scientific reports

Check for updates

OPEN

Unsupervised translation of vascular masks to NIR-II fluorescence images using Attention-Guided generative adversarial networks

Lu Fang[1,2], Huaixuan Sheng[3], Huizhu Li[3], Shunyao Li[3], Sijia Feng[3], Mo Chen[4], Yunxia Li[3], Jun Chen[3] & Fuchun Chen[1]✉

The second near-infrared window (NIR-II) fluorescence imaging is a crucial technology for investigating the structure and functionality of blood vessels. However, challenges arise from privacy concerns and the significant effort needed for data annotation, complicating the acquisition of near-infrared vascular imaging datasets. To tackle these issues, methods based on deep learning for data synthesis have demonstrated promise in generating high-quality synthetic images. In this paper, we propose an unsupervised generative adversarial network (GAN) approach for translating vascular masks into realistic NIR-II fluorescence vascular images. Leveraging an attention mechanism integrated into the loss function, our model focuses on essential features during the generation process, resulting in high-quality NIRII images without the need for supervision. Our method significantly outperforms eight baseline techniques in both visual quality and quantitative metrics, demonstrating its potential to address the challenge of limited datasets in NIR-II medical imaging. This work not only enhances the applications of NIR-II imaging but also facilitates downstream tasks by providing abundant, high-fidelity synthetic data.

Keywords Image synthesis, NIR-II imaging, Deep learning, Generative adversarial network

NIR-II fluorescence vascular imaging serves as an important tool for understanding the structure and function of human vasculature[1–3]. This technology allows for non-invasive observation of the microstructure of the vascular system, facilitating the exploration of internal physiological functions and aiding in disease diagnosis and the development of appropriate therapeutic regimens[4]. However, collecting NIR-II medical imaging datasets poses several challenges, including the need for specialized equipment and expertise, as well as the time-consuming and labor-intensive nature of obtaining and annotating medical images[5]. Therefore, methods are needed to effectively overcome the challenge to advance the application of NIR-II fluorescence vascular imaging technology in cardiovascular disease diagnosis and research.

Traditional data augmentation techniques are widely used in the medical field to increase data volume. By applying transformations such as image rotation, scaling, and cropping, these techniques enhance data diversity and partially alleviate data scarcity[6]. However, since these techniques generate data from existing samples, they often fail to capture the complete distribution of imaging data. In response, deep learning-based data synthesis methods have emerged, showcasing the ability to explore complex nonlinear relationships and generate highly diverse, realistic synthetic data. The retina contains a complex and delicate network of blood vessels, which are an integral part of the body's overall vascular system. As a result, retinal images bear structural resemblances to vascular images from other body parts. This connection implies that the principles and algorithms utilized in retinal fundus image generation can provide valuable guidance for vascular image generation[7]. The process of retinal fundus image generation primarily employs GAN and its variants to generate diverse and high-fidelity

[1]Chinese Academy of Sciences, Shanghai Institute of Technical Physics, Shanghai 200083, China. [2]University of Chinese Academy of Sciences, Beijing 100049, China. [3]Sports Medicine Institute of Fudan University, Department of Sports Medicine, Huashan Hospital, Fudan University, Shanghai 200040, China. [4]Department of Bone and Joint Surgery, Department of Orthopedics, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200001, China. ✉email: fuchun.chen@mail.sitp.ac.cn

images. For instance, Costa et al. implemented an adversarial autoencoder to generate realistic vascular trees and obtained relevant retinal images through a GAN[8]. Similarly, Zhao et al. introduced Tub-GAN and Tub-sGAN, which are suitable for small sample sizes, and capable of generating multiple realistic images based on the same tubular structured annotation[9]. R-sGAN integrates a nonlinear variant of the Gated Recurrent Unit, embedding the GAN generator within the recurrent neural network's unit structure. This architecture leverages its recursive properties to produce images in various styles, enhancing versatility[10]. Saeed et al. combined variational autoencoders (VAEs) and GAN architectures[11]. They employed multiple VAEs to generate vascular structures and optic discs separately. By employing the sharpening and varying vessels layer, the abundance of vessels in the reconstructed vascular tree is varied, which is then fed into the image translation network implemented by GAN to generate the fundus images. Additionally, Beji et al. proposed a novel framework called Seg2GAN, which combines Deep Convolutional Generative Adversarial Networks (DCGAN), U-Net, and Pix2pix to generate medical data to improve segmentation quality[12]. While significant advancements have been made in retinal image generation, research for infrared image synthesis remains limited in the medical field. Keivanmarz et al. introduced a conditional GAN (cGAN) to map RGB vein images to near-infrared (NIR) images for the first time, achieving high accuracy in vein length measurements post-skeletonization[13]. Researchers have also explored infrared image synthesis across various fields. To enhance target detection capabilities, Abbott et al. integrated object-specific loss into the CycleGAN architecture, facilitating the translation between unpaired RGB and Long-Wave Infrared (LWIR) datasets[14]. Addressing the scarcity of large annotated datasets for end-to-end tracking in Thermal Infrared (TIR) imaging, Zhang et al. generated a large annotated TIR dataset using image-to-image translation networks like Pix2pix and CycleGAN[15]. Their findings revealed that deep features trained on synthetic data outperformed those trained on conventional annotated TIR datasets. In another study, MWIRGAN incorporated perceptual loss into CycleGAN to transform visible light images to mid-wave infrared images effectively[16]. In remote sensing, Illarionova et al. developed a method based on Pix2pix for generating NIR band images from RGB satellite images[17]. Shukla et al. combines attention-based GANs with a super-resolution module to predict high-resolution NIR images from RGB images, specifically for large-scale plant data generation[18]. Unlike traditional methods that depend on spectral consistency, Contour GAN introduces spatial constraints for exploring cross-domain translation through contour consistency[19]. In the realm of automatic lip-reading, Lee et al. employed an enhanced U-Net network to estimate infrared and depth images from RGB inputs, achieving significantly higher recognition rates than using RGB images alone[20]. In addition to this, researchers have focused on asymmetric mode transformation from near-infrared to visible light. Yan et al. proposed a method for converting NIR images to RGB images using a U-Net-based subnetwork for texture extraction and a CycleGAN-based subnetwork for coloring, ultimately merging the outputs into a final RGB image[21]. Dou et al. proposed an asymmetric CycleGAN model that addresses the issue of mismatched complexity in transformation domains by employing generators of different sizes[22]. ACGANs utilize an asymmetric architecture based on cycleGAN, combining UNet and ResNet in the generator, while employing a feature pyramid network in the discriminator to colorize NIR images into RGB images[23].

Despite these advancements, the lack of available datasets hinders the development of deep learning-based downstream tasks for NIR-II fluorescence vascular imaging. Therefore, an unsupervised NIR-II fluorescence vascular image synthesis method is proposed, aimed at creating a reliable dataset that provides abundant high-quality data for subsequent research and development. The contributions of this work can be summarized as follows:

(1) A GAN-based model is designed for the unsupervised translation of vessel masks to NIR-II fluorescence vascular images.

(2) An attention mechanism is introduced that effectively integrates style patterns while preserving content structure, thereby enhancing the quality and detail of the generated images.

(3) Experimental results demonstrate that our method significantly outperforms eight baseline methods in both visual quality and quantitative analysis.

The structure of this paper is as follows: Firstly, related works are introduced. Next, a detailed explanation of the proposed method is provided. The Experiment section presents the experimental setup, including the datasets used and the values set for parameters in the experiments. The experimental results are then discussed, comparing our method with other baseline approaches in terms of visual quality and quantitative performance. Finally, the main contributions of this paper are summarized.

Related works

Generative adversarial network

GAN is a non-supervised learning model proposed by Goodfellow et al. in 2014, primarily consisting of a generator network and a discriminator network[24]. This innovative architecture achieves data synthesis through the adversarial interplay between these two networks. The key to GAN's success lies in the introduction of adversarial loss, as shown in Eq. 1, which optimizes the generator by minimizing the loss function of the discriminator, thus generating images that are difficult to distinguish from real ones[25].

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_{noise}(z)}[\log(1 - D(G(z)))] \tag{1}$$

where E(*) denotes the expected value of the distribution function, $p_{data}(x)$ is the distribution of real samples, and $p_{noise}(z)$ denotes the low-dimensional noise distribution. G and D are the generator and discriminator, respectively.

The development of GAN has led to several classical variants, such as DCGAN[26], WGAN (Wasserstein GAN)[27], CycleGAN[25], and Pix2Pix[28]. These methods have made significant contributions to improving the

quality of the generated images, increasing the stability of training, and broadening the scope of application. Today, GAN has been widely applied in various fields, including image synthesis, image restoration, and representation learning, exerting a profound influence in the field of computer vision.

Image synthesis

Image synthesis is a crucial task in the field of computer vision, aimed at creating new images from input data such as random noise or conditional information. Traditional image synthesis methods typically rely on models with explicit rules or generate images based on the pixels and feature descriptions of existing images. While these methods are relatively fast and do not require large datasets, they often produce images of lower quality and limited diversity, restricting their application in more complex tasks.

With the rapid advancement of deep learning technologies, the concept of end-to-end synthesis has been gradually introduced into the field of image synthesis, significantly enhancing the quality of generated images. Techniques based on GANs[29,30], VAEs[31], and autoregressive models[32] have played a vital role in this domain. Compared to traditional methods, these deep learning models not only generate higher-quality images but also exhibit greater diversity, driving innovation and progress in image synthesis technology.

The primary goal of this work is to generate NIR-II fluorescence vascular images that not only align with the style of the target domain but also retain the structural details of the input images. To better understand this task, we compare it with several benchmark methods. Below, we provide a detailed comparison of these methods in the context of our task, and the experiment comparison of subjective visual effect and object evaluation can be seen in Experiment section.

DCGAN is one of the earliest GAN architectures, designed for generating images from random noise[26], which employed convolution layers and transposed convolution layers as discriminators and generators, respectively[33,34]. While it demonstrates strong capabilities in producing realistic images, it does not perform domain translation and lacks mechanisms for preserving input image structure. As a result, it is not directly applicable to tasks requiring transformation between two domains. WGAN is a variant of the original GAN, which introduces the Wasserstein distance as the loss function to measure the difference between the generator distribution and the real data distribution[27,35]. WGAN-GP is an improved version of WGAN, it adds a weight-clipping trick to limit the range of parameters which contributes to the convergence of the loss function[36,37]. Both WGAN and WGAN-GP enhance training dynamics, but they still struggle to accurately capture complex vessel structures. SAGAN introduces a self-attention mechanism to better model long-range dependencies in images, improving the quality of generated images in terms of global consistency[38]. While this improves overall detail generation, it may overemphasize global features and neglect the local details necessary for vascular structure preservation. Eigengan represents a cutting-edge approach within the domain of GANs. It incorporates hierarchical feature-vector learning, which enables independent control over specific image attributes at each layer. This technique aims to enhance the understanding and representation of the data distribution, thereby facilitating more effective and stable training of GANs[39]. While this introduces greater flexibility, it may still fail to maintain the structural integrity of vascular features due to the lack of explicit input vessel guidance.

U-GAT-IT employs an encoder-decoder architecture with attention modules and Adaptive Layer-Instance Normalization, enabling it to perform unsupervised image-to-image translation[40]. The attention mechanism helps the model focus on important features, and the adaptive normalization contributes to the robustness and flexibility of the model, making it effective for tasks such as style transfer and domain adaptation[41,42]. However, it does not explicitly include a mechanism like cycle-consistency loss to enforce input-output structural consistency, which is crucial for our task. HiSD utilizes hierarchical style disentanglement, which breaks down the attributes of an image into hierarchical levels[43]. This allows for fine-grained control over different styles, facilitating more flexible and nuanced image transformations. These transformations can range from simple adjustments like color and texture changes to more complex manipulations[44]. While HiSD could effectively generate diverse outputs and perform nuanced transformations, it focuses on style diversity rather than structural preservation. Consequently, it is less suited for tasks where maintaining the structural integrity of vascular masks is critical.

In this work, we employ a CycleGAN-based network to perform domain translation, transforming vascular masks into NIR-II fluorescence vascular images. CycleGAN is specifically designed for unsupervised image-to-image translation tasks, which use the cycle-consistency loss to ensure that the translated images retain the structural content of the input images[25]. This makes it particularly suitable for tasks requiring style transfer while maintaining the structural fidelity of the input data. For vascular image synthesis, this is critical to preserving the intricate vessel structures during the translation process. In addition to CycleGAN's inherent advantages, the attention mechanism within the feature network plays a pivotal role in guiding the model to focus on the most important regions of the vascular masks, ensuring that fine details, such as smaller vessels, are preserved in the generated NIR-II fluorescence images. Compared to methods like U-GAT-IT, which have attention mechanisms, the combination of CycleGAN and the attention mechanism is more effective for retaining fine-grained structural features, resulting in more realistic-looking NIR-II fluorescence vascular images with enhanced detail representation.

Methods

This section presents the proposed method as shown in Fig. 1. In this work, we aim to learn how to map from tubular structure masks to realistic NIR-II fluorescence modality for vascular images. Specifically, this method employs a CycleGAN-based network architecture composed of two generators, two discriminators, and a feature network. The network uses vascular binary mask images and NIR-II vascular images as inputs. Generator G_AB is tasked with converting mask images into fake NIR-II images, while generator G_BA converts NIR-II images back into fake mask images, ensuring both forward and backward cycle consistency.

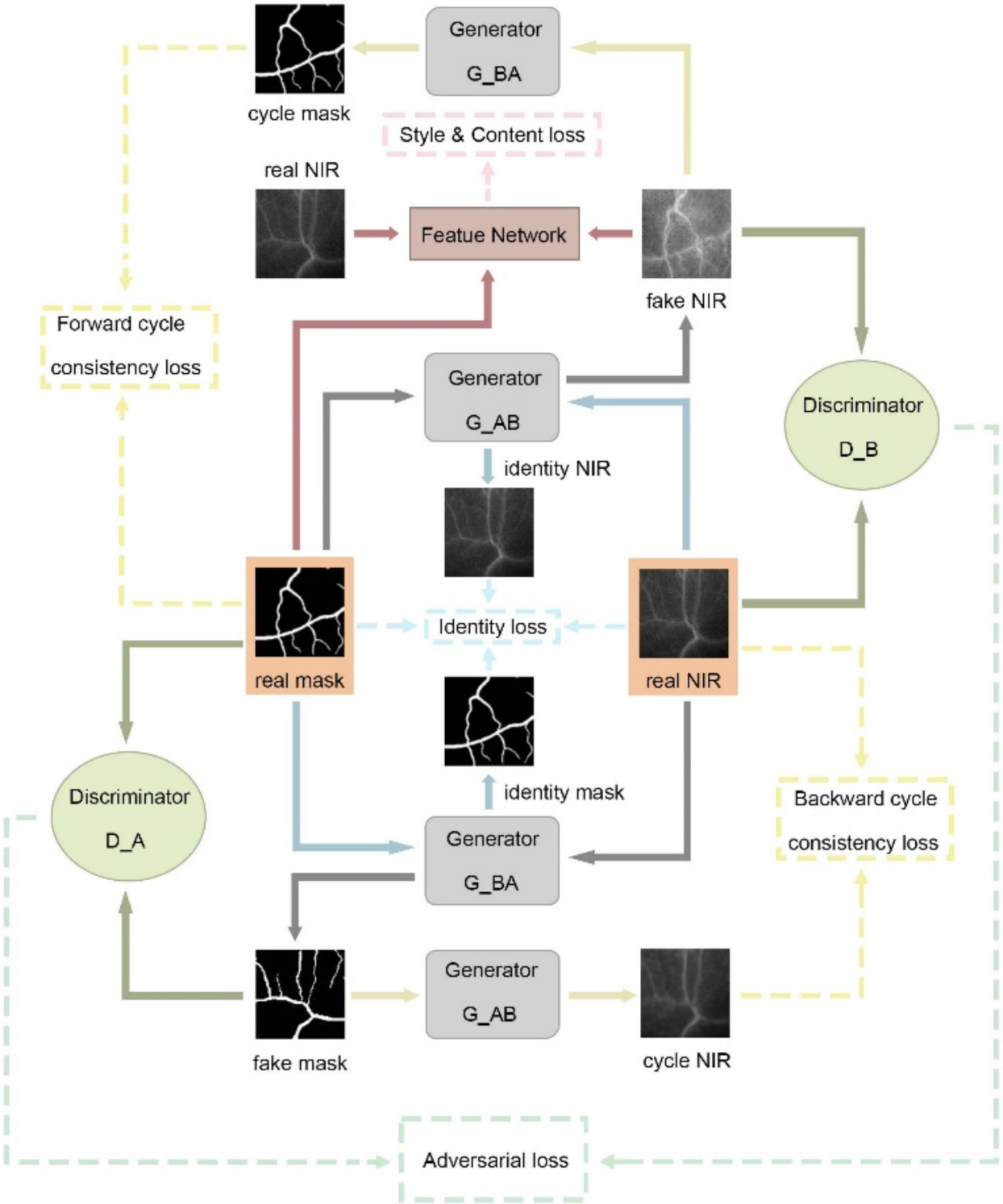

Fig. 1. The framework of the proposed approach.

Basic generative framework

This study employs CycleGAN as the primary framework, aiming to achieve unsupervised translation of unpaired images. The CycleGAN architecture consists of two generators and two discriminators, forming two adversarial networks. Through the adversarial interplay between the generators and discriminators, the model is able to generate high-quality NIR-II vascular images.

Generator and discriminator

According to our design, the generator consists of multiple convolutional layers, downsampling layers, and a series of residual blocks, effectively capturing and processing the essential features of the input data. The short-distance skip connections within the residual blocks facilitate smoother gradient flow, alleviating the issue of gradient disappearance. When the input is x, the feature representations extracted at each layer are as follows[45]:

$$\mathcal{F}_G^0 = Conv^1(x) \tag{2}$$

$$\mathcal{F}_G^1 = Down(Down(\mathcal{F}_G^0)) \tag{3}$$

$$\mathcal{F}_G^2 = \mathrm{Res}(\mathcal{F}_G^1) \tag{4}$$

$$\mathcal{F}_G^3 = Up(Up(\mathcal{F}_G^2)) \tag{5}$$

$$\mathcal{F}_G^4 = Conv^2(\mathcal{F}_G^3) \tag{6}$$

where $Conv^1$ represents a series of sequential operations, including padding, a convolutional layer (with a kernel size of 7), the instance normalization, and the ReLU activation function. Down denotes the downsampling operation, which includes a convolutional layer (kernel size = 3 and stride = 2), instance normalization, and the ReLU activation function. Res consists of 9 residual blocks. Up is the upsampling operation, composed of a transposed convolutional layer (with a kernel size of 3 and stride of 2), instance normalization, and ReLU activation function. The difference between $Conv^1$ and $Conv^2$ is that in $Conv^2$, the activation function is tanh.

With the generator designed to effectively produce high-quality outputs, the discriminator plays a crucial complementary role in the GAN-based model. It utilizes the PatchGAN architecture, comprising multiple convolutional layers and normalization operations, thereby enhancing the model's ability to capture fine-grained details while maintaining computational efficiency. This synergy between the generator and discriminator is essential for improving the overall performance of our system.

Loss function of cyclegan

During the training of CycleGAN, both adversarial loss and cycle consistency loss are introduced[25]. Adversarial loss encourages the generators to produce more realistic images, while cycle consistency loss ensures the reversibility of the transformation process, allowing the original input images to be effectively reconstructed after the transformation. This design not only improves the quality of the generated images but also improves the stability of the model. The details of the adversarial loss[25] are as follows:

$$\mathcal{L}_{GAN}(G_Y, D_Y, X, Y) = E_{y\sim p_{data}(y)}[\log D_Y(y)] + E_{x\sim p_{data}(x)}[\log(1 - D_Y(G_Y(x)))] \tag{7}$$

where the generator G_Y is responsible for translating images from the source domain X to the target domain Y, while the discriminator D_Y determines the authenticity of the data. Equation 7 is similar to Eq. 1. During the training process, the generator aims to minimize the adversarial loss, while the discriminator attempts to maximize it. Through this adversarial interplay, the generated data can become increasingly similar to the actual data distribution.

The adversarial loss for the transformation from the target domain Y back to the source domain X can be expressed as $L_{GAN}(G_X, D_X, Y, X)$. The total adversarial loss[25] is:

$$\mathcal{L}_{GAN} = \mathcal{L}_{GAN}(G_Y, D_Y, X, Y) + \mathcal{L}_{GAN}(G_X, D_X, Y, X) \tag{8}$$

The cycle consistency loss[25]can be expressed by the following:

$$\mathcal{L}_{cyc}(G_Y, G_X) = E_{x\sim p_{data}(x)}\left[\|G_X(G_Y(x)) - x\|_1\right] + E_{y\sim p_{data}(y)}\left[\|G_Y(G_X(y)) - y\|_1\right] \tag{9}$$

In an ideal scenario, the image x from the source domain X should remain unchanged after being processed by the generators G_Y and G_X. To achieve this, the generators actively strive to minimize the cycle consistency loss. Within the context of medical image processing, cycle consistency loss is essential, as it ensures reliable transformations between different image domains. By incorporating this loss, the model ensures that the generated images accurately preserve the characteristics of both the source and target domains, leading to transformations that closely align with the input images.

Additionally, we introduce identity loss[46] to encourage consistency in color composition between the input and output. The calculation of identity loss is as follows:

$$\mathcal{L}_{idt}(G_Y, G_X) = E_{y\sim p_{data}(y)}\left[\|G_Y(y) - y\|_1\right] + E_{x\sim p_{data}(x)}\left[\|G_X(x) - x\|_1\right] \tag{10}$$

By minimizing identity loss, the generator can approach identity mapping when processing real samples, thereby enhancing the overall quality of the generated images.

Feature network

This approach integrates a feature network into the basic architecture, leveraging a pre-trained VGG-19-based model to capture rich information from images. Previous studies[47–49]on attention mechanisms in style transfer often focused on deep CNN features while overlooking low-level textures, leading to local distortions. Inspired by the paper[50], this approach extracts features from various convolutional layers of the VGG-19-based

network[51], including both shallow and deep features, achieving a more comprehensive feature representation. This combination not only mitigates local distortions but also enhances the naturalness and realism of the generated images. Furthermore, incorporating multi-level features improves the effectiveness of our attention mechanism in capturing essential details and structures within the images. The structure diagram of the feature network is shown in Fig. 2.

Content loss

In the subsequent calculations, for the sake of convenience, the input source domain images are denoted as I_c, the input target domain image as I_s, and the generated image as I_{cs}. The feature network is represented as a series of consecutive CNN layers. When these images are input into the feature network, the features obtained from a specific layer λ are denoted as Ψ_c^λ, Ψ_s^λ, and Ψ_{cs}^λ, respectively.

To ensure that the generated image maintains content consistency with the source domain image, content loss is introduced. Specifically, the features of both the generated image and the source domain image are normalized to eliminate the effects of different scales and distributions. By comparing the mean squared error between these normalized features, we can effectively assess their content similarity. This process not only emphasizes the preservation of overall structure but also ensures the accurate conveyance of details, resulting in high-quality image generation. The content loss[52] can be calculated as follows:

$$\mathcal{L}_{cont} = \sum_{\lambda=1}^{4} \left\| Norm(\Psi_{cs}^\lambda) - Norm(\Psi_c^\lambda) \right\|_2 \tag{11}$$

where Norm here denotes channel-wise mean-variance normalization.

Style loss

First, based on the studies in[48,50,53], the global style loss is introduced to ensure that the generated image is consistent with the global statistical features of the target domain image and effectively transmits the overall style. This loss function enhances the similarity between the generated image and the target image by comparing their feature means and variances, ensuring a comprehensive representation of the style characteristics[48]:

$$\mathcal{L}_{gs} = \sum_{\lambda=1}^{4} \left(\left\| E(\Psi_{cs}^\lambda) - E(\Psi_s^\lambda) \right\|_2 + \left\| Std(\Psi_{cs}^\lambda) - Std(\Psi_s^\lambda) \right\|_2 \right) \tag{12}$$

where Std(*) represents the operation for calculating the standard deviation.

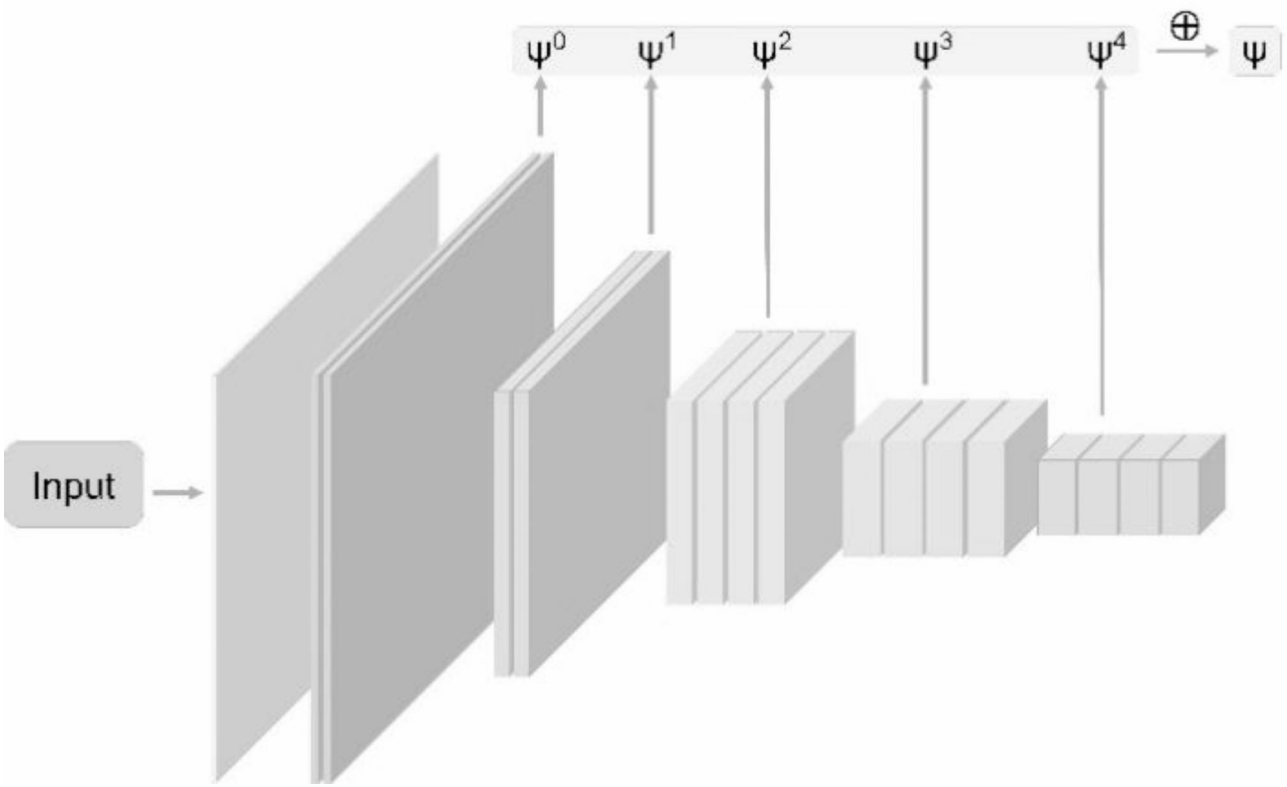

Fig. 2. The diagram of the feature network.

The introduction of local style loss aims to ensure that the model produces outputs that are consistent with the target image's style at the level of local features, making the generated images more visually appealing. This loss function is designed to leverage attention mechanisms proposed in[50], by calculating attention weights based on the content features generated from the source domain image and the style features from the target domain image.

The attention mechanism effectively combines content and style features by generating attention-weighted mean and variance maps from the content and style feature, resulting in the target feature map, as shown in Fig. 3. This method adjusts based on each feature point, enhancing the accuracy and flexibility of feature transfer. Specifically, the attention mechanism dynamically generates attention maps by calculating the similarity between content and style features, allowing the model to focus on important feature regions, thereby optimizing the generation results. The local loss function penalizes the differences between the target feature map and the generated image's feature maps, ensuring their distributions align for effective local modality transfer. The calculation flowchart of the target feature map can be seen as follows:

For $\psi^{\lambda} \in R^{C}{}_{\lambda} x H_{\lambda} W_{\lambda}$, the feature layers from layers below λ are first interpolated to match the size of ψ^{λ}, and then channel-wise mean-variance normalization is applied:

$$\mathcal{N}_c^{\lambda} = Norm(Concat(Interp(\Psi_c^{0:\lambda-1}), \Psi_c^{\lambda})) \tag{13}$$

where Interp refers to the interpolation of feature maps, and Concat is concatenating different feature maps along the channel dimension. By applying the same operation, we can obtain the output of the style feature map at this step.

Next, matrix multiplication on the two normalized feature maps is performed and the Softmax function is applied to obtain the attention map.

$$\mathcal{A} = \mathrm{Softmax}(\mathcal{N}_c^{\lambda} \otimes \mathcal{N}_s^{\lambda}) \tag{14}$$

The attention-weighted mean and attention-weighted variance are calculated as follows:

$$mean = \mathcal{A} \otimes \Psi_s^{\lambda} \tag{15}$$

$$Std = \sqrt{\mathrm{Relu}(\mathcal{A} \otimes (\Psi_s^{\lambda})^2 - mean^2)} \tag{16}$$

The obtained statistical information is utilized to compute the target features:

$$\Phi_{att}^{\lambda} = Std \cdot Norm(\Psi_c^{\lambda}) + mean \tag{17}$$

where denotes the Hadamard product of matrices.

The local style loss[50] is calculated as the difference between the feature maps of the generated and target images:

$$\mathcal{L}_{ls} = \sum_{\lambda=1}^{4} \left\| \Psi_{cs}^{\lambda} - \Phi_{att}^{\lambda} \right\|_2 \tag{18}$$

Total loss

The total loss is defined as follows:

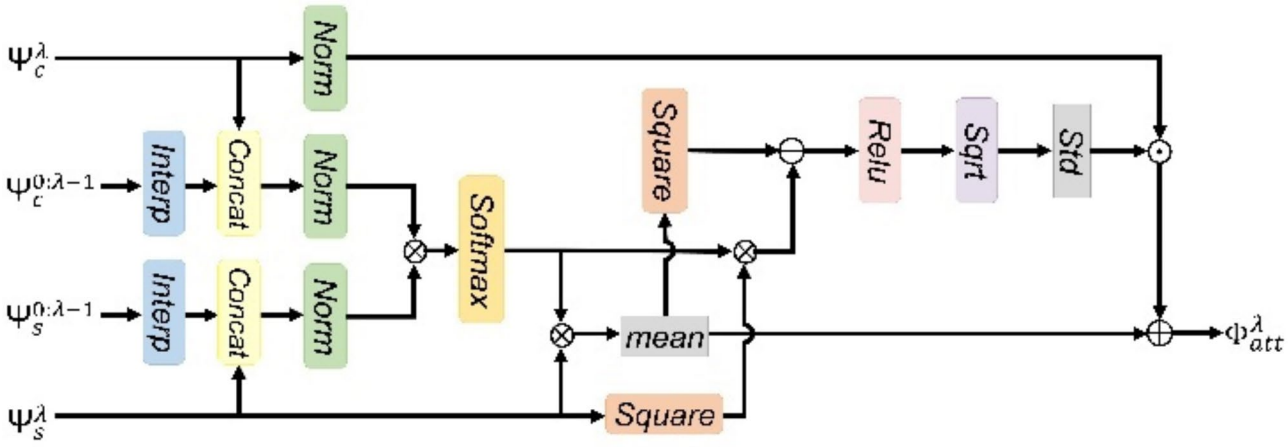

Fig. 3. The calculation flowchart of the target feature map.

$$\mathcal{L}_{total} = \mathcal{L}_{GAN} + \gamma_{cyc}\mathcal{L}_{cyc} + \gamma_{idt}\mathcal{L}_{idt} + \gamma_{cont}\mathcal{L}_{cont} + \gamma_{gs}\mathcal{L}_{gs} + \gamma_{ls}\mathcal{L}_{ls} \quad (19)$$

where the coefficients γ control the contribution of each loss term, allowing for flexible tuning of the model's performance based on specific tasks or dataset characteristics.

Experiment

Datasets and preprocessing

Data acquisition

The NIR-II fluorescence imaging was performed using an imaging system developed by the Shanghai Institute of Materia Medica, Chinese Academy of Sciences. Prior to the imaging procedure, the mice were anesthetized using isoflurane and intravenously injected with 0.35 mL of quantum dots (QDs) as a fluorescent tracer. The fluorescence was excited using two 808 nm fiber-coupled lasers, and the emitted fluorescence signal was captured by the NIRvana 640 InGaAs camera (Princeton, USA). To ensure accurate detection of the NIR-II fluorescence, a long-pass filter with a cutoff wavelength of 1250 nm was used. The animals were euthanized by overdose of anesthesia.

The segmentation masks from the DRIVE[54], STARE[55], and CHASE_DB1[38] retinal image datasets are selected as the source domain data. In the datasets, all mask images were manually segmented by experts using standard image annotation tools, ensuring high-quality and accurate delineation of the vascular structures. Specifically, the DRIVE dataset provides 40 segmentation masks (565 × 584 pixels), the STARE dataset includes 20 masks (700 × 605 pixels), and the CHASE_DB1 dataset contains 28 masks (999 × 960 pixels).

Image preprocessing

For preprocessing the source domain datasets, we first adjust the image contrast and apply OTSU thresholding to obtain the foreground mask when it is not provided. The radius of this circular mask is used to crop the image with the smallest enclosing rectangle of the optic disc. Given the varying sizes of these datasets, all datasets are resized to 512 × 512 pixels. To ensure the vessel sizes in the masks match those in the NIR-II images, a 400 × 400 region centered on the image is extracted and resized to 300 × 300 pixels. Subsequently, both source domain images (retinal masks) and target domain images (NIR-II images) are cropped into 128 × 128 pixel patches, resulting in a final dataset of 792 images from the retinal dataset and 406 images from the NIR-II dataset. The dataset is then divided into a training set and a test set at a 7:3 ratio.

Experimental and setup

Eight representative methods were used to compare with the model, including DCGAN[26], SAGAN[39], WGAN[27], WAN-GP[36], Eigengan[56], CycleGAN[25], U-GAT-IT[40], HiSD[43].

The Adam optimizer was used with an initial learning rate of 0.0002, a batch size of 1, and a total of 400 epochs for training. In this study, the hyperparameters are set as $\gamma_{cyc} = 10$, $\gamma_{idt} = 5$, and γ_{cont}, γ_{gs}, γ_{ls} are set to 0.01, 0.002, and 0.001, respectively.

Experimental results

Subjective visual effects

Figure 4 shows the results of different generative models in generating NIR-II fluorescence vascular images. The leftmost column displays the input vascular structures, clearly depicting the outlines and branches of the vessels, providing foundational information for subsequent generations. Next to it, the right column showcases the authentic NIR-II fluorescence images.

The images generated by DCGAN are relatively blurry, with unclear vessel edges and noise present in certain areas. Although they capture the basic shape of the vessels, the overall quality is inferior to that of the real images. In contrast, SAGAN shows improvement in generation effects, producing clearer images with richer details compared to DCGAN; however, some areas still appear unnatural, and the contrast is lacking. Images generated by WGAN show enhanced detail retention, with the vascular structures becoming more pronounced, though some parts still exhibit blurriness. WGAN-GP performs better in terms of clarity and contrast, resulting in a significant quality improvement in the generated images. On the other hand, Eigengan's generated images appear overly smooth, lacking detail, especially in the representation of small vascular structures, leading to a deficit in realism. CycleGAN's generated images show improved visual effects, adequately reproducing vascular structures with high detail clarity, but there remains room for enhancement in contrast, as some areas appear slightly blurred. U-GAT-IT generates images with outstanding detail and contrast, effectively reproducing the shapes and structures of the vessels, with an overall effect close to that of real images. Although HiSD generates images with high clarity, they are overly smooth, resulting in lower brightness of the vascular structures and a noticeable gap from the authentic images. Compared with the above methods, our method demonstrates better visual effects, with generated images closely resembling real NIR-II images in terms of detail and contrast, featuring clear vascular structures and natural shapes, showcasing superior realism and visual quality.

Objective evaluation

In this study, the Fréchet Inception Distance (FID)[57]and Kernel Inception Distance (KID)[58]are used to quantitatively evaluate the performance of the model in generating NIR-II fluorescence vascular images. FID is a widely used quality assessment metric in the field of image generation, measuring the distance between the distributions of generated images and real images in feature space, as extracted by the pre-trained Inception model[59]. KID is another important metric for assessing image quality in generative models. Similar to FID, KID utilizes a pre-trained Inception network to extract features, providing a measure of similarity between

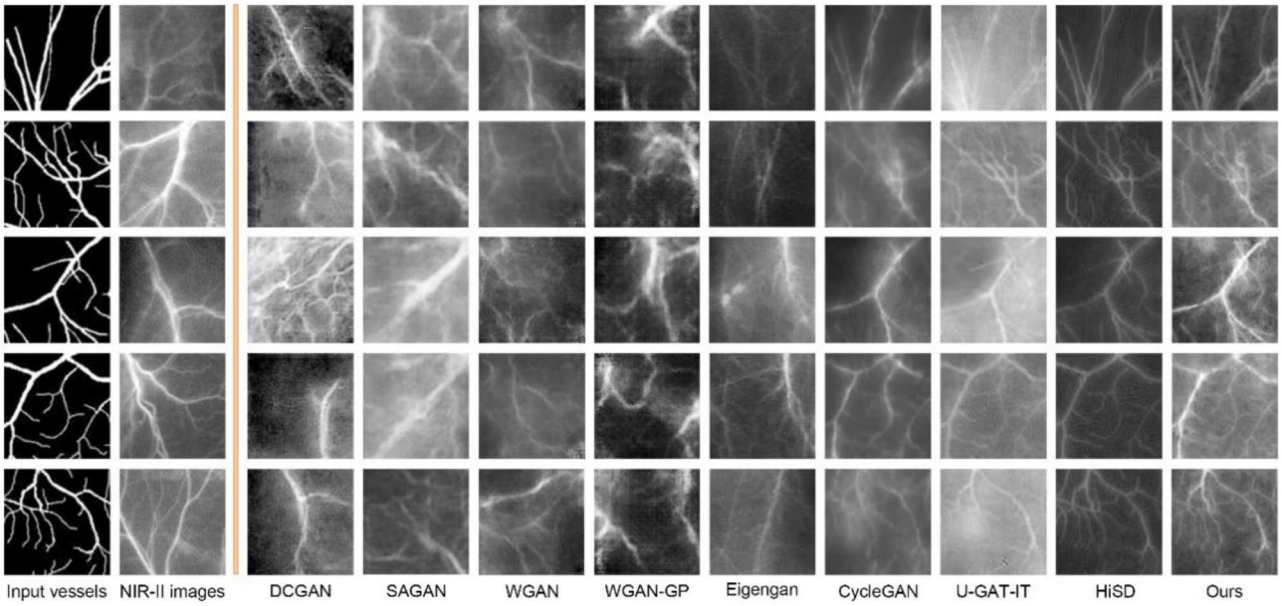

Fig. 4. The visual comparison results of different methods.

Methods	FID	KID
DCGAN	137.4243	7.8570 ± 0.2353
SAGAN	203.1509	12.2567 ± 0.2062
WGAN	198.2326	13.2762 ± 0.2943
WGAN-GP	216.2753	12.2758 ± 0.2012
Eigengan	177.5905	13.7936 ± 0.2949
CycleGAN	91.5453	6.6770 ± 0.2692
U-GAT-IT	63.6901	2.1281 ± 0.0955
HiSD	108.9341	4.7086 ± 0.1523
Ours	**55.7532**	**1.6025 ± 0.1249**

Table 1. The quantitative comparison results of different methods based on the FID and kid×100 ± std.×100.

the distributions of real and generated images. However, it employs the Maximum Mean Discrepancy (MMD) to evaluate the distance between these two distributions, using a kernel-based approach to capture their characteristics, which makes KID particularly reliable for evaluations with smaller sample sizes and offers clearer performance feedback. The lower FID and KID scores indicate a closer similarity between the distributions of real and generated images, suggesting better quality and diversity of the generated samples.

From Table 1, it can be seen that the proposed method achieves the lowest scores on the FID and KID metrics, with values of 55.75322 and 1.6025 ± 0.1249, respectively. This is consistent with the visual analysis results in Fig. 4, indicating the high similarity between the generated images and the real images.

Discussion

In this study, a GAN-based method is proposed to generate realistic NIR-II fluorescence vascular images from vascular masks. The primary objective of this work is to provide an effective solution to the challenge of limited NIR-II medical imaging datasets. By transforming vascular masks into high-quality synthetic NIR-II images, the method addresses the issues of data scarcity and the considerable effort required for manual image annotation. Ultimately, these advancements are expected to enhance the application of deep learning techniques in the field of NIR-II fluorescence imaging.

Though this method has achieved significant results, it is essential to take into account its limitations. The size of the dataset plays a critical role in the performance of deep learning models. In our study, given the potential scarcity of the NIR-II fluorescence vascular imaging dataset, it could pose challenges to the training and generalization capabilities of the model. Additionally, the computational resources also limit its application. The GAN-based model architecture requires significant computational resources. While the model performs well with the available dataset, scaling the model to handle even larger, more diverse datasets would require significant computational resources, which could limit its applicability in resource-constrained environments. What's more, the validation of the algorithm is limited to mice NIR-II vascular images. To improve the applicability

of the method in clinical settings, further validation is essential on more diverse datasets, including both animal and human images. In future research, we will focus on expanding the dataset by collecting additional NIR-II fluorescence vascular images from diverse sources, which will enhance the model's ability to generalize across various clinical scenarios. Additionally, we plan to explore model optimization techniques, such as pruning and transfer learning, as well as leverage more computationally efficient architectures, which will not only improve the model's scalability but also reduce its dependence on extensive computational resources, allowing it to effectively handle larger, more diverse datasets in future applications.

Conclusions

In this study, an unsupervised approach utilizing a GAN-based architecture is proposed for generating NIR-II fluorescence vascular images. This method effectively translates vessel masks into realistic NIR-II images, addressing the challenge of limited datasets in NIR-II medical imaging. By integrating an attention mechanism, the quality and detail of the generated images are enhanced, ensuring the preservation of both style and content. Experimental results demonstrate that our method is superior to existing baseline techniques in terms of visual quality and quantitative analysis, providing a robust solution for advancing NIR-II vascular imaging applications.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Received: 5 December 2024; Accepted: 20 February 2025
Published online: 25 February 2025

References

1. Yu, X. et al. Deciphering of cerebrovasculatures via ICG-assisted NIR-II fluorescence microscopy. *J. Mater. Chem. B* **7**, 6623-6629 (2019).
2. Lou, H. et al. A novel NIR-II nanoprobe for precision imaging of micro-meter sized tumor metastases of multi-organs and skin flap. *Chem. Eng. J.* **449,** 137848 (2022).
3. Cao, Z. et al. Thrombus-targeted nano-agents for NIR-II diagnostic fluorescence imaging-guided flap thromboembolism multi-model therapy. *J. Nanobiotechnol.* **20**, 447 (2022).
4. Zhang, H. et al. Imaging the deep spinal cord microvascular structure and function with High-Speed NIR-II fluorescence microscopy. *Small Methods* **6**, 2200155 (2022).
5. Salimi, M., Roshanfar, M., Tabatabaei, N. & Mosadegh, B. Machine Learning-Assisted Short-Wave infrared (SWIR) techniques for biomedical applications: towards personalized medicine. *J. Personalized Med.* **14** (2024).
6. Mendes, J. et al. Lung CT image synthesis using GANs. *Expert Syst. Appl.* **215**, 119350 (2023).
7. Go, S., Ji, Y., Park, S. J. & Lee, S. Generation of Structurally Realistic Retinal Fundus Images with Diffusion Models. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2335–2344 (2024).
8. Costa, P. et al. End-to-End adversarial retinal image synthesis. *IEEE Trans. Med. Imaging* **37**, 781-791 (2018).
9. Zhao, H., Li, H., Maurer-Stroh, S. & Cheng, L. Synthesizing retinal and neuronal images with generative adversarial Nets. *Med. Image Anal.* **49**, 14-26 (2018).
10. Zhao, H. et al. Supervised segmentation of Un-Annotated retinal fundus images by synthesis. *IEEE Trans. Med. Imaging* **38**, 46-56 (2019).
11. Saeed, A. Q., Sheikh Abdullah, S. N. H., Che-Hamzah, J., Abdul Ghani, A. T. & Abu-ain, W. A. k. Synthesizing retinal images using End-To-End VAEs-GAN Pipeline-Based sharpening and varying layer. *Multimed Tools Appl.* **83**, 1283-1307 (2024).
12. Beji, A., Blaiech, A. G., Said, M., Abdallah, A. B. & Bedoui, M. H. An innovative medical image synthesis based on dual GAN deep neural networks for improved segmentation quality. *Appl. Intell.* **53**, 3381-3397 (2023).
13. Keivanmarz, A., Sharifzadeh, H. & Fleming, R. Vein Pattern Visualisation using Conditional Generative Adversarial Networks. In *2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 1310–1316 (2020).
14. Abbott, R., Robertson, N. M., Rincon, J. M. & d.,Connor, B. Unsupervised object detection via LWIR/RGB translation. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 407–415 (2020).
15. Zhang, L., Gonzalez-Garcia, A., Weijer, J. & Danelljan, M. Synthetic data generation for End-to-End thermal infrared tracking. *IEEE Trans. Image Process* **28**, 1837-1850 (2019).
16. Uddin, M. S., Kwan, C. & Li, J. MWIRGAN: Unsupervised Visible-to-MWIR Image Translation with Generative Adversarial Network. *Electronics* **12**, 1039 (2023).
17. Illarionova, S., Shadrin, D., Trekin, A., Ignatiev, V. & Oseledets, I. Generation of the NIR spectral band for satellite images with convolutional neural networks. *Sensors* **21**, 5646 (2021).
18. Shukla, A., Upadhyay, A., Sharma, M., Chinnusamy, V. & Kumar, S. High-Resolution NIR Prediction from RGB Images: Application to Plant Phenotyping. In *2022 IEEE International Conference on Image Processing (ICIP)*, 4058–4062 (2022).
19. Lu, Y. & Lu, G. Bridging the Invisible and Visible World: Translation between RGB and IR Images through Contour Cycle GAN. In *2021 17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 1–8 (2021).
20. Lee, K. S. Improving the performance of automatic Lip-Reading using image conversion techniques. *Electronics* **13**, 1032 (2024).
21. Yan, L., Wang, X., Zhao, M., Liu, S. & Chen, J. A Multi-Model Fusion Framework for NIR-to-RGB Translation. In *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, 459–462 (2020).
22. Dou, H., Chen, C., Hu, X., Jia, L. & Peng, S. Asymmetric cyclegan for image-to-image translations with uneven complexities. *Neurocomputing* **415**, 114-122 (2020).
23. Sun, T., Jung, C., Fu, Q. & Han, Q. NIR to RGB domain translation using asymmetric cycle generative adversarial networks. *IEEE Access* **7**, 112459-112469 (2019).
24. Goodfellow, I. et al. Generative adversarial nets. In *Advances in neural information processing systems*, (2014).
25. Zhu, J. Y., Park, T., Isola, P. & Efros, A. A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, 2242–2251 (2017).
26. Radford, A., Metz, L. & Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. arXiv preprint at https://doi.org/10.48550/arXiv.1511.06434 (2016).
27. Martin, A. & Soumith, C. Wasserstein GAN. arXiv preprint at https://doi.org/10.48550/arXiv.1701.07875 (2017).
28. Isola, P., Zhu, J. Y., Zhou, T. & Efros, A. A. Image-to-Image Translation with Conditional Adversarial Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5967–5976 (2017).

29. Karras, T., Laine, S. & Aila, T. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4401–4410 (2019).
30. Karras, T. et al. Alias-free generative adversarial networks. In *Advances in neural information processing systems*, 852–863 (2021).
31. Kingma, D. Auto-Encoding variational Bayes. arXiv preprint at https://doi.org/10.48550/arXiv.1312.6114 (2013).
32. Van Den Oord, A., Kalchbrenner, N. & Kavukcuoglu, K. Pixel recurrent neural networks. In *International conference on machine learning*, 1747–1756 (2016).
33. Fujioka, T. et al. Virtual interpolation images of tumor development and growth on breast ultrasound image synthesis with deep convolutional generative adversarial networks. *J. Ultrasound Med.* **40**, 61-69 (2021).
34. Liu, X., Zou, Y., Xie, C., Kuang, H. & Ma, X. Bidirectional face aging synthesis based on improved deep convolutional generative adversarial networks. *Information* **10** (2019).
35. Dewi, C., Chen, R. C., Liu, Y. T., Jiang, X. & Hartomo, K. D. Yolo V4 for advanced traffic sign recognition with synthetic training data generated by various GAN. *IEEE Access* **9**, 97228-97242 (2021).
36. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V. & Courville, A. C. Improved training of wasserstein gans. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 5769–5779 (2017).
37. Chen, L. & Chan, H. Y. Generative Adversarial Networks With Data Augmentation and Multiple Penalty Areas for Image Synthesis. *Int Arab J Inf Techn* **20**, 428-434 (2023)
38. Fraz, M. M. et al. An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Trans. Biomed. Eng.* **59**, 2538–2548 (2012).
39. Han, Z., Ian, G., Dimitris, M. & Augustus, O. Self-Attention Generative Adversarial Networks. In *Proceedings of the 36th International Conference on Machine Learning* (eds Kamalika C. & Salakhutdinov R.), 7354–7363 (2019).
40. Kim, J., Kim, M., Kang, H. & Lee, K. U-GAT-IT: unsupervised generative attentional networks with adaptive Layer-Instance normalization for Image-to-Image translation. arXiv preprint at https://doi.org/10.48550/arXiv.1907.10830 (2019).
41. Matsuo, H. et al. Unsupervised-learning-based method for chest MRI–CT transformation using structure constrained unsupervised generative attention networks. *Sci. Rep.* **12**, 11090 (2022).
42. Komatsu, R. & Gonsalves, T. Multi-CartoonGAN with conditional adaptive Instance-Layer normalization for conditional artistic face translation. *AI* **3**, 37–52 (2022).
43. Li, X. et al. : Image-to-image translation via hierarchical style disentanglement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8639–8648 (2021).
44. Dai, Y., Fei, J., Huang, F. & Xia, Z. Face Omron ring: proactive defense against face forgery with identity awareness. *Neural Netw.* **180**, 106639 (2024).
45. Johnson, J., Alahi, A. & Fei-Fei, L. Perceptual losses for Real-Time style transfer and Super-Resolution. In *Computer Vision – ECCV 2016* (eds. Bastian Leibe B., Matas J., Sebe, N. & Welling M.), 694–711 (2016).
46. Zhong, Z., Zheng, L., Zheng, Z., Li, S. & Yang, Y. Camera Style Adaptation for Person Re-identification. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5157–5166 (2018).
47. Deng, Y. et al. Arbitrary style transfer via multi-adaptation network. In *Proceedings of the 28th ACM international conference on multimedia*, 2719–2727 (2020).
48. Huang, X. & Belongie, S. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE international conference on computer vision*, 1501–1510 (2017).
49. Yao, Y. et al. Attention-aware multi-stroke style transfer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1467–1475 (2019).
50. Liu, S. et al. Adaattn: Revisit attention mechanism in arbitrary neural style transfer. In *Proceedings of the IEEE/CVF international conference on computer vision*, 6649–6658 (2021).
51. Simonyan, K. Very deep convolutional networks for large-scale image recognition. arXiv preprint at https://doi.org/10.48550/arXiv.1409.1556 (2014).
52. Gatys, L. A. A neural algorithm of artistic style. arXiv preprint at https://doi.org/10.48550/arXiv.1508.06576 (2015).
53. Park, D. Y. & Lee, K. H. Arbitrary Style Transfer With Style-Attentional Networks. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5873–5881 (2019).
54. Staal, J., Abràmoff, M. D., Niemeijer, M., Viergever, M. A. & Van Ginneken, B. Ridge-based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imaging* **23**, 501-509 (2004).
55. Hoover, A., Kouznetsova, V. & Goldbaum, M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Trans. Med. Imaging* **19**, 203-210 (2000).
56. He, Z., Kan, M. & Shan, S. Eigengan: Layer-wise eigen-learning for gans. In *Proceedings of the IEEE/CVF international conference on computer vision*, 14408–14417 (2021).
57. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B. & Hochreiter, S. GANs trained by a two time-scale update rule converge to a local nash equilibrium. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 6629–6640 (2017).
58. Bińkowski, M., Sutherland, D. J., Arbel, M. & Gretton, A. Demystifying mmd gans. arXiv preprint at https://doi.org/10.48550/arXiv.1801.01401 (2018).
59. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2818–2826 (2016).

Author contributions

L.F. developed the concept and methodology, and drafted the original manuscript. H.S., H.L., and S.L. conducted investigations and provided resources. S.F. conducted investigations and curated data. M.C. worked on visualization. Y.L. curated data. J.C. conceptualized the study and supervised the work. F.C. conceptualized the study and methodology.

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval

This animal study was approved by the Animal Care Committee of the Laboratory Animal from Fudan University (Number:2020 华山医院 JS-646). Experiments in this study were performed in accordance with ARRIVE guidelines.

Additional information

Correspondence and requests for materials should be addressed to F.C.