# MATCHING THE CHARACTERISTICS OF FUNDUS AND SMARTPHONE CAMERA IMAGES

by

Sukesh Adiga V, Jayanthi Sivaswamy

in

*IEEE International Symposium on Biomedical Imaging*
(*ISBI-2019*)

Hilton Molino Stucky, Venice Italy

Report No: IIIT/TR/2019/-1

Centre for Visual Information Technology
International Institute of Information Technology
Hyderabad - 500 032, INDIA
April 2019

# MATCHING THE CHARACTERISTICS OF FUNDUS AND SMARTPHONE CAMERA IMAGES

*Sukesh Adiga V,* *Jayanthi Sivaswamy*

Center for Visual Information Technology, IIIT-Hyderabad, India

## ABSTRACT

Fundus imaging with a Smartphone camera (SC) is a cost-effective solution for the assessment of retina. However, imaging at high magnification and low light levels, results in loss of details, uneven illumination and noise especially in the peripheral region. We address these problems by matching the characteristics of images from SC to those from a regular fundus camera (FC) with an architecture called ResCycle-GAN. It is based on the CycleGAN with two significant changes: A residual connection is introduced to aid learning only the correction required; A structure similarity based loss function is used to improve the clarity of anatomical structures and pathologies. The proposed method can handle variations seen in normal and pathological images, acquired even without mydriasis, which is attractive in screening. The method produces consistently balanced results, outperforms CycleGAN both qualitatively and quantitatively, and has more pleasing results.

***Index Terms*—** Fundus image, Style mapping, Cycle-GAN, Unsupervised learning.

## 1. INTRODUCTION

Fundus images are commonly used by ophthalmologists to diagnose retinal diseases, with diabetic retinopathy being a major example. A fundus camera (FC) is a digital camera capable of high level of zoom due to the complex optics of a low power microscope at the front end. Thus, enabling high quality and high-resolution imaging of the fundus (or retina). It is therefore expensive and bulky. Recently, the smartphone camera (SC) has been explored for retinal imaging with a relatively low-cost lens attachment [1, 2]. This innovation has two significant advantages: much lower cost and a high degree of portability. However, even without a special lens, natural images captured by an SC and a standard DSLR camera differ in colour, definition/detail, especially of small objects. Imaging of the retina is even more challenging: calls for capturing a $45^{\circ}$ field of view (FOV) of the retina (spanning 132.32 sq. mm [3]) with an SC with a special lens, under illumination of a LED-based flash. This limits the ability to capture fine details such as capillaries.

Challenges in SC images include (i) noise due to low light conditions and CMOS sensors; (ii) uneven illumination, with typically darker periphery due to the curved retinal structure; (iii) dust/flash-induced artefacts; and (iv) variable image quality depending on camera specification of the mobile device. Both (i) and (ii) are acute in non-mydriatic imaging conditions.

Ophthalmic experts routinely see/read images in hospitals/clinics acquired by an FC. Hence, reading images acquired with an SC in screening scenarios will require some adaptation, without which screening can become erroneous with a slower throughput. Matching the standards/quality of the images from SC and FC is a solution. Standard image enhancement approaches proposed for FC images [4, 5] are inappropriate for this task, given the complex sources of problems in SC images. Kohler et al. [6] offer a solution to improve retinal image acquired with a custom-designed, low-cost camera with an adaptive and incremental frame averaging. Imperfect alignment of the frames blurs the image, and hence registration is done before averaging which increases the acquisition time.

In this paper, we propose a mapping solution to transform the SC retinal images (henceforth just referred to as SC images) such that its characteristics are closer or similar to those of FC images. The mapping will aim to preserve the integrity of structural details and introduce no artefacts. Noise removal is not within the scope of this work.

## 2. METHOD

The SC image requires illumination correction, structure enhancement (such as vessels, optic disk (OD), lesions) and flash artefact suppression for better clinical and automatic diagnosis. Further, it is also desired to match its characteristics to that of an FC image to facilitate experts who are used to reading FC images. Solving all these problems at once is very challenging and can be attempted by learning an appropriate mapping from SC to FC image. The problem at hand is similar image-to-image translation [7] which relies on paired image data. In the medical domain, acquisition of paired data is very challenging. Hence, the need is to learn image-to-image translation *without* paired data. Among the many solutions proposed for unsupervised image-to-image translation [8, 9, 10], the CycleGAN [11] has shown excellent results and hence, is taken as a source of inspiration for the proposed method.

Our aim is to learn mapping functions between SC and FC images (more compactly referred to as S and F respectively) in an unsupervised manner. The CycleGAN [11] learns to map an image from a source to the target domain with the two domains being quite different, for example, horse $\leftrightarrow$ zebra, winter $\leftrightarrow$ summer, etc. In our problem, the source and target domain is same (retina), and the aim is to only change the characteristics of an image without losing any structural details. Thus, the CycleGAN is modified by introducing a residual connection between the generator from input to the output end. The proposed architecture is called as ResCycleGAN (Fig. 1). It consists of two generators $G_F$ and $G_S$, which learn the mapping from S to F and F to S, respectively. Besides, two discriminators $D_S$ and $D_F$ learn to distinguish between real/fake S and F images, respectively. The ResCycleGAN is trained to minimise an objective function made of three terms: an adversarial loss [12], a cycle-consistency loss, and an identity loss. These are described next.

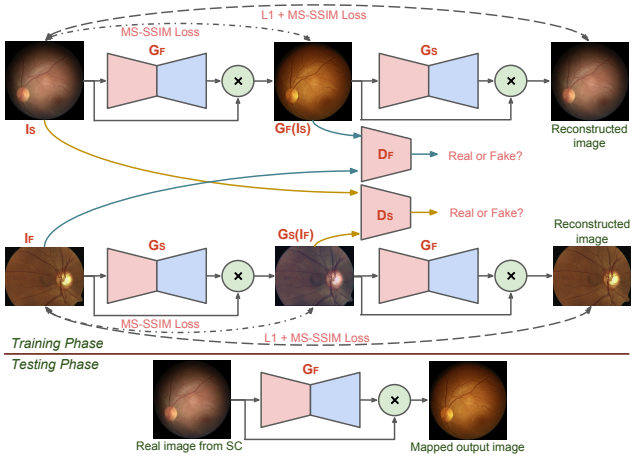


**Fig. 1:** Schematic of the proposed architecture

**Adversarial loss:** The adversarial loss generally serves to match the distribution of the generated output with the reference image. Here, it is used match the characteristics of SC to FC domain. This loss is applied to both the generator $G_F$ and $G_S$. A least-squares function [13] is used for adversarial loss for stable training and generating high-quality results. The adversarial loss for the generator $G_F$ and its corresponding discriminator $D_F$ is given as

$$\mathcal{L}_{GAN}(G_F, D_F) = D_F(G_F(I_S))^2 + (1 - D_F(I_F))^2$$

where $I_S$ and $I_F$ denote *unpaired* SC and FC images. In the training phase, $G_F$ tries to generate an image $G_F(I_S)$ close to real FC image, while $D_F$ tries to distinguish between the generated image $G_F(I_S)$ and real sample $I_F$. $G_F$ aims to minimize this loss against an adversary $D_F$ that tries to maximize it, i.e. $min_{G_F} max_{D_F} \mathcal{L}_{GAN}(G_F, D_F)$. Similarly an adversarial loss for generator $G_S$ and its discriminator $D_S$ are also defined, i.e $min_{G_S} max_{D_S} \mathcal{L}_{GAN}(G_S, D_S)$.

**Cycle-Consistency Loss:** This loss is used to measure the reconstruction capability of the network. i.e. The reconstructed images from $G_S(G_F(I_S))$ and $G_F(G_S(I_F))$ are needs to be identical to their inputs $I_S$ and $I_F$. The $l_1$ or $l_2$ norm is a popular choice for the loss function in a reconstruction problem, but they do not correlate well with the human perception, which is critical in our application as the end user can be a medical expert. The multi-scale, structure similarity index (MS-SSIM) [14] based loss addresses this issue while handling the variations in scale. Hence, we define the cycle-consistent loss function as a combination of $l_1$ norm and MS-SSIM and define it as follows

$$\begin{aligned}\mathcal{L}_{cycle}(G_F, G_S) = \delta_1 \cdot \mathcal{L}_{\text{MS}}(G_S(G_F(I_S)), I_S) \\ + (1 - \delta_1) \cdot \mathcal{L}_{l_1}(G_S(G_F(I_S)), I_S) \\ + \delta_2 \cdot \mathcal{L}_{\text{MS}}(G_F(G_S(I_F)), I_F) \\ + (1 - \delta_2) \cdot \mathcal{L}_{l_1}(G_F(G_S(I_F)), I_F)\end{aligned}$$

where $\mathcal{L}_{l_1}$ and $\mathcal{L}_{\text{MS}}$ are standard $l_1$ norm and MS-SSIM metric. The weights are set to $\delta_1 = \delta_2 = 0.85$ as per [15] and MS-SSIM is computed over three scales.

**Identity Loss:** This loss generally helps preserve colour composition between the input and generated images, whereas, in the application at hand, the colour palette is camera-dependent. The generator has to learn a mapping to either SC or FC fundus images while preserving the integrity of anatomical structures. Hence, a structure similarity function (or MS-SSIM) is suitable for identity loss. This is defined as

$$\mathcal{L}_{ss}(G_F, G_S) = \mathcal{L}_{\text{MS}}(G_F(I_S), I_S) + \mathcal{L}_{\text{MS}}(G_S(I_F), I_F)$$

MS-SSIM is once again computed over three scales.

**Overall training loss:** The overall training loss for the network is defined as a combination of the three losses as

$$\begin{aligned}\mathcal{L}(G_F, G_S, D_F, D_S) = \mathcal{L}_{GAN}(G_F, D_F) + \mathcal{L}_{GAN}(G_S, D_S) \\ + \lambda_1 \cdot \mathcal{L}_{cycle}(G_F, G_S) + \lambda_2 \cdot \mathcal{L}_{ss}(G_F, G_S)\end{aligned} \quad (1)$$

where $\lambda_1$ and $\lambda_2$ are weights for the loss terms.

## 3. IMPLEMENTATION

The architecture of our ResCycleGAN is adopted from CycleGAN [11]. The encoding layer in the generator had 4 blocks of $4 \times 4$ convolution (CONV) of stride 2 followed by LeakyReLU activation and Instance Normalization [16]. The decoding layer had blocks of $4 \times 4$ CONV of stride $\frac{1}{2}$, followed by ReLU activation and Instance Normalization. Skip connections were used from encoding to decoding layer for blocks having the same size. The final layer combined the decoded feature map with a $4 \times 4$ CONV with ReLU. The input and the final CONV layer are multiplied to derive the generator output as shown in Fig. 1. The final CONV layer learns the correction required for SC image to match to FC image.

The discriminator network has layers similar to the encoding layer, followed by a $4 \times 4$ CONV with ReLU.

The ResCycleGAN was trained to minimize the objective function $\mathcal{L}$ (Eq. 1) by alternatively updating $G_{F/S}$ with fixed $D_{F/S}$ and vice versa. The network was trained with patches of size $256 \times 256$ after normalisation to a range of [0,1]. The weights are set to $\lambda_1 = 10$ and $\lambda_2 = 1$. The optimisation was with an Adam solver [17] with an initial learning rate of 0.0002 and batch size of 1. The network was trained for 200000 iterations. The entire code was implemented in Keras library using python and executed on NVIDIA GTX 1080 GPU with 12GB RAM on a core i7 processor. In the testing phase, only the generator $G_F$ is used. The SC image with the original size is given to the generator $G_F$ to produce a mapped image (with characteristics similar to the FC images) is derived as shown in Fig. 1.

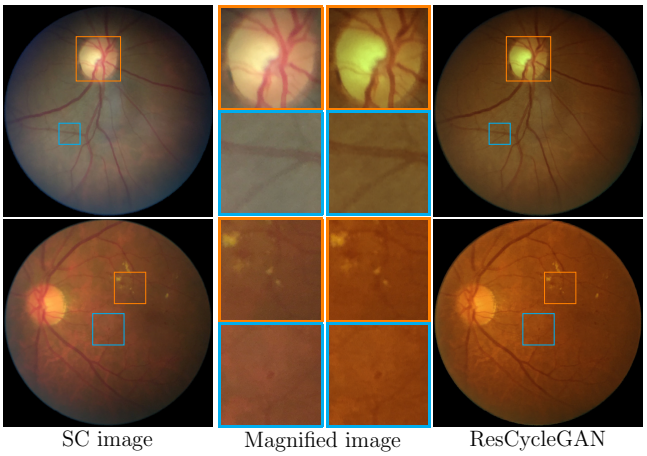


**Fig. 2:** Sample results for ResCycleGAN for images without (top) and with pathologies (bottom).

## 4. RESULTS

### 4.1. Dataset and Evaluation

265 FC images acquired (with mydriasis) with a Zeiss FF450 Plus camera were obtained from the authors of a Diabetic Retinopathy study [1]. A total of 540 SC images, the majority without mydriasis, were obtained from the *Fundus on Phone* (a product of Remidio Innovative Solutions Pvt. Ltd.) at $45^\circ$ FOV using iPhone 6. Both SC and FC images included pathological cases and were of varying quality. A 50% split was done to form the training and testing datasets for SC images. All FC images were used for training the network.

Both qualitative and quantitative evaluation of the proposed ResCycleGAN was done. A quantitative assessment was done using two metrics: $Q_v$ score [18] and the Bhattacharyya distance $D_b$ for comparing the characteristics (histograms) of mapped and FC image.

### 4.2. Performance analysis

Sample original SC images (first column) and their mapped results (last column) are shown in Fig. 2 along with magnified views of two sub-regions per image (middle two columns). The ResCycleGAN results (whole as well as sub-regions) in Row 1 indicate an improvement in contrast of structures such as OD and vessels as well as a reduction in bluish LED noise in the periphery. The horizontally oriented very thin vessels within OD and thin, dull vessels are distinguishable from the background in the magnified results. Similarly, the mapping is seen to improve the lesion (hard exudate in top and microaneurysm in the bottom sub-image) contrast in Row 2, which can be seen in the magnified image. Overall, the mapping is seen to change the colour profile and produce a balanced illumination and contrast.

In order to assess the effectiveness of the modification done to a CycleGAN, two mappings were generated: one with CycleGAN (trained with the same setting as ResCycleGAN) and the other with proposed ResCycleGAN. Two sample results are shown in Fig. 3. The images shown are cases of imaging with/without (top/bottom) mydriasis. The tissue background in CycleGAN results look more synthetic (Row 1) with heavy smoothing of the background erasing vessel, vessel reflections; the OD is also saturated. In the second example in Row 2, the CycleGAN produces a completely uncommon palette with optic cup disappearing, which is unacceptable. The result of ResCycleGAN on the other hand has structural details with a balanced illumination and contrast. The CycleGAN was trained for 400000 iteration which is twice the number of iterations for the ResCycleGAN. The shorter training for the latter is due to the residual connection which helps in learning.

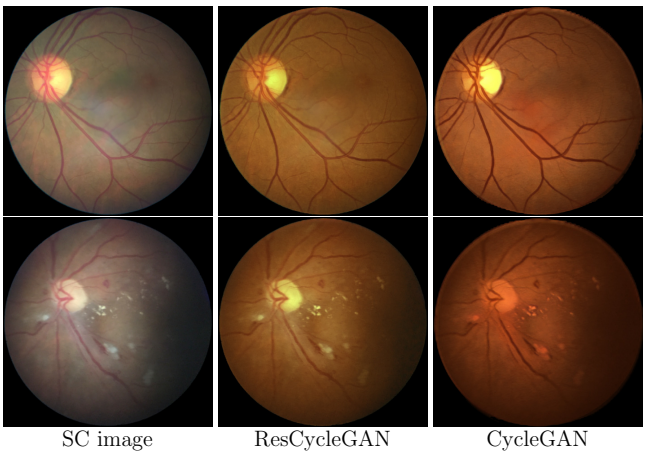


**Fig. 3:** Comparison of ResCycleGAN with CycleGAN outputs.

A quantitative assessment is challenging when no reference image is available. To make a meaningful evaluation of the mapped results, we use a metric to assess the vessel quality ($Q_v$ score [18]) and a metric to assess the similarity ($D_b$ Bhattacharyya distance) between the mapped results (de-

noted as O) and FC images. Higher $Q_v$ values indicate better quality in terms of noise and blur. This score was computed for 270 test images and is presented in Table 1. The similarity is assessed by computing $D_b$ between colour (HSI space) histograms. Average histograms were computed over 270 SC images, their mapped outputs and 265 FC images. $D_b(FC, X)$; $X = SC$ or $O$, is computed for the average histogram pairs and reported separately for the chromatic (C: H and S) and achromatic (AC: I) components in Table 1.

**Table 1:** Quantitative comparison of performance using $Q_v$ and $D_b$ on SC images.

| | $Q_v$ score | $D_b$ (C / AC) |
|---|---|---|
| SC images | $0.0189 \pm 0.0104$ | 0.1656 / 0.0883 |
| CycleGAN [11] | $0.0263 \pm 0.0143$ | 0.0058 / 0.0288 |
| ResCycleGAN | $\mathbf{0.0334 \pm 0.0175}$ | **0.0014 / 0.0166** |

The results indicate that ResCycleGAN outperforms CycleGAN in both $Q_v$ (the difference is statistically significant as $p < 0.05$) and $D_b$ values. This implies the mapping improves vessel contrast while attaining a good match with FC characteristics. Further, the match in characteristics is superior for both AC and C components.

**Fig. 4:** Comparison of standard retinal image enhancement with the proposed mapping. Left to right: SC image, results of our method and enhancement [5].

Finally, we present a comparison with a recently reported unsupervised enhancement method for retinal images [5]. Sample images (without mydriasis) along with the processed results are shown in Fig. 4. Since [5] essentially stretches luminosity and contrast, it leads to a heightened contrast and luminosity (last column) in the results without a colour shift. However, an unwanted bluish peripheral artefact is seen in the results. In contrast, our results (middle column) exhibit an overall balanced improvement.

## 5. CONCLUSION

A ResCycleGAN solution was proposed to match the characteristics of SC images to mydriatic FC images successfully. To the best of our knowledge, this is the first attempt to do such a mapping. The key strengths of our method are: it preserves the integrity of structures with a balanced illumination correction between the peripheral and centre region with no introduction of artefacts; the results are consistently good for images with/without pathologies as well as images acquired with/without mydriasis. Hence, our solution can aid ophthalmic experts; fast processing requiring 5.2 sec/image. One can also explore the method's use a preprocessing stage for adapting CAD systems developed for FC images.

## 6. ACKNOWLEDGEMENT

The authors thank Dr. A Sivaraman and Dr. R Rajalakshmi for providing fundus images for our experiments.

## 7. REFERENCES

[1] R. Rajalakshmi et al., "Validation of smartphone based retinal photography for diabetic retinopathy screening," *PLoS One*, vol. 10, no. 9, pp. e0138285, 2015.

[2] A. Bastawrous et al., "Clinical validation of a smartphone based adapter for optic disc imaging in kenya," *JAMA ophthalmology*, vol. 134, no. 2, pp. 151–158, 2016.

[3] H. Kolb et al., "Facts and figures concerning the human retina–webvision: The organization of the retina and visual system," 1995.

[4] G. D. Joshi et al., "Colour retinal image enhancement based on domain knowledge," in *Proc. of ICVGIP*. IEEE, 2008, pp. 591–598.

[5] M. Zhou et al., "Color retinal image enhancement based on luminosity and contrast adjustment," *IEEE Trans. on Biomedical Eng.*, vol. 65, no. 3, pp. 521–527, 2018.

[6] T. Köhler et al., "Quality-guided denoising for low-cost fundus imaging," in *Bildverarbeitung für die Medizin*, pp. 292–297. Springer, 2012.

[7] P. Isola et al., "Image-to-image translation with conditional adversarial networks," *Proc. of CVPR*, 2017.

[8] T. Kim et al., "Learning to discover cross-domain relations with generative adversarial networks," in *Proc. of ICML*. 2017, vol. 70, pp. 1857–1865, PMLR.

[9] Z. Yi et al., "Dualgan: Unsupervised dual learning for image-to-image translation.," in *Proc. of ICCV*, 2017, pp. 2868–2876.

[10] M. Liu et al., "Unsupervised image-to-image translation networks," in *Proc. of NIPS*, 2017, pp. 700–708.

[11] J. Zhu et al., "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. of ICCV*, 2017.

[12] I. Goodfellow et al., "Generative adversarial nets," in *Proc. of NIPS*, 2014, pp. 2672–2680.

[13] X. Mao et al., "Least squares generative adversarial networks," in *Proc. of ICCV*. IEEE, 2017, pp. 2813–2821.

[14] Z. Wang et al., "Multi-scale structural similarity for image quality assessment," in *Proc. of ACSSC*. IEEE, 2003, vol. 2, pp. 1398–1402.

[15] H. Zhao et al., "Loss functions for image restoration with neural networks," *IEEE Trans. on Computational Imaging*, vol. 3, no. 1, pp. 47–57, 2017.

[16] U. Dmitry et al., "Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis," in *Proc. of CVPR*. 2017, pp. 4105–4113, IEEE.

[17] D. Kinga et al., "A method for stochastic optimization," in *Proc. of ICLR*, 2015, vol. 5.

[18] T. Köhler et al., "Automatic no-reference quality assessment for retinal fundus images using vessel segmentation," in *26th Int. Symp. on Computer-Based Medical Systems*. IEEE, 2013, pp. 95–100.