Data-Driven Enhancement of Blurry Retinal Images via Generative Adversarial Networks

He Zhao, Bingyu Yang, Lvchen Cao, and Huiqi Li(✉)

Beijing Institute of Technology, Beijing, China
huiqili@bit.edu.cn

Abstract. In this paper, we aim at improving the quality of blurry retinal images that are caused by ocular diseases. The blurry images could affect clinical diagnosis for both ophthalmologists and automatic aided system. Inspired by the great success of generative adversarial networks, a data-driven approach is proposed to enhance the blurry images in a weakly supervised manner. That is to say, instead of paired blurry and high-quality images, our approach can be trained with two sets of unpaired images. The advantage of unpaired training setting makes our approach easily applicable, since the annotated data are very limited in medical images. Compared with traditional methods, our model is an end-to-end approach without human designed adjustments or prior knowledge. However, it achieves a superior performance on blurry images. Besides, a dynamic retinal image feature constraint is proposed to guide the generator to improve the performance and avoid over-enhancing the extremely blurry region. Our approach can work on large image resolution which makes it widely beneficial to clinic images.

1 Introduction

Retinal imaging is widely used by ophthalmologists for early disease detection and diagnosis including glaucoma, diabetic retinopathy, hypertensive retinopathy. However, the unsatisfied quality of retinal images such as poor illuminance, low contrast and blurriness makes it hard to distinguish different diseases and also decrease the accuracy of diagnosis for doctors [9]. Meanwhile, the poor quality image leads to an unsatisfied result for automatic image processing (*e.g.* segmentation, tracking) which may further influence the analysis of diseases. Recent years, there are many researchers trying to enhance the retinal images with low quality. Most of the methods focus on improving illuminance and contrast of retinal image using normalization techniques, while little effort is putting on deblurriness.

In this paper, we propose a novel deep learning approach to enhance the blurry images. Different from other retinal image enhancement methods with sophisticated adjustments of parameters, our approach is very straightforward with an end-to-end framework. We propose a solution of using an image-to-image

D. Shen et al. (Eds.): MICCAI 2019, LNCS 11764, pp. 75–83, 2019.
https://doi.org/10.1007/978-3-030-32239-7_9

translation pipeline with a two-way GAN that is restricted by feature constraint. In addition, our approach only requires two sets of unpaired blurry/high-quality images as inputs, which is applicable in many cases where paired medical images are rare and not accessible. The contribution of our approach can be summarized as follows. (1) To the best of our knowledge, our approach is the first end-to-end deep generative model to enhance the blurry retinal images, which is not based on prior knowledge and designed adjustment. (2) Our model can learn the mapping by a weakly supervision manner (*i.e.* no paired blurry/high quality images are required), which is appropriate to the situation that paired data are limited in medical images. (3) The proposed dynamic feature descriptor provides a feature constraint that helps the model to produce more reliable enhancement containing the core information and fewer artifacts. (4) Besides, our enhanced images are helpful on improving the performance of automatic processing, such as vessel segmentation and tracking.

2 Related Work

Retinal Image Enhancement. Image enhancement has been well studied in recent years, and various methods are proposed [3,8,10]. They have achieved a good performance on luminance and contrast enhancement. For medical images, image enhancement has been explored and methods are focusing on specific tasks. The histogram based method, contrast limited adaptive histogram equalization (CLAHE), is widely used and applied to improve the poor quality retinal images. A luminosity and contrast adjustment method is proposed in [13]. The luminance is enhanced by a luminance gain matrix from gamma correction and the contrast is enhanced by CLAHE in the Lab color space. This kind of methods is based on the knowledge of the neighborhood region. In [6], Fourier transformation is utilized to remove the opacity and CLAHE is used to enhance the contrast based on HIS color space. The blurriness of retinal images are modelled using scattering process in [11] and the images are enhanced with estimation of transmission map and background illuminance. Predefined parameters and prior knowledge such as enhancing details or region selected are required in the above models, and this may make their model sensitive.

Generative Adversarial Model. The generative adversarial model that is designed as a two-player zero-sum game between a discriminator and a generator [2], has been developed rapidly in the past five years. The generator is trained to generate samples as similar as the real ones, while the discriminator is designed to distinguish the generated samples. The model is utilized to generate realistic images of both natural images [4,14] and medical images [7,12]. For example, Isola *et al.* [4] present a general translator to transform image from one domain to another, which requires the paired training data. Later, Zhu *et al.* [14] loose the constraint by introducing a backward mapping from output to input. Application based on Zhu *et al.* is also studied to enhance the photos in terms of color and sharpness with modified model structure and training scheme for stability [1]. In medical imaging, the adversarial learning is widely used in other

tasks such as registration, reconstruction and segmentation and have achieved satisfied results, in which it is applied as an additional constraint.

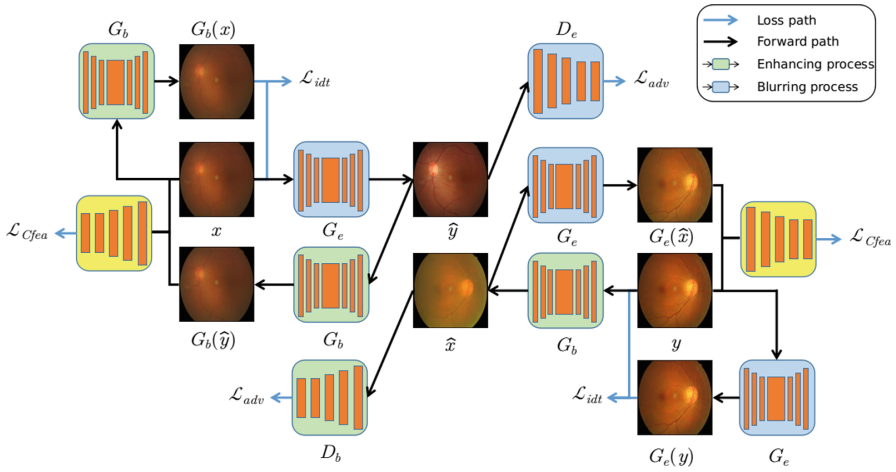


Fig. 1. Flowchart of our approach focusing on the low-quality to high-quality process. Blue boxes indicate the generator and discriminator for enhancing process, while green boxes are for blurring process. G_e and G_b represent enhanced generator and blurred generator respectively, while D_e and D_b indicate the corresponding discriminators. The yellow box refers to the dynamic feature descriptor for feature consistent constraint. x indicates the blurry image, while y refers to the high quality image. (Color figure online)

3 Method

In this section, we propose a retinal image enhancement model based on the generative adversarial networks with feature consistent constraint. Denote the training dataset as $\{\boldsymbol{x_i}, \boldsymbol{y_i}\}_{i=1}^{N}$, where $\boldsymbol{x} \in \mathbb{R}^{W \times H \times 3}$ refers to the blurry images and $\boldsymbol{y} \in \mathbb{R}^{W \times H \times 3}$ is the high quality images. Our goal is to enhance the blurry image $\boldsymbol{x}$ and produce the enhanced image $\hat{\boldsymbol{y}}$ to make it as clear as the normal retinal image $\boldsymbol{y}$. In order to achieve this target, two generative models are employed $G_e : \boldsymbol{x} \rightarrow \boldsymbol{y}$ and $G_b : \boldsymbol{y} \rightarrow \boldsymbol{x}$. The first generator G_e is used to enhance the image from low-quality to high-quality and G_b is used for providing a training reference by converting the high-quality to low-quality. This mechanism forms a feedback of information and makes the model trainable using a weakly supervised manner, where the feature descriptor is proposed for a perceptual constraint.

3.1 Model Structure

The flowchart of our approach is displayed in Fig. 1. The two generators G_e and G_b share the same model structure but with different tasks, while the two

discriminators D_e and D_b also have the same model structure. The D_e aims to distinguish the enhanced images $\hat{\boldsymbol{y}}$ from the real high quality images $\boldsymbol{y}$, while the D_b is utilized for checking out whether the input images are blurry images $\boldsymbol{x}$ or the synthesized $\hat{\boldsymbol{y}}$ by G_b. Besides the generators and discriminators, we also introduce a dynamic feature descriptor to extract the feature description of each image. In practice, the discriminators D_e and D_b are selected as dynamic feature descriptors for the enhanced images and the blurred images respectively. The parameters of these two feature descriptors are updated for each training iteration. In this way, the features designed for describing the real image characteristics become stronger and stronger during training and it's less computation expensive for this setting.

Our generator consists of 3 convolutional and deconvolutional layers and a residual bottleneck module. The input image is passed into the first convolutional layer with kernel size of $7 \times 7 \times 64$ and stride 1, which is followed by two downsampling convolutional layers whose kernel sizes are $3 \times 3 \times 128$ and $3 \times 3 \times 256$ respectively. The deconvolutional layers have the same kernel sizes as the corresponding convolutional layers. The bottleneck module contains 9 residual blocks, in which two convolutional layers and one skip connection are employed. Our discriminator consists of 5 convolutional layers, where the convolutional kernels with 3×3 and stride 2 are applied for 4 times downsampling. The Patch-GAN [4] is utilized here which classifies whether the image patches are real or fake instead of the whole image. This setting decreases the number of parameters and improves the ability of the discriminators.

3.2 Objective Function

The entire enhancement process can be described as $G_e(x) : \boldsymbol{x} \rightarrow \hat{\boldsymbol{y}}$, while a discriminator function is defined as $D_e : \boldsymbol{X} \rightarrow d \in [0, 1]$. When the input $\boldsymbol{X}$ is the real high quality image $\boldsymbol{y}$, d should be closed to 1 and when $\boldsymbol{X}$ is the enhanced image $\hat{\boldsymbol{y}}$, d should be closed to 0. Different from the traditional GAN, another pair of generator G_b and discriminator D_b is employed here to train the model using a weakly supervised manner. We follow the GAN's idea and combine the two pairs of G and D, and the optimization problem which needs to be solved is:

$$\min_{G_e, G_b} \max_{D_e, D_b} \mathcal{L} = \omega_{adv}\mathcal{L}_{adv} + \omega_{Cfea}\mathcal{L}_{Cfea} + +\omega_{idt}\mathcal{L}_{idt} \tag{1}$$

where the adversarial loss is $\mathcal{L}_{adv} = (D_e(\boldsymbol{y})-1)^2 + D_e(\hat{\boldsymbol{y}})^2 + (D_b(\boldsymbol{x})-1)^2 + D_b(\hat{\boldsymbol{x}})^2$, with $\mathcal{L}_{Cfea}$ and $\mathcal{L}_{idt}$ being the feature consistent constraint and identity loss respectively. Here, we use a least-squares loss [5] to obtain a more stable training process and better results. These three terms form our final loss function and will be introduced in the following.

Feature Consistent Constraint. In general, the generator G_e can produce many "enhanced" outputs if the adversarial loss is the only restriction. In order to make G_e enhance the blurry images in the direction we expect, the consistent mapping

function is introduced. The intuition of this restriction is that one blurry image $\hat{\boldsymbol{x}}$ generated by G_b from $\boldsymbol{y}$ should be converted back by the enhanced function G_b. Similarly, for each enhanced image $\hat{\boldsymbol{y}}$ generated by G_e should also satisfy the cycle consistency, where $G_b(\hat{\boldsymbol{y}}) \rightarrow \boldsymbol{x}$. So the consistent constraint can be defined as:

$$\mathcal{L}_C = \|G_b(\hat{\boldsymbol{y}}) - \boldsymbol{x}\|_1 + \|G_e(\hat{\boldsymbol{x}}) - \boldsymbol{y}\|_1 \tag{2}$$

where $\hat{\boldsymbol{y}} = G_e(\boldsymbol{x})$ and $\hat{\boldsymbol{x}} = G_b(\boldsymbol{y})$ indicate the outputs of enhanced and blurry result respectively. This consistent strategy creates a supervisory signal to train the enhanced generator G_e and the blurred generator G_b, which somehow makes this unpaired training task into a "pair-wise" learning. However, due to the complex background and foreground of retinal images, image-level consistent constraint is not enough and leads to unsatisfactory artifacts on the enhanced images. A feature consistent constraint is further proposed to overcome this issue. In practice, we choose to measure the difference of feature maps coming from convolutional layers. For specific layer l and the convolutional function F, the final feature consistent constraint is defined as:

$$\mathcal{L}_{Cfea} = \left\|F^l(G_b(\hat{\boldsymbol{y}})) - F^l(\boldsymbol{x})\right\|_1 + \left\|F^l(G_e(\hat{\boldsymbol{x}})) - F^l(\boldsymbol{y})\right\|_1 \tag{3}$$

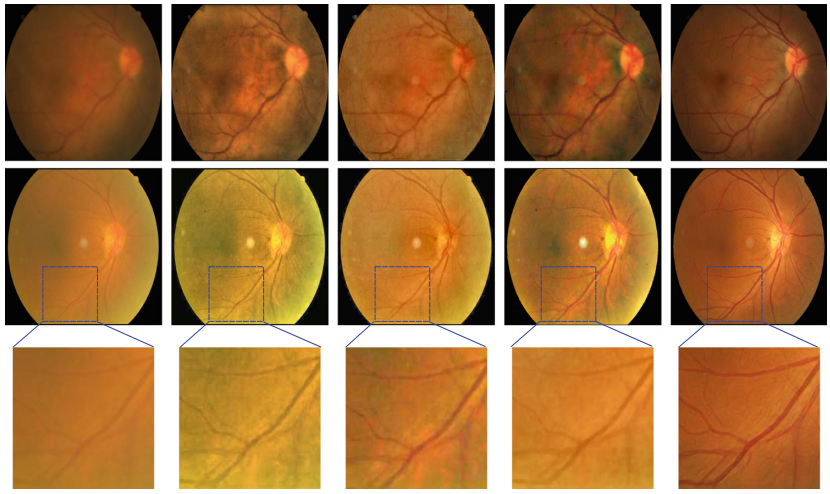
Identity Loss. The identity loss is applied to regularize the generator G_e to produce an identity mapping when a high-quality image $\boldsymbol{y}$ is fed into G_e. The same operation is also applied to the output of another generator G_b. Then the final identity loss is defined as:

$$\mathcal{L}_{idt} = (\|G_e(\boldsymbol{y}) - \boldsymbol{y}\|_1 + \|G_b(\boldsymbol{x}) - \boldsymbol{x}\|_1) \tag{4}$$

Intuitively, the same image should be obtained when a high-quality image is fed into the enhanced generator. This additional loss is very helpful to preserve the color between the input and output by adding a restriction to the generator. Without the identity loss, the model is free to generate and the enhanced image $\hat{\boldsymbol{y}}$ is corrupted with unsatisfied color appearance.

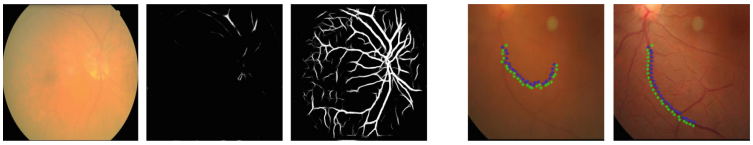
4 Experiment and Results

In practice, we set $\omega_{adv} = 1$, $\omega_{Cfea} = 10$, $\omega_{idt} = 10$ as the weights of each single loss. Empirically, all convolutional layers in discriminator are selected to computing $\mathcal{L}_{Cfea}$. The dataset used for training and testing are from hospitals. There are 550 blurry images and 550 high quality images used for training and 60 blurry images are used for testing. The blurry images are from cataract patients and the high quality are from normal people. There are two subsets of testing set. The first one contains only fifty blurry images, while the second one consists of ten images with ground-truths. Images of cataract patients are blurry due to the opacity of lens, and the images after surgery are regarded as ground-truth. To evaluate our retinal image enhancement performance, we carry out the following experiments including visual and quantitative evaluation.



(a) Blurry image (b) Mitra *et al.*[5] (c) Zhou *et al.*[14] (d) Xiong *et al.*[12] (e) Ours

Fig. 2. Visual comparison between our approach and other methods. Two different image samples are shown in the first and the second rows, while the last row is the zoom-in views of selected regions.



(a) Blurry image (b) Segmentation on blury image (c) Segmenetation on enhanced image by ours (d) Vessel tracking on blury image (e) Vessel tracking on enhanced image by ours

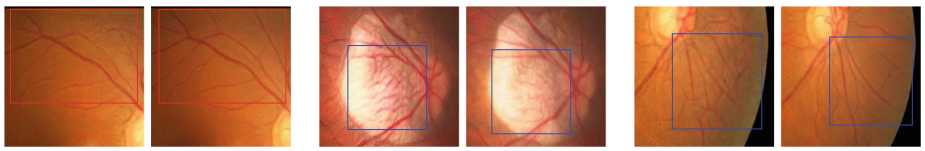
Fig. 3. Results of two retinal image processing tasks on blurry and enhanced images. (a–c) show the segmentation task, while (d–e) display the tracking task.

4.1 Enhancement Evaluation – Visual and Quantitative

In this section, two kinds of image quality assessment are adopted, full-reference and no-reference evaluation. For no-reference assessment, Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE), Natural Image Quality Evaluator (NIQE) and Entropy are chosen to assess each enhanced image and its original blurry retinal image. Fifty blurry images are employed in the no-reference evaluation. All these no-reference quality metrics give us an absolute image quality but not the proximity to a reference. As to full-reference assessment, Signal-to-Noise Ratio (PSNR) and structural similarity index measure (SSIM) are utilized. SSIM and PSNR give the comparison between the enhanced image and the ground truth. For this assessment, images of ten cataract patients before and after cataract surgery are used for evaluation. The quantitative results are displayed in Table 1. The lower values of BRISQUE and NIQE indicate better

Table 1. Quantitative comparison between our approach and other methods, where full-reference and no-reference assessments are included. The lower the score of BRISQUE and NIQE, the better, while opposite for others.

	No-reference			Full-reference	
	BRISQUE	NIQE	Entropy	PSNR	SSIM
Mitra *et al.* [6]	45.16	3.32	6.69	16.38	0.78
Zhou *et al.* [13]	46.13	4.30	6.74	17.73	0.73
Xiong *et al.* [11]	43.61	3.87	6.67	17.26	0.87
Ours w/o feature constraint	41.39	2.78	6.67	19.03	0.88
Ours	**40.62**	**2.74**	**6.89**	**19.24**	**0.89**



(a) (b) (c)

Fig. 4. Visual comparison between feature-level constraint and image-level constraint on enhanced retinal images. For each panel, images on the left are outputs from model with image-level constraint and images on the right are the outputs from model with feature constraint. (Color figure online)

image quality, while higher entropy scores refer to better quality. Our approach achieves the best BRISQUE, NIQE and entropy scores which are 40.62, 2.74 and 6.89 respectively. As to the full-reference assessment, our approach still obtains the highest value of PSNR with 19.24 and SSIM with 0.89. This superior performance can also be supported by the visual comparison of enhanced results shown in Fig. 2, in which three methods (Mitra *et al.* [6], Zhou *et al.* [13] and Xiong *et al.* [11]) are compared with ours. All these four methods can produce an overall good performance and enhance vessels more or less. But the results obtained by our approach are visually better. Our approach can generate a clean image with a similar appearance of raw image, while other methods may produce irregular color appearance. Besides, the boundaries of vessels are sharper than other methods and the background is cleaner without noises. The results obtained by the other methods contain more blurred vessels. This is more obvious in the zoomed-in views of the selected patches.

The enhanced images can benefit other retinal image processing, such as vessel segmentation or vessel tracking. To demonstrate the improvement, we conduct experiments on both blurry and enhanced images for different tasks. Figure 3 displays the results of vessel segmentation and tracking with the same model for each task. As we can see, more vessels can be segmented out on our enhanced image, while the wrong vessel tracking will be corrected.

4.2 Ablation Test on Feature Constraint

To evaluate the performance of our dynamic feature constraint compared with image-level constraint, we measure the results of these two models separately. The second last row of Table 1 displays the quantitative results of model without feature constraint. It's a bit worse than the final model of our approach both on no-reference and full-reference assessment. Figure 4 shows the visual comparison of these two models.

Overall, the results enhanced by image-level constraint are satisfied. However, the image-level model usually produces artifacts when the area is extremely unclear (shown with blue boxes in figure), and some parts are not well enhanced compared with our final model (shown with red boxes in figure). Retinal structures can hardly be seen on the extreme blurred images, which becomes a barrier for image-level model to guide the enhanced generator. On the other hand, the pattern in the feature space can still provide the guidance, this is also the reason why our proposed feature-level constraint works better than image-level one.

5 Conclusion

We propose a novel deep learning approach to enhance blurry retinal images in a data-driven fashion. It is trained in a weakly supervised manner without the requirement of paired data, which is extremely important for medical images where not much paired data are available. These two advantages (*i.e.* data-driven fashion and unpaired data) make our approach flexible and could be easily used. The proposed approach outperforms the state-of-art retinal image enhancement methods in both visual and quantitative evaluation. Furthermore, it can be used to improve the performance of image processing, including retinal vessel segmentation and tracking. The proposed method can be extend from retinal images to other medical images.

References

1. Chen, Y.S., Wang, Y.C., Kao, M.H., Chuang, Y.Y.: Deep photo enhancer: unpaired learning for image enhancement from photographs with gans. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 6306–6314 (2018)
2. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in Neural Information processing Systems, pp. 2672–2680 (2014)
3. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE Trans. Pattern Anal. Mach. Intell. **33**(12), 2341–2353 (2011)
4. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 5967–5976. IEEE (2017)
5. Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z., Paul Smolley, S.: Least squares generative adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2794–2802 (2017)

6. Mitra, A., Roy, S., Roy, S., Setua, S.K.: Enhancement and restoration of non-uniform illuminated fundus image of retina obtained through thin layer of cataract. Comput. Methods Program. Biomed. **156**, 169–178 (2018)
7. Nie, D., et al.: Medical image synthesis with context-aware generative adversarial networks. In: Descoteaux, Maxime, Maier-Hein, Lena, Franz, Alfred, Jannin, Pierre, Collins, D.Louis, Duchesne, Simon (eds.) MICCAI 2017. LNCS, vol. 10435, pp. 417–425. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_48
8. Polesel, A., Ramponi, G., Mathews, V.J.: Image enhancement via adaptive unsharp masking. IEEE Trans. Image Process. **9**(3), 505–510 (2000)
9. Sevik, U., Kose, C., Berber, T., Erdol, H.: Identification of suitable fundus images using automated quality assessment methods. J. Biomed. Opt. **19**(4), 046006 (2014)
10. Starck, J.L., Murtagh, F., Candès, E.J., Donoho, D.L.: Gray and color image contrast enhancement by the curvelet transform. IEEE Trans. Image Process. **12**(6), 706–717 (2003)
11. Xiong, L., Li, H., Xu, L.: An enhancement method for color retinal images based on image formation model. Comput. Methods Program. Biomed. **143**, 137–150 (2017)
12. 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)
13. Zhou, M., Jin, K., Wang, S., Ye, J., Qian, D.: Color retinal image enhancement based on luminosity and contrast adjustment. IEEE Trans. Biomed. Eng. **65**(3), 521–527 (2018)
14. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: International Conference on Computer Vision, pp. 2242–2251. IEEE (2017)