



REStoring Clarity: Unpaired Retina Image Enhancement Using Scattering Transform

Ellen Jieun Oh¹, Yechan Hwang¹, Yubin Han¹, Taegeun Choi²,
Geunyoung Lee², and Won Hwa Kim¹(✉)

¹ Pohang University of Science and Technology (POSTECH), Pohang, South Korea
{jiejunoh,yechan99,yubin,wonhwa}@postech.ac.kr

² Mediwhale, Seoul, South Korea

Abstract. Retina images are non-invasive and highly effective in the diagnosis of various diseases such as cardiovascular and ophthalmological diseases. Accurate diagnosis depends on the quality of the retina images, however, obtaining high-quality images can be challenging due to various factors, such as noise, artifacts, and eye movement. Methods for enhancing retina images are therefore in high demand for clinical purposes, yet the problem remains challenging as there is a natural trade-off between preserving anatomical details (e.g., vessels) and increasing overall image quality other than the content in it. Moreover, training an enhancement model often requires paired images that map low-quality images to high-quality images, which may not be available in practice. In this regime, we propose a novel Retina image Enhancement framework using Scattering Transform (REST). REST uses unpaired retina image sets and does not require prior knowledge of the degraded factors. The generator in REST enhances retina images by utilizing the Anatomy Preserving Branch (APB) and the Tone Transferring Branch (TTB) with different roles. Our model successfully enhances low-quality retina images demonstrating commendable results on two independent datasets.

1 Introduction

Microvasculature and neural tissue in the retina can be directly and non-invasively visualized in vivo [19], and retina images are used for the diagnosis of various diseases including cardiovascular diseases [18], Alzheimer's disease [23], and ophthalmological diseases such as glaucoma and cataract [9]. Therefore, screening diseases using retina images has significant advantages over painful invasive methods such as blood tests or biopsies. However, the screening accuracy highly depends on the quality of the image [21] which can be compromised by various factors such as noise, low contrast, uneven illumination, blurriness, camera model, and experience level of the clinician taking the image [6, 16, 17, 22].

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43999-5_45.

These issues make it difficult to discriminate important indicators and biomarkers, leading to the failure of early detection of diseases and proper intervention. Therefore, image enhancement of low-quality retina images is highly necessary. However, retina image enhancement is a challenging task due to several reasons. Intuitively, it requires paired low and high-quality retina images to learn to map low-quality images to their high-quality ones and such data are difficult to acquire in practice [1]. Moreover, enhancing images according to signal-to-noise ratio might discard important anatomical structures, e.g., the shape of the optic disc, vessels, and disease-related features, that are critical for disease screening.

To deal with the issues above, a structure-preserving guided retina image filter (SGRIF) was proposed to enhance unpaired retina images with prior knowledge of clouding effect [4]. While dealing with unpaired data, it is biased with de-clouding the artifact caused by the cataract. Several parametric methods were recently proposed including [25] and [24] with simple CycleGAN [26] structure which require authentic unpaired data only. Although they can map the tone of the input image such as color, contrast, and illuminance, to those of high-quality images, their ability to preserve crucial anatomical structures that are essential for accurate image examination [6] remains limited. Later, ISE-CRET [5] and PCENet (Pyramid Constraint Enhancement Network) [13] were proposed, which utilize supervised learning with synthesized low-quality images and authentic high-quality image pairs. These models preserve detailed features effectively, however, generating degraded images with priors remains a labor-intensive process, and trained models are specific to the dataset in question.

In this context, we propose an unpaired Retina image Enhancement with Scattering Transform (REST) which preserves the anatomical structure (e.g., vessels, optic disc, and cup) and maps the tone (e.g., color and illuminance) effectively by utilizing an unpaired dataset. Our model contains a genuine generator that includes two branches: the Anatomy Preserving Branch (APB) and the Tone Transferring Branch (TTB). The APB incorporates scattering transform, which effectively captures anatomic structures and the TTB employs a multi-layer convolution to refine the tone of the image as that of high-quality images. Constructing a cyclic architecture [26] with the proposed generators and discriminators for consistency regularization, the REST successfully learns how to enhance the low-quality retina images with the following contributions: **1)** REST only uses authentic unpaired data and does not require any prior knowledge, **2)** Successful preservation of anatomic structures is achieved with scattering transform, **3)** REST is extensively evaluated qualitatively and quantitatively on two independent datasets. Experiments on UK Biobank and EyeQ [6] datasets demonstrate that the REST can adequately enhance retina images by restoring dark and uncertain regions without compromising anatomical structures.

2 Related Works

The SGRIF proposed in [4] is a non-parametric method that utilizes a filter with degradation equations describing clouding effects caused by cataracts on

the retina image. SGRIF pose and tackles the enhancement problem as a dehazing problem in computer vision. Using the degradation equation, global structure transfer filter, and global edge-preserving smoothing filter, SGRIF de-clouds the retina image. Although this method does not require image pairs, its performance depends on prior functions for degradation which causes a lack of generalizability. Parametric methods, especially those based on deep learning, perform better at generalization by identifying and improving low-quality factors in input images with the parameters. Simple CycleGAN-based models, such as [25] and [24], use only the unpaired images and do not require prior knowledge but face challenges in preserving detailed information. To address this limitation, recent approaches synthesize low-quality images paired with authentic high-quality images for supervised learning, as proposed in [5, 12, 13]. Although such methods preserve structural information, it is prone to being dataset-specific by the degrading model, and synthesizing low-quality images is time-consuming due to the difficulty in designing an accurate degradation model.

3 Methods

We propose a generative framework for unpaired retina image enhancement, which translates low-quality retina images X to high-quality images Y through two separate branches: APB and TTB. As illustrated in Fig. 1, an input image is simultaneously fed into both branches with encoder-decoder structures, and their outputs are combined with a kernel to obtain an enhanced image.

3.1 Anatomy Preserving Branch

In order to preserve fine anatomical details, such as the shape of an optic disc, an optic cup, and vessels, the model must accurately capture high-frequency components, i.e., edges. To extract these high-frequency components, we designed the APB, which employs a wavelet scattering transform over multiple encoding layers. Wavelet scattering transform extracts high-frequency factors and invariants with respect to translation and rotation [3] by comprising wavelet filters Ψ with the set of frequency and phase indices Λ and J number of scales, a low pass filter Φ , and a modulus operator [2, 11]. Given an image x as an input, the scattering transform consisting of one layer, results in 0-th and first-order scattering coefficients, S_J^0 and S_J^1 , as

$$S_J x = [S_J^0 x, S_J^1 x] = [x * \Phi, |x * \Psi_\lambda| * \Phi], \lambda \in \Lambda \quad (1)$$

The 0-th order scattering coefficient $S_J^0 x$ is computed by convolving $(*)$ x with Φ . The first-order coefficients set $S_J^1 x$ is obtained by convolving x with a set of wavelet filters Ψ_λ and modulus operation followed by a low pass filter Φ . Since the wavelet filter sets Ψ_λ have diverse frequencies and angles (i.e., Λ) within 2D image space as well as scales (i.e., J), the filters ensure the scattering transform to capture the structural information with respect to the frequency and direction [3].

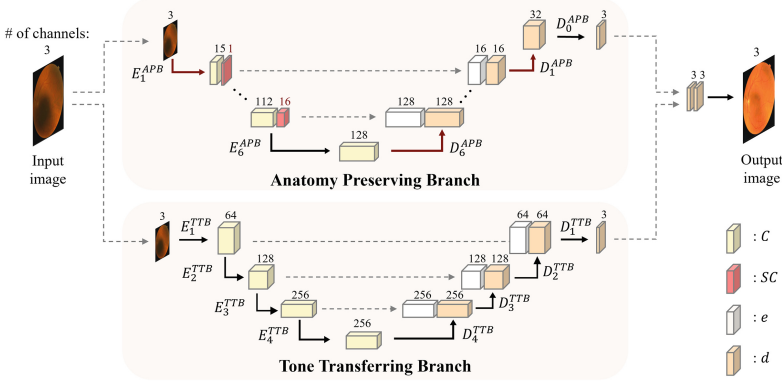


Fig. 1. The architecture of the generator in REST with APB (top) and TTB (bottom). Input images are fed into two branches and the two outputs are concatenated and convolved by 1×1 kernel to generate the output images. The symbols on the legend will be explained in Sect. 3.1 and Sect. 3.2.

The design of APB with scattering transform can be seen in Fig. 1 (top). First, the input image, denoted as x , is processed through a sequence of N encoding operations E_i^{APB} , where i denotes the layer index as

$$e_i^{\text{APB}} = \begin{cases} x, & i = 0 \\ E_i^{\text{APB}}(e_{i-1}^{\text{APB}}) = C_i \parallel SC_i, & 0 < i < N \\ E_i^{\text{APB}}(e_{i-1}^{\text{APB}}) = C_i, & i = N \end{cases} \quad (2)$$

where,

$$C_i = e_{i-1}^{\text{APB}} * k_{E_i}^{\text{APB}}, \quad SC_i = (S_1 e_{i-1}^{\text{APB}}) * k_{1 \times 1}. \quad (3)$$

The input of each layer e_{i-1}^{APB} undergoes both the scattering transform S_1 and convolution with a trainable kernel $k_{1 \times 1}$ and $k_{E_i}^{\text{APB}}$, except for the last layer in the encoding process. The C_i and SC_i are concatenated (\parallel) to yield an input for a subsequent layer. In the final layer, the input undergoes only convolution with a kernel $k_{E_i}^{\text{APB}}$.

After the completing N encoding processes, the decoding process, denoted as D_i^{APB} , commences. The final encoded feature map of the encoder, e_N^{APB} , undergoes $N + 1$ decoding processes. The decoding process is composed of a resizing $R(\cdot)$ and convolution with trainable kernel $k_{D_i}^{\text{APB}}$ as

$$d_i^{\text{APB}} = \begin{cases} D_i^{\text{APB}}(e_i^{\text{APB}}) = R(e_i^{\text{APB}}) * k_{D_i}^{\text{APB}}, & i = N \\ D_i^{\text{APB}}(d_{i+1}^{\text{APB}}, e_i^{\text{APB}}) = R(d_{i+1}^{\text{APB}} \parallel e_i^{\text{APB}}) * k_{D_i}^{\text{APB}}, & 0 < i < N \\ D_i^{\text{APB}}(d_{i+1}^{\text{APB}}) = d_{i+1}^{\text{APB}} * k_{D_i}^{\text{APB}}, & i = 0. \end{cases} \quad (4)$$

In the upscaling phase ($0 < i \leq N$), a combination of resizing and convolution techniques is implemented instead of transpose convolution. This approach prevents the occurrence of checkerboard artifacts, which can arise from uneven overlap during the transpose convolution [14]. The artifacts are particularly problematic in the APB since the APB deals with high-frequency features which are

easily affected them. The use of a combination of resizing and convolution techniques prevents the occurrence of these artifacts and ultimately produces clearer images. More specifically, we utilized interpolation with a factor of 2 and convolution with the $k_{D_i}^{\text{APB}}$ kernel. Additionally, to preserve anatomical structures such as vessels and an optic disc of the input image, we introduce skip connections [20] between the encoder and decoder. These connections involve the concatenation of the output of each encoder layer, e_i ($0 < i < N$), with the corresponding decoder layer's input, as illustrated in Fig. 1 (top).

3.2 Tone Transferring Branch

As the quality of a retina image depends on the tone of the image such as color, illumination, and contrast in addition to anatomical structures [6], it is crucial to generate synthetic images that have a similar tone to high-quality images. To ensure that the synthetic images resemble the tone of high-quality images, we designed TTB, which consists of multiple convolutional layers with a U-Net [20] architecture as in Fig. 1 (bottom). Similar to the APB, the input x is subject to a sequence of N encoding and decoding operations, denoted as E_i^{TTB} and D_i^{TTB} , respectively. For the encoding layers E_i^{TTB} , the input e_{i-1}^{TTB} is convolved with the trainable kernel k_i^{TTB} as

$$e_i^{\text{TTB}} = \begin{cases} x, & i = 0 \\ E_i^{\text{TTB}}(e_{i-1}^{\text{TTB}}) = e_{i-1}^{\text{TTB}} * k_i^{\text{TTB}}, & 0 < i \leq N. \end{cases} \quad (5)$$

After the encoding process, the N sequence of decoding operations D_i^{TTB} commences with the layer $i = N$. In the first decoding layer, the output of the encoder e_N^{TTB} is up-scaled using the transposed convolution operator (\otimes). For subsequent layers ($1 \leq i < N$), the output of the previous decoding layer d_{i+1}^{TTB} and the corresponding output of the encoding layer e_i^{TTB} are concatenated processed through the transposed convolution operator as

$$d_i^{\text{TTB}} = \begin{cases} D_i^{\text{TTB}}(e_i^{\text{TTB}}) = e_i^{\text{TTB}} \otimes k_i^{\text{TTB}}, & i = N \\ D_i^{\text{TTB}}(d_{i+1}^{\text{TTB}}, e_i^{\text{TTB}}) = (d_{i+1}^{\text{TTB}} \parallel e_i^{\text{TTB}}) \otimes k_i^{\text{TTB}}, & 1 \leq i < N. \end{cases} \quad (6)$$

Along with our objective function containing GAN Loss [7], which will be illustrated in the next section, Sect. 3.3, the sequence of convolution and transposed convolution layers encourage the distribution of the generated image, including the tone of the image, to match that of the high-quality images [26]. Moreover, skip connections that provide direct information from the encoder to the decoder constrain the branch from modifying the input image beyond recognition [20].

3.3 Loss Function for Unpaired Image Enhancement

The popular adversarial training architecture for unpaired image translation [26] is devised to train an enhancement model G comprised of APB and TTB, from Sect. 3.1 and Sect. 3.2, for synthesizing high-quality images Y from low-quality

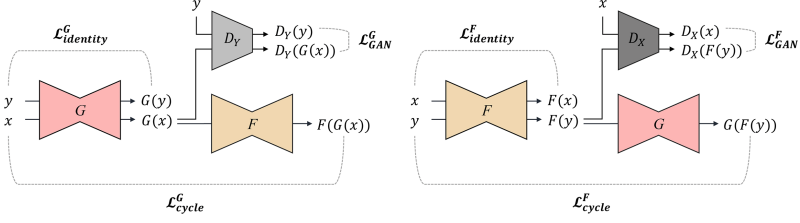


Fig. 2. The overall architecture of the REST is composed of cycles of two generators: G, F , and two discriminators: D_X, D_Y .

images X . As seen in Fig. 2, the overall architecture is composed of two generators $G: X \rightarrow Y$ and $F: Y \rightarrow X$, and two discriminators D_X and D_Y which discriminate authentic images and synthetic images of low-quality and high-quality respectively. The two generators share the same architecture as Fig. 1 and the two discriminators are also identically modeled with multiple convolution layers. To successfully train the enhancement model $G: X \rightarrow Y$, we composed our loss function \mathcal{L} with GAN losses $\mathcal{L}_{GAN}^G, \mathcal{L}_{GAN}^F$, cycle consistency losses $\mathcal{L}_{cycle}^G, \mathcal{L}_{cycle}^F$ and identity losses $\mathcal{L}_{identity}^G, \mathcal{L}_{identity}^F$ [26] as follows:

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{GAN}^G + \mathcal{L}_{GAN}^F + \mathcal{L}_{cycle} + \mathcal{L}_{identity}. \quad (7)$$

To ensure that generated images look like genuine images, the GAN losses $\mathcal{L}_{GAN}^G, \mathcal{L}_{GAN}^F$ are defined as [7]

$$\begin{aligned} \mathcal{L}_{GAN}^G &= \mathbb{E}_{y \in Y} [\|D_Y(y)\|_2^2] + \mathbb{E}_{x \in X} [\|1 - D_Y(G(x))\|_2^2], \\ \mathcal{L}_{GAN}^F &= \mathbb{E}_{x \in X} [\|D_X(x)\|_2^2] + \mathbb{E}_{y \in Y} [\|1 - D_X(F(y))\|_2^2]. \end{aligned} \quad (8)$$

To preserve the contents of the input image, we aim to make $F(G(x)) \approx x$ and $G(F(y)) \approx y$ with the cycle consistency losses as [26]

$$\mathcal{L}_{cycle}^G = \lambda_{c1} \mathbb{E}_{x \in X} [\|F(G(x)) - x\|_1], \quad \mathcal{L}_{cycle}^F = \lambda_{c2} \mathbb{E}_{y \in Y} [\|G(F(y)) - y\|_1] \quad (9)$$

where λ_{c1} and λ_{c2} are hyperparameters. To make sure that the generators avoid translation when there is no need, i.e., a high-quality image as an input to G should lead to a high-quality image without changes, identity losses are defined as [26]

$$\mathcal{L}_{identity}^G = \lambda_{i1} \mathbb{E}_{y \in Y} [\|G(y) - y\|_1], \quad \mathcal{L}_{identity}^F = \lambda_{i2} \mathbb{E}_{x \in X} [\|F(x) - x\|_1] \quad (10)$$

with the hyperparameters λ_{i1} and λ_{i2} . With the aforementioned losses, we aim to obtain generators G^*, F^* that minimize the loss function while discriminators D_X^*, D_Y^* that maximize it as

$$G^*, F^*, D_X^*, D_Y^* = \arg \max_{D_X, D_Y} \min_{G, F} \mathcal{L}(G, F, D_X, D_Y). \quad (11)$$

After training, G^* is utilized at the inference stage to enhance retina images.

4 Experiments

4.1 Datasets and Experimental Setup

Datasets. From **UKB**, 2000 images, i.e., 1000 high-quality and 1000 low-quality images respectively, were randomly sampled and split into train and test sets by a ratio of 2:1. For the **EyeQ**, we utilized 23252 labeled images, i.e., 6434 low-quality images labeled ‘usable’ and 16818 high-quality images labeled ‘good’ [6] and followed the data splitting protocol into the train, validation, and test sets as in [5]. As ISECRET and PCENet require low-quality pairs and a mask for every image during training, degraded pairs were generated between high-quality ones and masks with the given degrading and masking methods in [5, 13].

Table 1. FIQA, WFQA, the number (#) of parameters, usage of paired data, and training samples (# of data) for UKB (left) and EyeQ (right) are compared.

	Parameters	Pair	Samples	FIQA	WFQA		Pair	FIQA
Real high	-	-	-	0.989	1.978	Real high	-	0.980
Real low	-	-	-	0.051	0.141	Real low	-	0.198
CycleGAN (ResNet) [7, 8]	28.29 M	-	1332	0.926 ± 0.04	1.853 ± 0.07	cGAN [10]	-	0.304
CycleGAN (U-Net) [7, 20]	114.35 M	-	1332	0.893 ± 0.04	1.765 ± 0.09	CycleGAN [7]	-	0.890
ISECRET [5]	15.36 M	✓	1998	0.898 ± 0.02	1.765 ± 0.09	CutGAN [15]	-	0.919
PCENet [13]	26.65 M	✓	1332	0.389 ± 0.06	0.782 ± 0.12	ISECRET [5]	✓	0.937
REST (ours)	16.35 M	-	1332	0.940 ± 0.01	1.881 ± 0.03	PCENet [13]	✓	0.894
REST w/o APB (ours)	13.70 M	-	1332	0.920 ± 0.02	1.843 ± 0.54	REST (ours)	-	<u>0.936</u>
REST w/o TTB (ours)	8.18 M	-	1332	0.223 ± 0.05	0.454 ± 0.09			

Evaluation Setup. To quantitatively evaluate the quality of the generated retina images, Fundus Image Quality Assessment (FIQA) [6] and Weighted FIQA (WFQA) [13] scores were adopted. For FIQA, a Multiple Color-space Fusion Network (MCF-Net) [6] which labels each input image of the retina as ‘Good’, ‘Usable’, or ‘Reject’ is utilized. With the output of the MCF-Net, the FIQA score is calculated. The FIQA is calculated as the ratio of the number of images labeled as ‘Good’ to the total number of inputs, while the WFQA is determined by the ratio of the weighted sum of the number of images labeled as ‘Good’ and ‘Usable’ to the total number of inputs, with weights of 2 and 1 assigned to ‘Good’ and ‘Usable’ respectively. For evaluation, the FIQA and WFQA scores were calculated by inputting output images from enhancement models to pretrained MCF-Net. For **UKB** data, both quantitative and qualitative evaluations were done for our model (REST) and four baseline models, i.e., two traditional unpaired image translation models, CycleGANs with ResNet and U-Net [8, 20, 26] and two latest models designed for unpaired retina image enhancement, ISECRET [5] and PCENet [13]. For **EyeQ** data, experiments were done for REST and PCENet following the settings in [5], and results are compared with traditional paired image translation model cGAN [15] and unpaired image translation models, CycleGAN [26], CutGAN [15], and ISECRET [5], reported in [5].

Implmenementation. The batch size was set to 4 and the Adam optimizer was adopted with the initial learning rate of 0.0002 linearly decaying. For scattering transform, a Morlet wavelet was used with eight different angles within 2D image space and J was set to 1. A 2D Gabor filter was used as a low-pass filter. Parameters in losses, i.e., λ_{c1} , λ_{c2} , λ_{i1} and λ_{i2} , were set to 10, 10, 5 and 5 respectively.

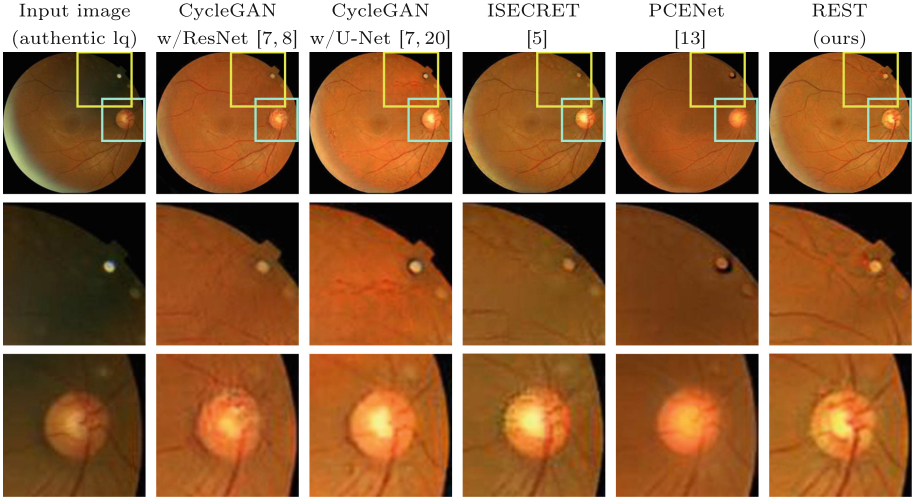


Fig. 3. Examples of the low-quality input image and the outputs of different models. The zoomed-in figures from yellow and blue boxes (2nd and 3rd rows) show that our model REST better preserves the anatomical structures (vessels, optic disc, and cup) and transfers the tone. (Color figure online)

4.2 Results

Regarding the experiments on **UKB**, the result of the qualitative evaluation is illustrated in Table 1 (left). In the Table 1, (1) the number of parameters, (2) the requirement for the paired data while training, (3) the total number of images used for training, and (4) the FIQA score and WFQA score for each model are illustrated with the best score denoted in bold, and the second-best in underlined. REST achieved the best FIQA with the smallest standard deviation. Compared with the second-best model, excluding the ablation study, cycleGAN with Resnet backbone, the FIQA score of REST is 0.014 (1.4%p) higher with a lot smaller number of total parameters. Qualitative evaluation on a UKB sample is illustrated in Fig. 3. In Fig. 3, the low-quality input image and synthesized high-quality images by each model are shown sequentially. The second and third rows in Fig. 3 show the details of the image marked with a box in the first row. As can be seen in Fig. 3, REST preserved the anatomical structures such as

vessels both in and out of the optic disc and the shape of the optic disc and cup better than the baselines. The overall tone of the image was also successfully transferred to make the contrast and the illuminance become even.

For the experiment on the **EyeQ** dataset, the result of the qualitative evaluation is illustrated in Table 1 (right). REST achieved the second-best FIQA score slightly behind by only 0.001p from ISECRET. However, notice that ISECRET requires synthetic paired images while REST uses only unpaired images. As the low-quality images utilized in the EyeQ experiment, i.e., labeled as ‘Usable’, are already deemed to be of relatively high quality, the capability of REST that shows superior preservation of anatomical structure in contaminated images was not sufficiently manifested as in the UKB experiment. Still, the results are highly competitive and demonstrate potentials to be used for real applications.

5 Conclusion

In the study, we proposed a novel approach to unpaired retina image enhancement, i.e., REST. While using only the authentic unpaired images, the proposed method effectively preserves anatomical structures during the enhancement process. This is done by utilizing APB for preserving the details by the scattering transform and TTB for transferring the tone of the images. Notably, REST has demonstrated commendable performance on both two different datasets.

Acknowledgement. This research was supported by NRF-2022R1A2C2092336 (50%), IITP-2022-0-00290 (30%), IITP-2019-0-01906 (AI Graduate Program at POSTECH, 10%), and IITP-2022-2020-0-01461 (ITRC, 10%) funded by MSIT in South Korea.

References

1. Ancuti, C.O., Ancuti, C., Vasluianu, F.A., Timofte, R.: Ntire 2021 nonhomogeneous dehazing challenge report. In: CVPR, pp. 627–646 (2021)
2. Andreux, M., Angles, T., Exarchakisgeo, G., et al.: Kymatio: scattering transforms in python. *J. Mach. Learn. Res.* **21**(1), 2256–2261 (2020)
3. Bruna, J., Mallat, S.: Invariant scattering convolution networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(8), 1872–1886 (2013)
4. Cheng, J., Li, Z., Gu, Z., Fu, H., Wong, D.W.K., Liu, J.: Structure-preserving guided retinal image filtering and its application for optic disk analysis. *IEEE Trans. Med. Imaging* **37**(11), 2536–2546 (2018)
5. Cheng, P., Lin, L., Huang, Y., Lyu, J., Tang, X.: I-SECRET: importance-guided fundus image enhancement via semi-supervised contrastive constraining. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12908, pp. 87–96. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87237-3_9

6. Fu, H., Wang, B., Shen, J., Cui, S., Xu, Y., Liu, J., Shao, L.: Evaluation of retinal image quality assessment networks in different color-spaces. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 48–56. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_6
7. Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al.: Generative adversarial networks. *Commun. ACM* **63**(11), 139–144 (2020)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR, pp. 770–778 (2016)
9. Huang, Y., et al.: Automated hemorrhage detection from coarsely annotated fundus images in diabetic retinopathy. In: ISBI, pp. 1369–1372. IEEE (2020)
10. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: CVPR, pp. 1125–1134 (2017)
11. Kim, W.H., Ravi, S.N., Johnson, S.C., et al.: On statistical analysis of neuroimages with imperfect registration. In: ICCV, pp. 666–674 (2015)
12. Li, H., Liu, H., Fu, H., et al.: Structure-consistent restoration network for cataract fundus image enhancement. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13432, pp. 487–496. Springer, Heidelberg (2022). https://doi.org/10.1007/978-3-031-16434-7_47
13. Liu, H., Li, H., Fu, H., et al.: Degradation-invariant enhancement of fundus images via pyramid constraint network. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13432, pp. 507–516. Springer, Heidelberg (2022). https://doi.org/10.1007/978-3-031-16434-7_49
14. Odena, A., Dumoulin, V., Olah, C.: Deconvolution and checkerboard artifacts. *Distill* **1**(10), e3 (2016)
15. Park, T., Efros, A.A., Zhang, R., Zhu, J.-Y.: Contrastive learning for unpaired image-to-image translation. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12354, pp. 319–345. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58545-7_19
16. Pérez, A.D., Perdomo, O., Rios, H., Rodríguez, F., González, F.A.: A conditional generative adversarial network-based method for eye fundus image quality enhancement. In: Fu, H., Garvin, M.K., MacGillivray, T., Xu, Y., Zheng, Y. (eds.) OMIA 2020. LNCS, vol. 12069, pp. 185–194. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-63419-3_19
17. Raj, A., Tiwari, A.K., Martini, M.G.: Fundus image quality assessment: survey, challenges, and future scope. *IET Image Process.* **13**(8), 1211–1224 (2019)
18. Rim, T.H., Lee, C.J., Tham, Y.C., et al.: Deep-learning-based cardiovascular risk stratification using coronary artery calcium scores predicted from retinal photographs. *Lancet Dig. Health* **3**(5), e306–e316 (2021)
19. Rim, T.H., Lee, G., Kim, Y., Tham, Y.C., et al.: Prediction of systemic biomarkers from retinal photographs: development and validation of deep-learning algorithms. *Lancet Dig. Health* **2**(10), e526–e536 (2020)
20. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
21. Şevik, U., Köse, C., Berber, T., Erdöl, H.: Identification of suitable fundus images using automated quality assessment methods. *J. Biomed. Opt.* **19**(4), 046006–046006 (2014)
22. Shen, Z., Fu, H., Shen, J., Shao, L.: Modeling and enhancing low-quality retinal fundus images. *IEEE Trans. Med. Imaging* **40**(3), 996–1006 (2020)

23. Snyder, P.J., Alber, J., Alt, C., et al.: Retinal imaging in Alzheimer's and neurodegenerative diseases. *Alzheimer's Dementia* **17**(1), 103–111 (2021)
24. You, Q., Wan, C., Sun, J., Shen, J., Ye, H., Yu, Q.: Fundus image enhancement method based on cyclegan. In: EMBC, pp. 4500–4503. IEEE (2019)
25. Zhao, H., Yang, B., Cao, L., Li, H.: Data-driven enhancement of blurry retinal images via generative adversarial networks. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 75–83. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_9
26. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: ICCV, pp. 2223–2232 (2017)