of all generators and discriminators to be optimized at the same time. Although the reliability of super-resolution outcomes increases and more training time can be saved, it is inevitable that training this network requires a larger number of parameters, resulting in greater memory requirements, which imposes higher demands on the GPU used for training and parameters storage. As for the output, while the background does not contain meaningful structural information, we find that noise within a predominantly white background is not easily reducible, which interferes with the partial visual performance. In the future, we will carry out targeted treatment for this phenomenon, so as to improve the noise suppression effects.

5. Conclusion

To enhance the resolution and suppress noise of pCLE images, in this paper we propose a novel unsupervised SISR methodology. Considering the flexibility and applicability for clinical practice, first we investigate an innovative *degradation module* as a preprocess of the task, which makes use of our proposed DAM to randomly generate fibers' position and thus can be easily adapted to imaging probes with

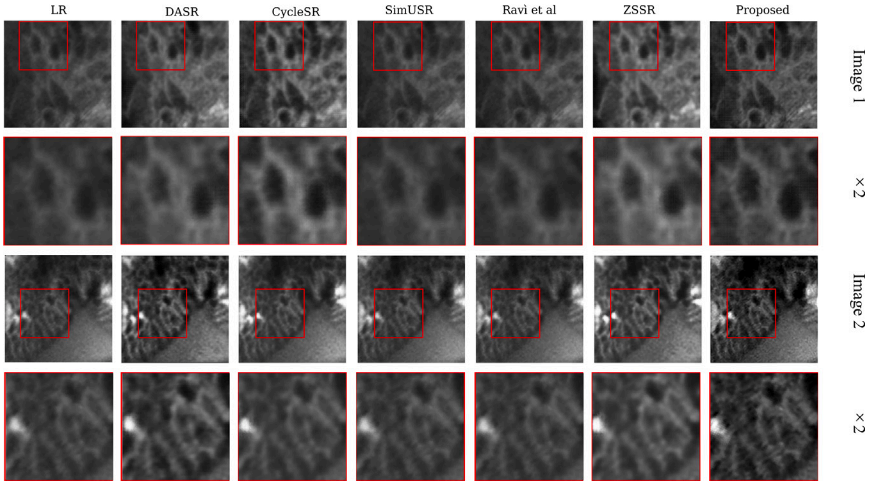

Fig. 9. Example of visual results obtained by our approach in comparison with other state-of-the-art approaches in real pCLE images domain on external dataset (no ground truth is available).

different fiber patterns. In this way, our approach solves the problem of generalizable training with different pCLE probes and unknown real fiber positions. Additionally, we propose an unsupervised deep learning framework available for pCLE image super-resolution, which combines a *style transformation module* with a *super resolution module* and does not require one-to-one alignment between LR and HR. Here, we also introduce a unique integration of loss functions, which assists in increasing the effect and reliability of super-resolution outcomes. Both quantitative and qualitative experiments with internal and external datasets demonstrate the outstanding performance and reliability of the whole pipeline. Besides, focusing on degradation process, we novelly implement a comparison between our proposed method and previous modules. In general, as a time-saving and widely applicable post-processing framework which enhances pCLE image quality with a prominent effect, our proposed methodology has the capability to assist in better real-time and accurate diagnosis of gastrointestinal and other diseases in clinical settings. Simultaneously, we hope that this method will facilitate further research into developing generalized unsupervised learning methodology for SISR tasks.

CRediT authorship contribution statement

Linghao Meng: Writing – review & editing, Writing – original draft, Formal analysis, Data curation. **Yangxi Li:** Writing – review & editing, Formal analysis, Data curation. **Yuchao Zheng:** Formal analysis. **Yu Feng:** Data curation. **Fang Chen:** Writing – review & editing, Investigation, Formal analysis. **Longfei Ma:** Visualization. **Hongen Liao:** Writing – review & editing, Supervision, Investigation, Funding acquisition, Conceptualization.

Ethics statement

The pCLE images were acquired using a device that have already been approved for the market (CLE-1000, BIOPSEE Medical Technology Co. Ltd.). Collection of data from human have been approved by the ethics committee of Beijing Tsinghua Changgung Hospital affiliated to Tsinghua University (No. 24336-0-02) and have the informed consents from patients.

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.

Acknowledgment

The authors acknowledge supports from National Natural Science Foundation of China (grant No. 82027807).

References

[1] N.D. Pilonis, W. Januszewicz, M. di Pietro, Confocal laser endomicroscopy in gastro-intestinal endoscopy: technical aspects and clinical applications, Transl. Gastroenterol. Hepatol. 7 (2022).

[2] R.L. Siegel, K.D. Miller, N.S. Wagle, A. Jemal, Cancer statistics, 2023, CA: Cancer J. Clin. 73 (1) (2023).

[3] V. Pleasant, A. Sammarco, G. Keeney-Bonthrone, S. Bell, R. Saad, M.B. Berger, Use of X-ray to assess fecal loading in patients with gastrointestinal symptoms, Dig. Dis. Sci. 64 (2019) 3589–3595.

[4] A. Dohan, M. Boudiaf, X. Dray, E. Samaha, C. Cellier, M. Camus, C. Eveno, R. Dautry, P. Soyer, Detection of small-bowel tumours with CT enteroclysis using carbon dioxide and virtual enteroscopy: a preliminary study, Eur. Radiol. 28 (2018) 206–213.

[5] C.A. Lamb, N.A. Kennedy, T. Raine, P.A. Hendy, P.J. Smith, J.K. Limdi, B. Hayee, M.C. Lomer, G.C. Parkes, C. Selinger, et al., British society of gastroenterology consensus guidelines on the management of inflammatory bowel disease in adults, Gut 68 (Suppl 3) (2019) s1–s106.

[6] C.S. De Jonge, A.J. Smout, A.J. Nederveen, J. Stoker, Evaluation of gastrointestinal motility with MRI: Advances, challenges and opportunities, Neurogastroenterol Motility 30 (1) (2018) e13257.

[7] S. Giannarou, C. Xu, A. Roddan, Endomicroscopy, in: Biophotonics and Biosensing, Elsevier, 2024, pp. 269–284.

[8] C. Xu, H. Xu, S. Giannarou, Distance regression enhanced with temporal information fusion and adversarial training for robot-assisted endomicroscopy, IEEE Trans. Med. Imaging (2024).

[9] A. Perperidis, K. Dhaliwal, S. McLaughlin, T. Vercauteren, Image computing for fibre-bundle endomicroscopy: A review, Med. Image Anal. 62 (2020) 101620.

[10] D. Ravì, A.B. Szczotka, S.P. Pereira, T. Vercauteren, Adversarial training with cycle consistency for unsupervised super-resolution in endomicroscopy, Med. Image Anal. 53 (2019) 123–131.

[11] C. Zhang, Y. Gu, G.-Z. Yang, Contrastive adversarial learning for endomicroscopy imaging super-resolution, IEEE J. Biomed. Heal. Inform. 27 (8) (2023) 3994–4005.

[12] D. Qiu, Y. Cheng, X. Wang, Medical image super-resolution reconstruction algorithms based on deep learning: A survey, Comput. Methods Programs Biomed. 238 (2023) 107590.

[13] S. Chen, Z. Han, E. Dai, X. Jia, Z. Liu, L. Xing, X. Zou, C. Xu, J. Liu, Q. Tian, Unsupervised image super-resolution with an indirect supervised path, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 468–469.

[14] Y. Wei, S. Gu, Y. Li, R. Timofte, L. Jin, H. Song, Unsupervised real-world image super resolution via domain-distance aware training, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 13385–13394.

[15] J. Liu, H. Zhang, J.-H. Tian, Y. Su, Y. Chen, Y. Wang, R2d2-GAN: Robust dual discriminator generative adversarial network for microscopy hyperspectral image super-resolution, IEEE Trans. Med. Imaging (2024).

[16] J. Chi, Z. Sun, L. Meng, S. Wang, X. Yu, X. Wei, B. Yang, Low-dose CT image super-resolution with noise suppression based on prior degradation estimator and self-guidance mechanism, IEEE Trans. Med. Imaging (2024).

[17] D. Mishra, O. Hadar, Accelerating neural style-transfer using contrastive learning for unsupervised satellite image super-resolution, IEEE Trans. Geosci. Remote Sens. 61 (2023) 1–14.

[18] H. Lin, J. Zou, K. Wang, Y. Feng, C. Xu, J. Lyu, J. Qin, Dual-space high-frequency learning for transformer-based mri super-resolution, Comput. Methods Programs Biomed. 250 (2024) 108165.

[19] S. Dong, Z. Cai, G. Hangel, W. Bogner, G. Widhalm, Y. Huang, Q. Liang, C. You, C. Kumaragamage, R.K. Fulbright, et al., A flow-based truncated denoising diffusion model for super-resolution magnetic resonance spectroscopic imaging, Med. Image Anal. 99 (2025) 103358.

[20] Y. Huang, T. Miyazaki, X. Liu, K. Jiang, Z. Tang, S. Omachi, Learn from orientation prior for radiograph super-resolution: Orientation operator transformer, Comput. Methods Programs Biomed. 245 (2024) 108000.

[21] D. Qiu, Y. Cheng, X. Wang, Progressive feedback residual attention network for cardiac magnetic resonance imaging super-resolution, IEEE J. Biomed. Heal. Inform. 27 (7) (2023) 3478–3488.

[22] D. Mishra, O. Hadar, Self-FuseNet: Data free unsupervised remote sensing image super-resolution, IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 16 (2023) 1710–1727.

[23] A.B. Szczotka, D.I. Shakir, M.J. Clarkson, S.P. Pereira, T. Vercauteren, Zero-shot super-resolution with a physically-motivated downsampling kernel for endomicroscopy, IEEE Trans. Med. Imaging 40 (7) (2021) 1863–1874.

[24] M. Eadie, J. Liao, W. Ageeli, G. Nabi, N. Krstajić, Fiber bundle image reconstruction using convolutional neural networks and bundle rotation in endomicroscopy, Sensors 23 (5) (2023) 2469.

[25] K. Yang, H. Zhang, Y. Qiu, T. Zhai, Z. Zhang, Self-supervised joint learning for pCLE image denoising, Sensors 24 (9) (2024) 2853.

[26] J. Zhou, X. Dong, Q. Liu, Context-aware dynamic filtering network for confocal laser endomicroscopy image denoising, Phys. Med. Biol. 68 (19) (2023) 195014.

[27] X. Zhang, M. Tan, M. Nabil, R. Shukla, S. Vasavada, S. Anandasabapathy, M.A. Anastasio, E. Petrova, Deep-learning-based image super-resolution of an end-expandable optical fiber probe for application in esophageal cancer diagnostics, J. Biomed. Opt. 29 (4) (2024) 046001–046001.

[28] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2223–2232.

[29] T. Park, A.A. Efros, R. Zhang, 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.

[30] B. Lim, S. Son, H. Kim, S. Nah, K. Mu Lee, Enhanced deep residual networks for single image super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 136–144.

[31] J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, in: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, the Netherlands, October 11-14, 2016, Proceedings, Part II 14, Springer, 2016, pp. 694–711.

[32] M.S. Sajjadi, B. Scholkopf, M. Hirsch, Enhancenet: Single image super-resolution through automated texture synthesis, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 4491–4500.

[33] A.A. Borkowski, M.M. Bui, L.B. Thomas, C.P. Wilson, L.A. DeLand, S.M. Mastorides, Lung and colon cancer histopathological image dataset (lc25000), 2019, arXiv preprint arXiv:1912.12142.

[34] N. Ahn, J. Yoo, K.-A. Sohn, Simusr: A simple but strong baseline for unsupervised image super-resolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 474–475.

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

[36] X. Yi, E. Walia, P. Babyn, Generative adversarial network in medical imaging: A review, Med. Image Anal. 58 (2019) 101552.

[37] R. Zhang, P. Isola, A.A. Efros, E. Shechtman, O. Wang, The unreasonable effectiveness of deep features as a perceptual metric, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 586–595.

[38] K. Li, S. Yang, R. Dong, X. Wang, J. Huang, Survey of single image super-resolution reconstruction, IET Image Process. 14 (11) (2020) 2273–2290.

[39] R.M. Haralick, K. Shanmugam, I.H. Dinstein, Textural features for image classification, IEEE Trans. Syst. Man Cybern. (6) (1973) 610–621.

[40] Y. Li, Y. Fan, H. Liao, Self-supervised speckle noise reduction of optical coherence tomography without clean data, Biomed. Opt. Express 13 (12) (2022) 6357–6372.