Research article

A novel multi-modal fundus image fusion method for guiding the laser surgery of central serous chorioretinopathy

Jianguo Xu[1], Cheng Wan[2], Weihua Yang[3,*], Bo Zheng[4], Zhipeng Yan[3] and Jianxin Shen[1,*]

[1] College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
[2] College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
[3] The Affiliated Eye Hospital of Nanjing Medical University, Nanjing, 210029, China
[4] School of Information Engineering, Huzhou University, Huzhou, 313000, China

* Correspondence: Email: cadatc@nuaa.edu.com, benben0606@139.com.

Abstract: The angiography and color fundus images are of great assistance for the localization of central serous chorioretinopathy (CSCR) lesions. However, it brings much inconvenience to ophthalmologists because of these two modalities working independently in guiding laser surgery. Hence, a novel fundus image fusion method in non-subsampled contourlet transform (NSCT) domain, aiming to integrate the multi-modal CSCR information, is proposed. Specifically, the source images are initially decomposed into high-frequency and low-frequency components based on NSCT. Then, an improved deep learning-based method is employed for the fusion of low-frequency components, which helps to alleviate the tedious process of manually designing fusion rules and enhance the smoothness of the fused images. The fusion of high-frequency components based on pulse-coupled neural network (PCNN) is closely followed to facilitate the integration of detailed information. Finally, the fused images can be obtained by applying an inverse transform on the above fusion components. Qualitative and quantitative experiments demonstrate the proposed scheme is superior to the baseline methods of multi-scale transform (MST) in most cases, which not only implies its potential in multi-modal fundus image fusion, but also expands the research direction of MST-based fusion methods.

Keywords: central serous chorioretinopathy; image fusion; non-subsampled contourlet transform; generative adversarial network; pulse-coupled neural network

1. Introduction

CSCR is a common macular disease characterized by neurosensory retinal detachment (NRD) with or without pigment epithelium detachment (PED) [1,2], which often leads to blurred vision, central scotoma and deformed vision, and even results in permanent visual loss in serious cases. At present, laser surgery intervention is an effective measure for the treatment of CSCR, among which the angiography and color fundus images are the key auxiliary tools. As shown in Figure 1, the angiography image can clearly display the location and shape of the leakage areas (hereinafter referred to as lesions), but it has obvious color difference with the actual fundus field of view. On the contrary, the color fundus image cannot provide the location and shape information of the lesions, but it has the advantage of reflecting the actual color of fundus field of vision. In view of the characteristics of multi-modal fundus images, during the laser surgery procedure, ophthalmologists have to read the angiography image carefully to memorize the location and shape of the lesions, and then the corresponding lesion information will be marked on the color fundus image, and the marked positions of which are consistent with the angiography image. With the help of the marked color fundus image, the location and range of lesions in the real field of view are obtained in time by ophthalmologists through memory, which can provide guidance information for the location and area of laser surgery. However, it can be found that the multi-modal fundus images work independently when guiding the laser surgery of CSCR, that is, the ophthalmologists need to process the two kinds of images separately in order to get the final lesion information of real field of view, which not only increases the burden of ophthalmologists, but also causes inaccurate lesion information due to the cross-modal hand-marked deviation. Therefore, by utilizing multi-modal fundus information fusion strategy, we attempt to provide more accurate lesion information for the laser surgery of CSCR and simultaneously reduce the pressure of ophthalmologists. To the best of our knowledge, the multi-modal fundus image fusion for CSCR has not been done yet. Fortunately, the research of natural image fusion and brain image fusion has been widely carried out, which can provide some inspiration for the work of this paper.
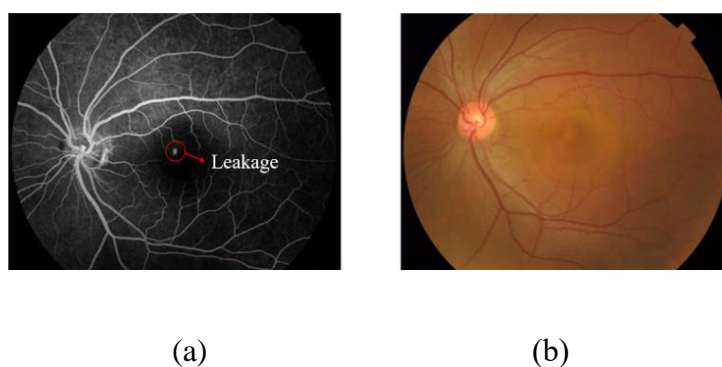

(a)   (b)

Figure 1. The multi-modal fundus images. (a) Angiography image; (b) Color fundus image.

Generally, image fusion can be divided into three categories, namely, pixel level, feature level, and decision level, of which the pixel-level image fusion performs directly on the source images and can retain more information. Next, the related research of this fusion scheme is described briefly. The pixel-level image fusion is mainly composed of spatial domain and transform domain [3]. In terms of the former, it aims to synthesize an image using techniques such as averaging the source images [4],

the principal component analysis [5], the intensity-hue-saturation transform [6] and so on. Although spatial domain methods achieve better results, spectral degradation and contrast reduction [7] may be followed. Then the latter, namely transform domain, has emerged as an excellent fusion way to alleviate the above issues. The representation learning-based and multi-scale transform (MST) based methods are typical representatives of this fusion scheme. Yang et al. [8] applied the sparse representation (SR) theory to image fusion field. Specifically, the sparse coefficients of source images were firstly gained using an overcomplete dictionary. A choose-max fusion rule was then employed to combine the coefficients. With the help of global dictionary, the final fused image can be reconstructed accordingly. Li et al. [9] presented a low-rank representation-based fusion method. The low-rank representation was used instead of sparse representation to extract features from source images, and after that the L1-norm and the same fusion strategy as [8] were jointly adopted to reconstruct the fused image. In addition, the MST-based methods have marked predominance in multi-modal image fusion, where the techniques such as discrete wavelet transform (DWT) [10], Shearlet [11], Contourlet transform (CT) [12], NSCT [13] and non-subsampled shearlet transform (NSST) [14] are frequently applied to decompose the source images into high-frequency and low-frequency components. In particular, the NSCT has achieved great success in multi-modal image fusion by combining with various methods. Yong et al. [15] utilized a type-2 fuzzy logic-based rule to fuse the high-frequency components and a local energy algorithm to combine the low-frequency components, and the inverse NSCT was then followed to construct the final brain image. Similarly, a multi-modal image fusion method [16] combined with SR and PCNN was raised based on the NSCT decomposition. Especially, they employed the number of ignition times of PCNN to adjust the fusion parameters for the high-frequency components and acquired more detailed information. By applying the maximum entropy of square of the coefficients and the maximum weighted sum-modified laplacian in NSCT domain, respectively, Ganasala et al. [17] achieved the fusion of different components successfully. A fuzzy-adaptive reduced PCNN-based fusion way in NSCT domain was presented by Das et al. [18], the difference of which was that the fusion of high-frequency components and low-frequency components adopted the same method, namely using PCNN-based strategy with fuzzy-adaptive linking strengths.

NSCT is a shift-invariant, multiscale, and multi-direction expansion with fast implementability [19]. However, it cannot be better competent for specific image fusion tasks independently and may cause the degradation of image quality. Therefore, the combination with other fusion techniques like [15–18] is a common solution for NSCT and have achieved promising performance. Nevertheless, in order to make the fusion result possess both the content and details of the source images and shield the noise to the greatest extent, manually designing the fusion rules of different components plays an important role in MST-based methods, which makes these methods complicated and tediously. Fortunately, deep learning provides an effective solution for this problem and some research has preliminarily proved its potential. Hou et al. [20] put forward an unsupervised image fusion framework for adaptively extracting, fusing and reconstructing the deep features from the source images, which shows the applicability in infrared and visible image fusion task. Li et al. [21] also explored the deep learning-based method for image fusion, but only for extracting the multi-layer features of source images. After that the max selection strategy was still carried out on the candidate features generated by the L1-norm and weighted-average rule. Moreover, a fusion scheme [22] based on the generative adversarial network (GAN), was intelligently designed. In their method, the idea of generative adversarial training was cleverly integrated into the image fusion process where the fusion parameters of the source images were obtained in an end-to-end way, and achieved gratifying performance.

The aforementioned deep learning methods are beneficial to alleviate the multi-modal image fusion issue in NSCT domain, among which the GAN-based method has given us the most inspiration. In view of the dilemma of information acquisition of CSCR lesions in the laser surgery and the solid research foundation of image fusion mentioned above, this paper attempts to seek an adaptive parameter fusion method without too much manual intervention for multi-modal CSCR fundus image fusion and aims to obtain an assistant tool with the characteristics of both the angiography and color fundus images. The significance of this work is not only to relieve the pressure of ophthalmologists to capture the lesion information of CSCR in traditional laser surgery, but also to further facilitate the safety and efficiency of automatic laser surgery by providing clearer image data. In general, the main contributions of our research are as follows: 1) To the best of our knowledge, the image fusion method is firstly employed in the multi-modal CSCR fundus image scene for promoting to obtain more accurate information of the CSCR lesions, which provides a feasible scheme for better guiding the laser surgery of CSCR. 2) The data set of low-frequency and high-frequency components for the angiography and color fundus images is established though NSCT, which lays the foundation for better boosting the fusion of image content (structure) and detail information in NSCT domain. 3) An improved deep learning-based fusion method is employed to obtain the adaptive fusion parameters of low-frequency components in an end-to-end learning way, which enhances the smoothness of the fused images. Additionally, the PCNN-based fusion scheme is adopted to combine the high-frequency components, which facilitates the integration of detailed information. 4) Qualitative and quantitative experiments are performed on the multi-modal CSCR fundus images to verify the effectiveness of the proposed image fusion method in NSCT domain.

The rest of the paper is organized as follows. Section 2 presents the related work of NSCT, GAN and PCNN. Section 3 describes the details of the proposed image fusion method. Section 4 shows the experiments and discussions. Section 5 concludes this paper.

2. Related work

2.1. Non-subsampled contourlet transform

NSCT is a multi-scale transform technique developed on the basis of CT, which can effectively suppress pseudo-Gibbs phenomena [23]. Driven by non-subsampled pyramid filter bank (NSPFB) and non-subsampled directional filter bank (NSDFB), NSCT achieves multi-scale and multi-directional successfully [13]. Specifically, on the one hand, NSPFB can complete the iterative decomposition task on the lower-frequency component obtained by the previous decomposition, which ensures the multi-scale characteristics of NSCT. And on the other hand, NSDFB performs multi-directional decomposition with l directional levels on the high-frequency components, which can make NSCT obtain 2^l sub-images with more precise directional details on the current scale. Besides, due to the removal of down-sampling and up-sampling operations in the process of multi-scale transform and multi-directional decomposition processes, NSCT has the advantage of shift-invariance property. It is also the above superiorities that make NSCT shine in a variety of image fusion scenes.

2.2. Generative adversarial network

GAN [24] enhances its ability to capture data distribution by skillfully establishing an adversarial

game between the generator G and the discriminator D. Since its birth, it has been widely used in various visual tasks [25,26]. Yu et al. [25] effectively enhanced the capability of cervical cell detection model by introducing images generated from GAN (i.e., data augmentation) when training the convolutional neural network. In the aspect of image fusion, with the help of adversarial training, the integration of infrared and visible images was achieved successfully [26]. Whether it is the process of data augmentation or image fusion, the application of GAN are essentially a minimax two-player game between G and D from a mathematical point of view, which can be described as the following formula:

$$\min_{G} \max_{D} V_{GAN}(G, D) = E_{x \sim P_{data}(x)}[log\ D(x)] + E_{z \sim P_z(z)}[log\ (1 - D(G(z)))] \tag{2.1}$$

where x is a sample from a real distribution $P_{data}(x)$ and z denotes a sample from a concrete prior distribution $P_z(z)$. Because it is difficult to express $P_{data}(x)$ explicitly, so GAN tries its best to fit the distribution of $P_{data}(x)$ by fooling D using the generated samples based on z. The whole process can be further described separately as follows:

$$D^* = \max_{D} V_{GAN}(G, D) = E_{x \sim P_{data}(x)}[log\ D(x)] + E_{z \sim P_z(z)}[log\ (1 - D(G(z)))] \tag{2.2}$$

$$G^* = \min_{G} V_{GAN}(G, D) = E_{z \sim P_z(z)}[log\ (1 - D(G(z)))] \tag{2.3}$$

where G^* and D^* represent the optimal G and D, respectively. Alternating optimization drives G to generate more realistic data distribution close to $P_{data}(x)$, and the capability of D to distinguish true from false is gradually enhanced. Once the adversarial training reaches equilibrium, the final generative model can be applied in the corresponding scene.

2.3. Pulse-coupled neural network

PCNN is an artificial neural network that simulates the information processing mechanism of mammalian visual cortex, which processes the input signal autonomously through the cooperation of three sub-modules, namely, the receiving field, modulation field and pulse generator module, and has been well implemented in the field of image fusion [27,28]. To perform object detection, the PCNN-based image fusion network [27] was firstly chosen as an architecture for the fusion of the original and filtered images, and then the target can be highlighted consequently and thus be easily detected. In order to enhance the ability to capture edge information, Tan et al. [28] properly introduced multi-scale morphological gradient into the PCNN-based fusion method to improve its performance in remote sensing image fusion task. The same technique is also used in multi-modal brain image fusion task for integrating the information from the fine-structure and coarse-structure layers [29]. It can be found that PCNN has performed well in the above fields, especially in the fusion task, which also encourages us to yearn for it in our multi-modal CSCR fundus image fusion scene.

3. Proposed methodology

The CSCR image fusion scheme proposed in this paper is an innovative application of GAN-based and PCNN-based image fusion methods in NSCT domain. Firstly, NSCT decomposes the CSCR images into high-frequency and low-frequency components, which promotes the smooth fusion of information in the angiography and color fundus images. Secondly, the GAN-based technique is

employed for the fusion of low-frequency components, which can determine the adaptive fusion parameters in a learning way and greatly avoid the manual intervention. Simultaneously, PCNN is adopted to fuse the high-frequency components, which can effectively integrate the edge and contour information from the two modal images. Finally, the required fused CSCR image can be obtained by applying a color conversion operation on the fused image Y_{AB} achieved based on the inverse transform on the above fusion components. The flow chart of the proposed method is shown in Figure 2.
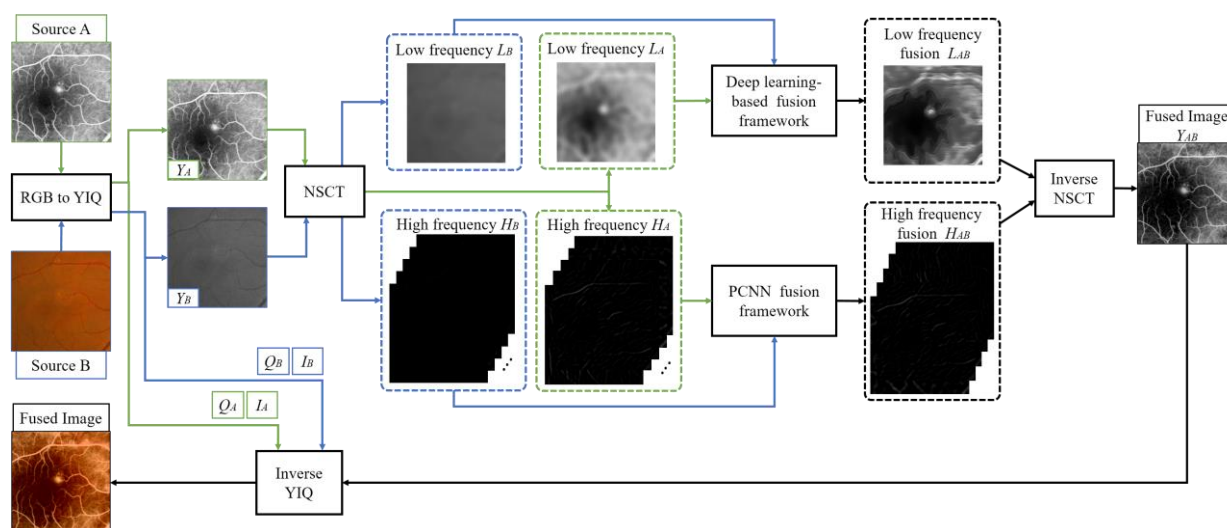

Figure 2. An overview of the proposed method.

3.1. Preprocessing

It should be noted that the preprocessing operation needs to be done before the fusion task is formally executed, the parts of which mainly involve the registration of multi-modal fundus images, the clipping of macular region, and the establishment of data set. Generally, registration is used for the alignment of sequence images or multi-modal images to ensure the consistency of spatial information of image content, which is very important for the analysis of medical images [30]. Considering the accuracy and efficiency of registration, the scheme in [31] is adopted in our CSCR fundus image registration task. After that, in order to speed up the fusion of multi-modal information, the macular region is roughly localized by clipping, which can limit the image fusion range to the occurrence area of CSCR. Following that is the color space conversion and NSCT operation, among which the YIQ color conversion is applied to separate the brightness and color channels and followed by the NSCT operation on the brightness channel to establish the data set based on the high-frequency and low-frequency components. Furthermore, the low-frequency components will be further divided into patches with size of 24 in this paper, which can not only expand the training samples, but also reduce the training cost.

3.2. Deep learning-based fusion framework

In recent years, deep learning techniques have been involved in various image fusion scenes and achieved good performance [20–22,26]. Among them, the GAN-based fusion schemes [22,26] have shown great advantages, which provides a new solution for the determination of adaptive fusion

parameters. This section is the work extension of GAN-based image fusion scheme (as shown in Figure 3), which is embodied in three aspects, namely, a new attempt in application domain, a new incorporation in the MST domain and a new adjustment in technique domain. Firstly, we apply the GAN-based fusion scheme to the multi-modal CSCR fundus image information integration task, which broadens its application field and scope. Secondly, the scheme is intriguingly integrated into the MST domain, aiming at capturing the data distribution of the low-frequency components in an end-to-end learning way and driving the generation of adaptive fusion parameters to enhance the smoothness of the fused images. Thirdly, due to the computational cost of NSCT, not all details are filtered into the high-frequency components [13], which means that the structure and detailed information need to be processed synchronously in the fusion of low-frequency components. To this issue, the cycle-consistency loss term, inspired by [32], is integrated into the deep learning-based fusion scheme to further enhance the learning ability of GAN model for low-frequency component data distribution.
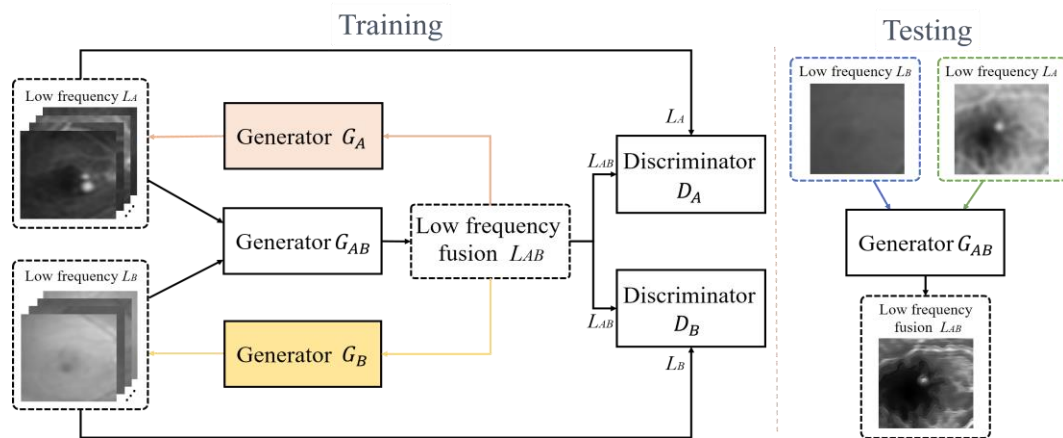

Figure 3. The deep learning-based fusion framework.

Considering the loss of information, the up-sampling and down-sampling operations in [22] are removed in our work and other network structures remain unchanged. Finally, the objective optimization framework corresponding to Eq (2.1) is as follow:

$$
\min_{G_{AB}} \max_{D_A, D_B} V_{GAN}(G, D_A, D_B) = E[\log D_A(L_A)] + E[\log (1 - D_A(G_{AB}(L_A, L_B)))]
$$
$$
+ E[\log D_B(L_B)] + E[\log (1 - D_B(G_{AB}(L_A, L_B)))]
$$
(3.1)

where G_{AB} is the generator for obtaining the adaptive fusion parameters and generating fusion results, and D_A and D_B denote the discriminators that drive G_{AB} to capture the data distributions of the low-frequency components L_A and low-frequency component L_B synchronously. Similarly, corresponding to Eqs (2.2) and (2.3), we define the loss functions for D_A, D_B and G_{AB} as follows:

$$
Loss\_D_A = E[\log D_A(L_A)] + E[\log (1 - D_A(G_{AB}(L_A, L_B)))]
$$
(3.2)

$$
Loss\_D_B = E[\log D_B(L_B)] + E[\log (1 - D_B(G_{AB}(L_A, L_B)))]
$$
(3.3)

$$
Loss\_G_{AB} = E[\log (1 - D_A(G_{AB}(L_A, L_B)))] + E[\log (1 - D_B(G_{AB}(L_A, L_B)))]
$$
(3.4)

As mentioned above, we adopt the cycle-consistency term to improve the performance of GAN model, that is to say, to enhance the learning ability of the generator G_{AB} for low-frequency component data distribution. Specifically, through two generators with the same encoding and decoding structures as G_{AB}, the fused image is translated back to the corresponding modals respectively to force G_{AB} to grasp the data distribution of the source images in detail. The cycle-consistency loss can be written as follows:

$$Loss_{AB \to A} = E[\|G_A(G_{AB}(L_A, L_B)) - L_A\|_1] \tag{3.5}$$

$$Loss_{AB \to B} = E[\|G_B(G_{AB}(L_A, L_B)) - L_B\|_1] \tag{3.6}$$

$$Loss\_cycle = \eta Loss_{AB \to A} + \beta Loss_{AB \to B} \tag{3.7}$$

where G_A and G_B represent the generators of corresponding modals respectively, and $\|\cdot\|_1$ is the L1-norm term that reduces blurs in the fused images. Then, the total loss function can be defined as follows:

$$Loss\_T = Loss\_G_{AB} + Loss\_D_A + Loss\_D_B + \lambda Loss\_cycle \tag{3.8}$$

where the η, β and λ denote the trade-off parameters adjusting the proportion of the corresponding term, which are set to 0.05, 1.5 and 0.8, respectively. Then, the final optimized framework Eq (3.1) can be rewritten as:

$$G_{AB}^*, D_A^*, D_B^* = \min_{G_{AB}} \max_{D_A, D_B} Loss\_T \tag{3.9}$$

where G_{AB}^*, D_A^* and D_B^* represent the optimal G_{AB}, D_A and D_B respectively. In practice, the image fusion method based on GAN is not stable in the training process. In order to reduce the model oscillation, this paper follows the training strategy in [22], that is, the training quota is allocated according to the performance of the generators and discriminators instead of training in turns.

3.3. PCNN fusion framework

The existing PCNN model is developed on the basis of biological neuron model [33]. Due to simulating the information processing mechanism of mammalian visual cortex, the signal processed by the model is closer to the characteristics of human vision. However, setting a large number of parameters is a defect of the traditional PCNN model. Therefore, a simplified PCNN [34] model is introduced into our multi-modal CSCR fundus image fusion research, which is specifically used for the information integration of the high-frequency components in this section. The expressions of the simplified PCNN model can be defined as follows:

$$F_{ij}^A = S_{ij}^A, \quad F_{ij}^B = S_{ij}^B \tag{3.10}$$

$$L_{ij}^A(n) = e^{-a_L^A} L_{ij}^A(n-1) + V_L^A \sum W_{ijkl}^A Y_{kl}^A(n-1), \quad L_{ij}^B(n) = e^{-a_L^B} L_{ij}^B(n-1) + V_L^B \sum W_{ijkl}^B Y_{kl}^B(n-1) \tag{3.11}$$

where S_{ij}^A and S_{ij}^B denote the external stimuli determined by the high-frequency components in this paper, and F_{ij}^A and F_{ij}^B are the feedbacks input of the above inputs respectively. L_{ij}^A and L_{ij}^B represent the linking parameters of neurons, and the corresponding amplification and attenuation

coefficients are a_L^A, V_L^A, a_L^B, V_L^B, among which a_L^A equals a_L^B and V_L^A equals V_L^B in this paper. Y_{kl}^A and Y_{kl}^B are the corresponding outputs of neuron (k, l). W_{ijkl}^A and W_{ijkl}^B are the connection matrixes between neuron (i, j) and neuron (k, l) in the respective low-frequency components, and they can be express as follow:

$$W_{ijkl}^A = \frac{1}{\sqrt{(i^A - k^A)^2 + (j^A - l^A)^2}}, \quad W_{ijkl}^B = \frac{1}{\sqrt{(i^B - k^B)^2 + (j^B - l^B)^2}} \tag{3.12}$$

where (i^A, j^A) and (i^B, j^B) denote the current neurons in the high-frequency component A and the high-frequency component B respectively, and (k^A, l^A) and (k^A, l^A) are the corresponding neighbor neurons.

$$U_{ij}^A(n) = F_{ij}^A(1 + \beta^A L_{ij}^A(n-1)), \quad U_{ij}^B(n) = F_{ij}^B(1 + \beta^B L_{ij}^B(n-1)) \tag{3.13}$$

$$E_{ij}^A(n) = e^{-a_E^A} E_{ij}^A(n-1) + V_E^A Y_{ij}^A(n-1), \quad E_{ij}^B(n) = e^{-a_E^B} E_{ij}^B(n-1) + V_E^B Y_{ij}^B(n-1) \tag{3.14}$$

$$Y_{ij}^A(n) = \begin{cases} 1, & if \ E_{ij}^A(n) \le U_{ij}^A(n) \\ 0, & else \end{cases}, \quad Y_{ij}^B(n) = \begin{cases} 1, & if \ E_{ij}^B(n) \le U_{ij}^B(n) \\ 0, & else \end{cases} \tag{3.15}$$

where U_{ij}^A and U_{ij}^B denote the modulation output results, and β^A and β^B are the link strength factors that take the same value. E_{ij}^A and E_{ij}^B are dynamic thresholds used to adjust the firing frequency of neurons, and the corresponding amplification and attenuation coefficients are a_E^A, V_E^A, a_E^B, V_E^B, among which a_E^A equals a_E^B and V_E^A equals V_E^B in this paper. Then, based on the step function, the pulse value Y_{ij}^A and Y_{ij}^B of PCNN model can be obtained. Finally, we acquire the fusion map of the high-frequency components by the following formula:

$$F_A = \sum_{n=1}^{N} Y_{ij}^A(n), \ F_B = \sum_{n=1}^{N} Y_{ij}^B(n), \ F_{map} = \begin{cases} 1, & if \ F_A \ge F_B \\ 0, & else \end{cases} \tag{3.16}$$

where F_A and F_B denote the sum of the pulses with N iterations. Based on Eq (3.16), the final fusion result of the high-frequency components can be obtained, that is, H_{AB} shown in Figure 4.

3.4. The inverse transformation and color space conversion

Through Sections 3.2 and 3.3, the low-frequency components and high-frequency components of multi-modal CSCR fundus images can be integrated by deep learning-based fusion method and PCNN-based fusion method respectively, which greatly reduces the manual intervention of fusion parameters and strengthens the integration of detail information. On this basis, the preliminary fusion results can be obtained by performing the NSCT inverse transform on the low-frequency fusion and high-frequency fusion results, followed by the YIQ-based color space conversion on the above results to obtain the final fusion image based on the RGB color space.
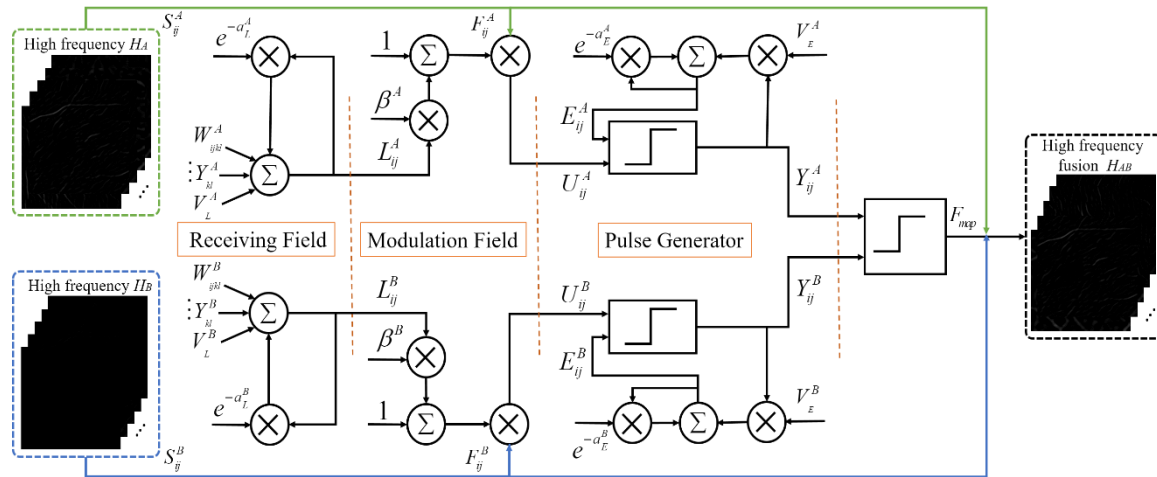
Figure 4. The PCNN fusion framework.

4. Experimental results and discussions

4.1. Datasets and experimental settings

In the task of multi-modal CSCR fundus image fusion in this paper, a total of 396 angiography and color fundus image pairs are used to verify the performance of the proposed image fusion method based on NSCT. By relying on the NVIDIA-2080ti GPU with the Tensorflow framework, 356 image pairs are utilized to train the deep learning-based fusion method and the remaining image pairs are used for testing. Moreover, in order to expand the training samples and reduce the training cost, the low-frequency components are further divided into 297,457 patches with the size of 24. Meanwhile, in the training process, the learning rate, batch size and learning decay rate are set to 0.0002, 64 and 0.9 respectively, and the size of all test images is adjusted to 256 to ensure the fusion effect of the whole CSCR area in the test process.

4.2. Quantitative criteria

Quantitative analysis of image fusion results is one of the important ways to evaluate the fusion methods effectively. In this paper, six commonly used evaluation indexes are applied to analyze the fusion effect of multi-modal CSCR fundus image, namely, average gradient (AG), entropy (EN) [35], mutual information (MI) [36], Qabf [37], Nabf [38], structural similarity (SSIM) [39]. The details are as follows:

• AG: AG is the average value of the image gradient, which represents the clarity of the fused image and can be expressed as follow:

$$AG = \frac{1}{(M-1)\times(N-1)} \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \sqrt{((F(i+1,j)-F(i,j))^2 + (F(i,j+1)-F(i,j))^2)/2} \qquad (4.1)$$

where F denotes the final fused image, and M and N represent the size of F.

• EN: EN is an objective evaluation to measure the amount of information contained in the fused image. The larger the value is, the more information the fused image carries. The formula is as follow:

$$EN = -\sum_{l=1}^{W} p(l)\log_2 p(l) \tag{4.2}$$

where W denotes all gray levels in the final fused image and $p(l)$ represents the proportion of gray value l to all gray levels.

• MI: MI can indicate how much information the fused image carries about the source images, and the greater the MI value is, the better the fusion effect is. The formula of MI can be expressed as follow:

$$MI = MI_{AF} + MI_{BF} \tag{4.3}$$

$$MI_{AF} = \sum_{l}^{W}\sum_{p}^{W} p_{AF}(l,p)\log_2\left(\frac{p_{AF}(l,p)}{p_A(l)p_F(p)}\right), \quad MI_{BF} = \sum_{l}^{W}\sum_{p}^{W} p_{BF}(l,p)\log_2\left(\frac{p_{BF}(l,p)}{p_B(l)p_F(p)}\right) \tag{4.4}$$

where MI_{AF} and MI_{BF} denote the mutual information between the source image A and the fused image F, the source image B and the fused image F, respectively. $p_{AF}(l,p)$ and $p_{BF}(l,p)$ are the corresponding joint distribution of gray levels.

• Qabf and Nabf: Qabf reflects the quality of the visual information obtained from the input image, a higher value of which means that the quality of the fused image is better. Details of the formula can be found in [37]. Nabf represents the noise or artifacts added in fused image due to fusion process, a smaller value of the index indicates a higher image quality. Refer to reference [38] for details.

• SSIM: SSIM is a measure to characterize the structural similarity between the fused image and the source images. A larger value means that the fused image is more consistent with the source images in terms of structure. The formula of SSIM is shown as follow:

$$SSIM = (SSIM_{AF} + SSIM_{BF})/2 \tag{4.5}$$

$$SSIM_{AF} = \frac{(2\mu_A\mu_F + C_1)(2\sigma_{AF} + C_2)}{(\mu_A^2 + \mu_F^2 + C_1)(\sigma_A^2 + \sigma_F^2 + C_2)}, \quad SSIM_{BF} = \frac{(2\mu_B\mu_F + C_1)(2\sigma_{BF} + C_2)}{(\mu_B^2 + \mu_F^2 + C_1)(\sigma_B^2 + \sigma_F^2 + C_2)} \tag{4.6}$$

where $SSIM_{AF}$ and $SSIM_{BF}$ denote the structural similarity between the source image A and the fused image F, the source image B and the fused image F, respectively. μ_A, μ_B and μ_F are the corresponding mean intensity respectively. σ_A and σ_B are the standard deviations, and σ_{AF} and σ_{BF} denote the covariances between the source images and the fused image F respectively. $C_1 = (K_1 * 255)^2$ and $C_2 = (K_2 * 255)^2$ are two constants to prevent an unstable results, and K_1 and K_2 are set to 0.01 and 0.03 respectively.

4.3. Comparison of the proposed method with or without the cycle-consistency module

In this section, we conduct the ablation experiment to verify the effectiveness of the cycle-consistency loss module in the GAN-based fusion method for the low-frequency components. The results of quantitative comparison are shown in Figure 5.

Significantly, in the five evaluation indexes, namely, EN, Qabf, MI, Nabf and SSIM, the deep learning-based fusion scheme equipped with the cycle-consistency loss module has achieved good results in most cases. Especially the MI index, in the absence of the cycle-consistency loss module, the

final fusion images obtain less information from the source images, which may lead to the lack of key lesion information and is very unfavourable for the multi-modal CSCR fundus image fusion task. In addition, in terms of the AG index, it can be found that the low-frequency fusion scheme integrated with the cycle-consistency loss module will be slightly inferior to that without the same module, which means a slight difference in image clarity. However, on the basis of other indexes, it can be observed that this is at the cost of image information reduction and noise increase, which is obviously not desirable.
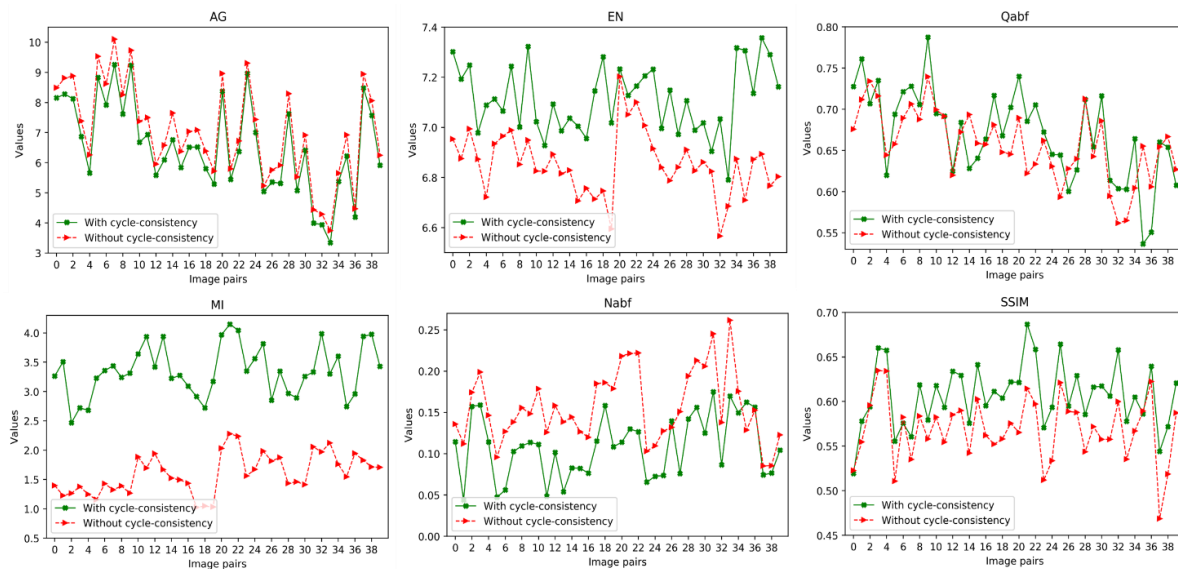
Figure 5. The quantitative comparison of the proposed method with or without the cycle-consistency module.
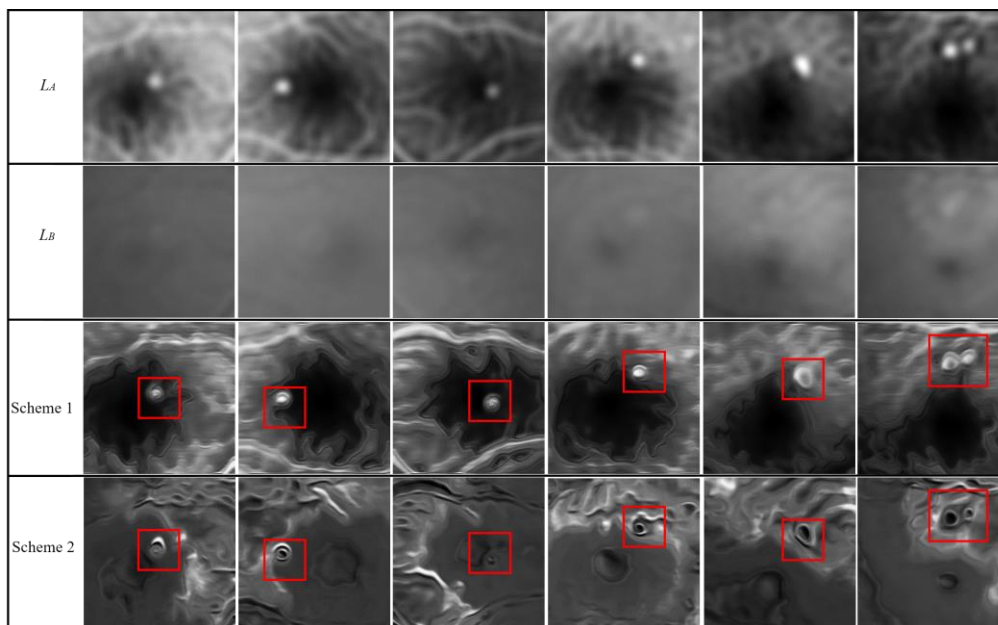

Figure 6. The fusion results of low-frequency components based on the GAN method.

Besides, we show the effect of the cycle-consistency loss module in the fusion scheme based on

deep learning through the specific low-frequency fusion experiments. As shown in Figure 6, the first line and the second line are the low-frequency components of the angiography and color fundus images, represented by L_A and L_B respectively. The third line and the fourth line denote the GAN-based fusion method with and without the cycle-consistency loss module respectively, represented by Scheme 1 and Scheme 2. Under the same experimental setup, Scheme 1 can better learn the data distribution of L_A and L_B, and the final low-frequency fusion results not only contain normal structural information, but also better retain the lesion content. But the performance of Scheme 2 is on the contrary. Specifically, the low-frequency fusion images under this scheme are generally not smooth enough, and the fusion quality of the lesion information is poor, which seriously affects the ophthalmologists' judgment of the shape and location of the lesions, and thus cannot guide the CSCR laser surgery well.
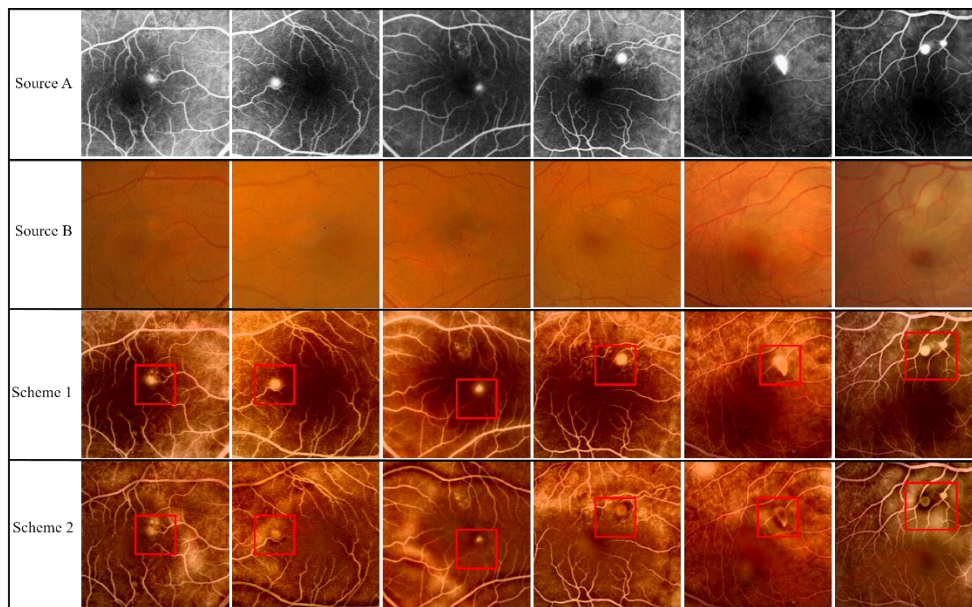

Figure 7. The fused CSCR images of the proposed method equipped with or without the cycle-consistency loss module.
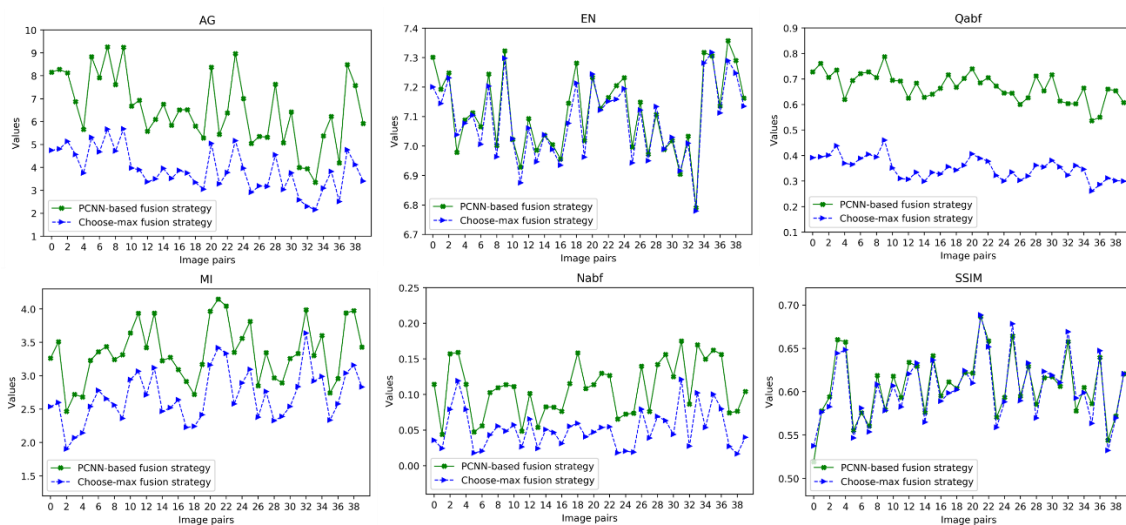

Figure 8. The quantitative comparison of the PCNN-based and choose-max fusion strategies.

As shown in Figure 7, it further conveys the difference of the final fusion results obtained by the proposed fusion method using the above two low-frequency fusion schemes. It can be clearly observed that part of the lesion information is lost in Scheme 2 and the vessels and macular area are not smooth enough, which are consistent with the performance in Figure 6 and confirm the necessity of introducing the cycle-consistency loss module. On the contrary, Scheme 1 not only captures the distribution of lesions in the source images, but also better integrates the normal structural information. More importantly, the transition of each region of the fused image is natural and smooth, which can provide better reference for ophthalmologists to identify the lesion information.

4.4. Comparison of the fusion strategies for high-frequency components

After the previous section, the importance of the cycle-consistency loss module in the multi-modal CSCR fundus image fusion task is reflected. On this basis, this section further discusses the fusion strategy of high-frequency components, specifically involving the PCNN-based fusion strategy (i.e., strategy 1) and the commonly used choose-max fusion strategy (i.e., strategy 2). On the premise that the GAN-based fusion scheme is the low-frequency component fusion strategy, the results shown in Figure 8 can be obtained by using the above two high-frequency component fusion strategies. Apparently, except for the Nabf index, the PCNN-based fusion strategy can make the proposed fusion method more attractive in the task of this paper. In particular, the three indexes, namely AG, Qabf and MI, show that the strategy can promote the whole fusion method to obtain high-quality and high-definition fusion results which inherit more information from the source images. Admittedly, the choose-max fusion strategy can indeed make the fusion results avoid too much noise, but it cannot make up for the losses of other aspects of the fusion results, such as the image clarity, image quality and the source image information carried by the fused images.
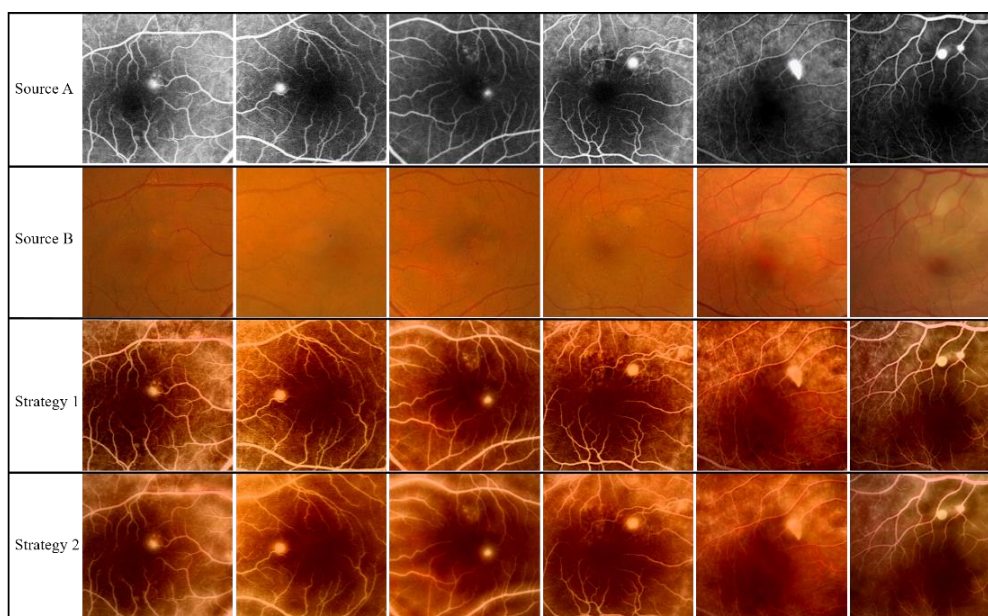

Figure 9. The fused CSCR images based on the two strategies.

In addition, corresponding to the above analysis, it can be clearly observed from Figure 9 that the fused images obtained by strategy 1 are significantly better than those obtained by strategy 2 in terms

of clarity. Moreover, the latter results in a weak contrast of image fusion results, which makes it difficult to distinguish the fundus background area, normal structure and lesion information to a certain extent. On the contrary, the performance of strategy 1 in this aspect is commendable. The above discussions prove the applicability and feasibility of the PCNN-based fusion strategy in the multi-modal CSCR fundus image fusion task in this paper, which also lay the foundation for the combination of the PCNN-based and deep learning-based fusion strategies.

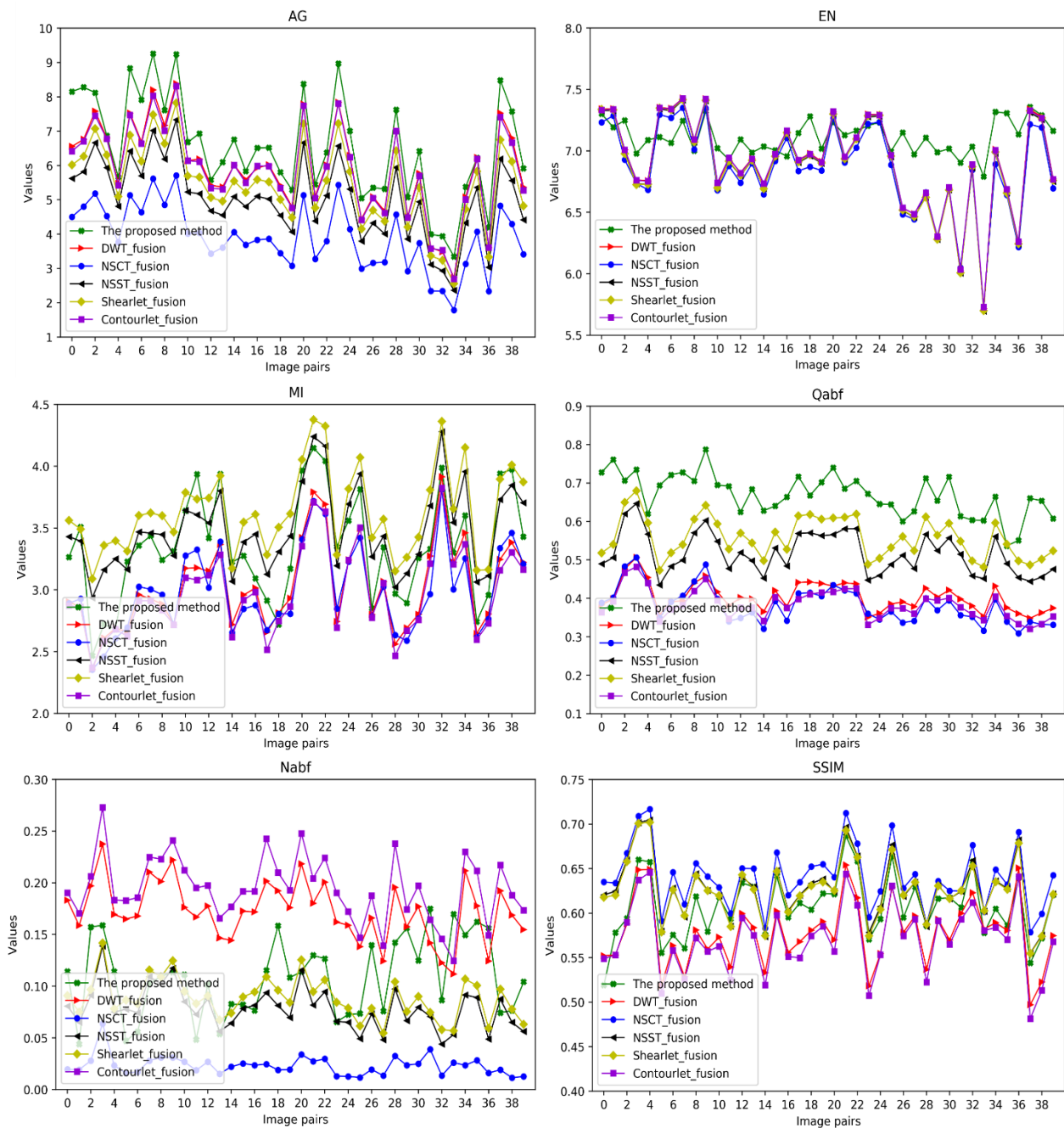## 4.5. Comparison of the proposed and baseline methods

Figure 10. The quantitative comparison of the proposed and baseline methods.

In this part, the proposed multi-modal CSCR fundus image fusion method in NSCT domain is compared with the MST baseline methods, and the fusion techniques of low-frequency components and high-frequency components in the baseline methods adopt the common fusion strategies [8,40]. Additionally, both the proposed and baseline methods perform image fusion tasks in the Y channel of YIQ color space and acquire the final fusion images after the same color space conversion operation. The results of quantitative comparison are shown in Figure 10. It can be found that the AG, EN and Qabf values of the proposed method are better than the baseline methods in most cases. Among them, the AG and Qabf indexes have significant advantages, which imply that the fused images are clearer and of good quality. Compared with the Shearlet and NSST fusion methods, it should be pointed out that the proposed fusion scheme is competitive in terms of the MI index. Furthermore, although there are still some gaps in the Nabf and SSIM indexes between our proposed fusion method and some baseline schemes, this slight difference of which also confirms the potential of the proposed method. At the same time, in order to display the fusion performance of multi-modal CSCR fundus images with different fusion methods more conveniently, the average fusion indexes of 40 pairs of CSCR images are specially counted and summarized in Table 1. The quantitative results in the table show that the proposed method is superior to other baseline fusion methods in the average values of AG, EN and Qabf and projects its competitiveness in most cases in other indexes, which is consistent with the above analysis.

Table 1. The average value of each index of the proposed and baseline fusion methods.

	DWT	NSCT	NSST	Shearlet	Contourlet	Proposed
AG	5.9231	3.9179	5.0483	5.4747	5.8684	**6.5489**
EN	6.9128	6.8538	6.8928	6.8963	6.9133	**7.1129**
MI	3.0255	2.9946	3.4880	**3.6318**	2.9755	3.3504
Qabf	0.4039	0.3823	0.5174	0.5583	0.3843	**0.6699**
Nabf	0.1730	**0.0233**	0.0786	0.0883	0.1946	0.1083
SSIM	0.5776	**0.6403**	0.6273	0.6260	0.5705	0.6067

After the same color space conversion, the MST baseline methods and the proposed method can obtain the final CSCR fusion images. The specific fusion results are shown in Figure 11. From the perspective of image clarity, the results gained by the proposed fusion method based on NSCT are relatively clear, and the contours of vessels and lesions are easy to identify. From the perspective of contrast, the proposed method can make the final fusion images have better contrast, and make the vessels, macular area and background easy to distinguish. From the perspective of preserving image information, all the methods can achieve the fusion task of the angiography and color fundus images, but on the whole our scheme can better deal with the fusion of structure and detail information of the source images.

Generally speaking, compared with the MST baseline fusion methods, the proposed fusion scheme based on NSCT can be able to better realize the information integration of multi-modal CSCR fundus images, which shows its potential value. The fusion scheme not only innovatively realizes the combination of the deep learning-based and PCNN-based fusion techniques, but also provides a new

idea for the research of fusion rules based on MST fusion methods. Moreover, through the information integration of multi-modal fundus images, such an image fusion scheme provides a more convenient auxiliary tool characterized by similar fundus color and clear lesion and structure information for guiding the operation of traditional CSCR laser surgery.
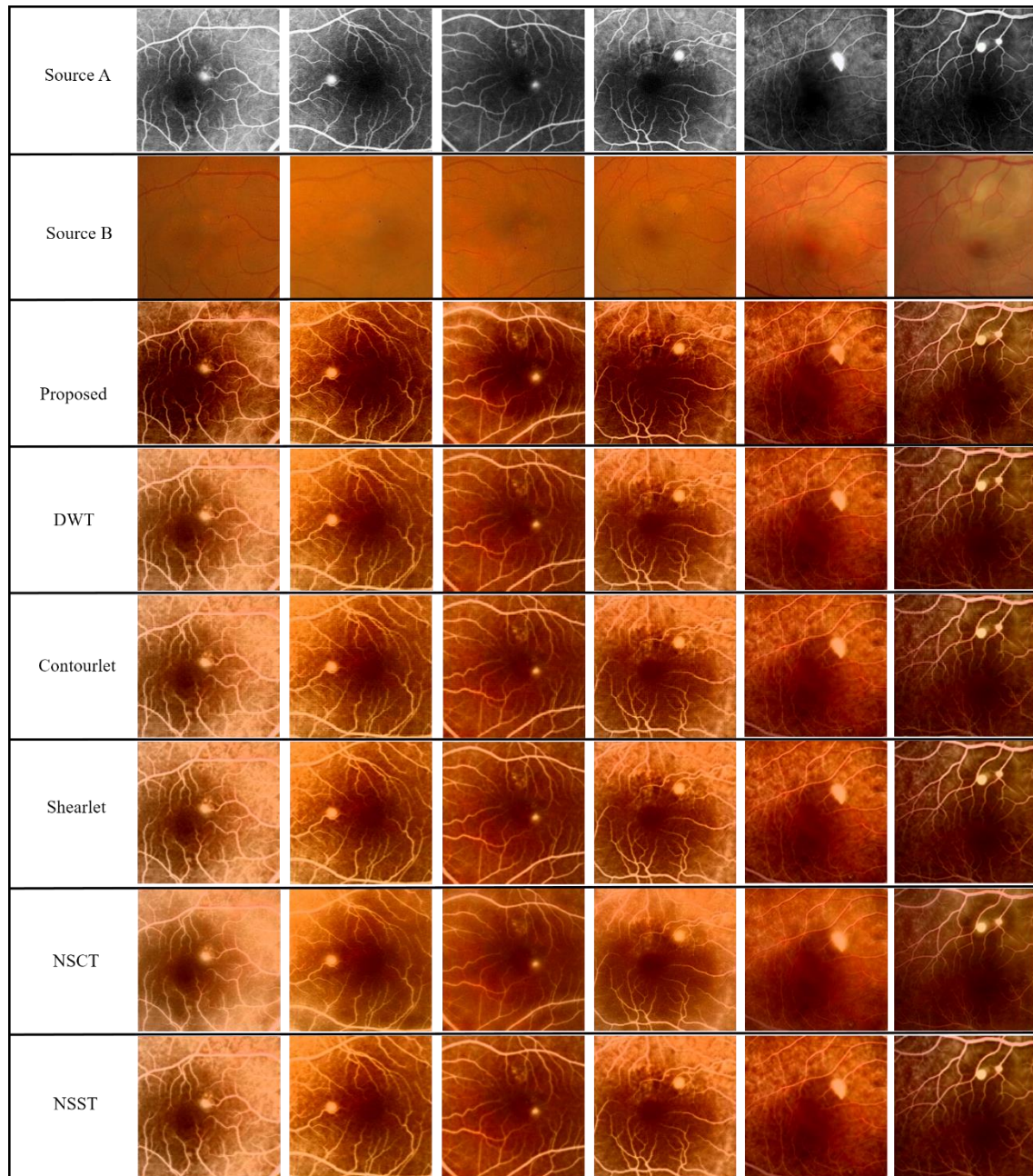

Figure 11. The fused CSCR images of the proposed and baseline methods.

5. Conclusions

In this paper, a multi-modal CSCR fundus image fusion method in NSCT domain is presented to assist ophthalmologists to identify the CSCR lesion information and normal fundus structure more

convenient so as to better guide the laser surgery. This fusion idea is an intriguing combination of the GAN-based and PCNN-based fusion schemes, which is not only the application extension of deep learning-based fusion technique in the low-frequency component fusion task, but also the reappearance of PCNN fusion rule in the information integration field of high-frequency components. On the one hand, in order to alleviate the tedious process of manually designing fusion rules, we introduce the GAN-based method into the fusion scene of low-frequency components and improve its learning ability of the data distribution by integrating the cycle-consistency loss module. On the other hand, the PCNN-based method is employed to promote to fuse the high-frequency components, aiming to enhance the integration of detail information of multi-modal CSCR fundus images. Quantitative and qualitative experiments demonstrate the effectiveness and competitiveness of the proposed method in the fusion scene of CSCR fundus images. Although the proposed fusion method achieves considerable performance and provides a new idea for the research of MST-based fusion methods, we will continue to make efforts from the following two aspects that include: 1) further optimizing the training strategy of the GAN fusion method to improve the training efficiency and 2) further improving the fusion rules of the low-frequency components and high-frequency components to enhance the quality of fused images.

Acknowledgments

Conflict of interest

The authors declare that we have no conflict of interests regarding this paper.

References

1. A. Daruich, A. Matet, A. Dirani, E. Bousquet, M. Zhao, N. Farman, et al., Central serous chorioretinopathy: Recent findings and new physiopathology hypothesis, *Prog. Retinal Eye Res.*, **48** (2015), 82–118.
2. J. Yu, C. Jiang, G. Xu, Study of subretinal exudation and consequent changes in acute central serous chorioretinopathy by optical coherence tomography, *Am. J. Ophthalmol.*, **158** (2014), 752–756.
3. P. Balasubramaniam, V. P. Ananthi, Image fusion using intuitionistic fuzzy sets, *Inf. Fusion*, **20** (2014), 21–30.
4. H. Yin, Tensor sparse representation for 3-D medical image fusion using weighted average rule, *IEEE Trans. Biomed. Eng.*, **65** (2018), 2622–2633.
5. J. Li, M. Song, Y. Peng, Infrared and visible image fusion based on robust principal component analysis and compressed sensing, *Infrared Phys. Technol.*, **89** (2018), 129–139.
6. Y. Leung, J. Liu, J. Zhang, An improved adaptive intensity-hue-saturation method for the fusion of remote sensing images, *IEEE Geosci. Remote Sens. Lett.*, **11** (2014), 985–989.
7. Y. Yang, S. Tong, S. Huang, P. Lin, Multi-focus image fusion based on NSCT and focused area detection, *IEEE Sens. J.*, **15** (2015), 2824–2838.

8. B. Yang, S. Li, Multi-focus image fusion and restoration with sparse representation, *IEEE Trans. Instrum. Meas.*, **59** (2010), 884–892.

9. H. Li, X. Wu, Multi-focus image fusion using dictionary learning and low-rank representation, in *International Conference on Image and Graphics*, Springer, **10666** (2017), 675–686.

10. A. B. Hamza, Y. He, H. Krim, A. Willsky, A multiscale approach to pixel-level image fusion, *Integr. Comput.-Aided Eng.*, **12** (2005), 135–146.

11. L. Wang, B. Li, L. Tian, Multi-modal medical image fusion using the inter-scale and intra-scale dependencies between image shift-invariant shearlet coefficients, *Inf. Fusion*, **19** (2014), 20–28.

12. M. N. Do, M. Vetterli, Contourlets: a directional multiresolution image representation, in *Proceedings. International Conference on Image Processing*, Rochester, NY, USA, 2002.

13. Z. Zhu, M. Zheng, G. Qi, D. Wang, Y. Xiang, A phase congruency and local Laplacian energy based multi-modality medical image fusion method in NSCT domain, *IEEE Access*, **7** (2019), 20811–20824.

14. G. Easley, D. Labate, W.Q. Lim, Sparse directional image representations using the discrete shearlet transform, *Appl. Comput. Harmonic Anal.*, **25** (2008), 25–46.

15. Y. Yang, Y. Que, S. Huang, P. Lin, Multimodal sensor medical image fusion based on type-2 fuzzy logic in NSCT domain, *IEEE Sens. J.*, **16** (2016), 3735–3745.

16. J. Xia, Y. Chen, A. Chen, Y. Chen, Medical image fusion based on sparse representation and PCNN in NSCT domain, *Comput. Math. Methods Med.*, **2018** (2018), 1–12.

17. P. Ganasala, V. Kumar, CT and MR image fusion scheme in non-subsampled contourlet transform domain, *J. Digit. Imag.*, **27** (2014), 407–418.

18. S. Das, M. K. Kundu, A neuro-fuzzy approach for medical image fusion, *IEEE Trans. Biomed. Eng.*, **60** (2013), 3347–3353.

19. A. L. D. Cunha, J. Zhou, M. N. Do, The non-subsampled contourlet transform: Theory, design, and applications, *IEEE Trans. Image Process.*, **15** (2006), 3089–3101.

20. R. Hou, D. Zhou, R. Nie, D. Liu, L. Xiong, Y. Guo, et al., VIF-Net: an unsupervised framework for infrared and visible image fusion, *IEEE Trans. Comput. Imaging*, **6** (2020), 640–651.

21. H. Li, X. Wu, J. Kittler, Infrared and visible image fusion using a deep learning framework, in *2018 24th international conference on pattern recognition (ICPR)*, Beijing, China, (2018), 2705–2710.

22. J. Ma, H. Xu, J. Jiang, X. Mei, X. Zhang, DDcGAN: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion, *IEEE Trans. Image Process.*, **29** (2020), 4980–4995.

23. W. Kong, Y. Lei, X. Ni, Fusion technique for grey-scale visible light and infrared images based on non-subsampled contourlet transform and intensity-hue-saturation transform, *IET Signal Process.*, **5** (2011), 75–80.

24. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, et al., Generative adversarial networks, *Adv. Neural Inf. Process. Syst.*, **3** (2014), 2672–2680.

25. S. Yu, S. Zhang, B. Wang, H. Dun, L. Xu, X. Huang, et al., Generative adversarial network based data augmentation to improve cervical cell classification model, *Math. Biosci. Eng.*, **18** (2021), 1740–1752.

26. J. Ma, W. Yu, P. Liang, C. Li, J. Jiang, FusionGAN: A generative adversarial network for infrared and visible image fusion, *Inf. Fusion*, **48** (2019), 11–26.

27. R. P. Broussard, S. K. Rogers, M. E. Oxley, G. L. Tarr, Physiologically motivated image fusion for object detection using a pulse coupled neural network, *IEEE Trans. Neural Networks*, **10** (1999), 554–563.

28. W. Tan, P. Xiang, J. Zhang, H. Zhou, H. Qin, Remote sensing image fusion via boundary measured dual-channel pcnn in multi-scale morphological gradient domain, *IEEE Access*, **8** (2020), 42540–42549.

29. W. Tan, W. Thitøn, P. Xiang, H. Zhou, Multi-modal brain image fusion based on multi-level edge-preserving filtering, *Biomed. Signal Process. Control*, **64** (2021), 102280.

30. D. Paquin, D. Levy, E. Schreibmann, L. Xing, Multiscale image registration, *Math. Biosci. Eng.*, **3** (2006), 389–418.

31. J. Wang, J. Chen, H. Xu, S. Zhang, X. Mei, J. Huang, et al., Gaussian field estimator with manifold regularization for retinal image registration, *Signal Process.*, **157** (2019), 225–235.

32. K. Li, L. Yu, S. Wang, P. A. Heng, Unsupervised retina image synthesis via disentangled representation learning, in *International Workshop on Simulation and Synthesis in Medical Imaging*, Springer, Cham, **11827** (2019), 32–41.

33. E. M. Izhikevich, Class 1 neural excitability, conventional synapses, weakly connected networks, and mathematical foundations of pulse-coupled models, *IEEE Trans. Neural Networks*, **10** (1999), 499–507.

34. S. Yang, M. Wang, Y. Lu, W. Qi, L. Jiao, Fusion of multi-parametric SAR images based on SW-non-subsampled contourlet and PCNN, *Signal Process.*, **89** (2009), 2596–2608.

35. J. V. Aardt, Assessment of image fusion procedures using entropy, image quality, and multispectral classification, *J. Appl. Remote Sens.*, **2** (2008), 023522.

36. G. Qu, D. Zhang, P. Yan, Information measure for performance of image fusion, *Electron. Lett.*, **38** (2002), 313–315.

37. C. S. Xydeas, V. S. Petrovic, Objective image fusion performance measure, *Electron. Lett.*, **36** (2000), 308–309.

38. B. K. S. Kumar, Multifocus and multispectral image fusion based on pixel significance using discrete cosine harmonic wavelet transform, *Signal, Image Video Process.*, **7** (2013), 1125–1143.

39. 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** (2004), 600–612.

40. H. Li, X. Wu, T. Durrani, Multi-focus noisy image fusion using low-rank representation, preprint, arXiv: 1804.09325.