Deblurring adaptive optics retinal images using deep convolutional neural networks

XIAO FEI,1,2 JUNLEI ZHAO,1,2 HAOXIN ZHAO,1,2 DAI YUN,1,2,* AND YUDONG ZHANG1,2

1The Laboratory on Adaptive Optics, Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu 610209, China
2Key Laboratory on Adaptive Optics, Chinese Academy of Sciences, Chengdu 610209, China
*daiyunqq@163.com

Abstract: The adaptive optics (AO) can be used to compensate for ocular aberrations to achieve near diffraction limited high-resolution retinal images. However, many factors such as the limited aberration measurement and correction accuracy with AO, intraocular scatter, imaging noise and so on will degrade the quality of retinal images. Image post processing is an indispensable and economical method to make up for the limitation of AO retinal imaging procedure. In this paper, we proposed a deep learning method to restore the degraded retinal images for the first time. The method directly learned an end-to-end mapping between the blurred and restored retinal images. The mapping was represented as a deep convolutional neural network that was trained to output high-quality images directly from blurry inputs without any preprocessing. This network was validated on synthetically generated retinal images as well as real AO retinal images. The assessment of the restored retinal images demonstrated that the image quality had been significantly improved.

© 2017 Optical Society of America under the terms of the OSA Open Access Publishing Agreement

OCIS codes: (100.1455) Blind deconvolution; (100.3010) Image reconstruction techniques; (170.4470) Ophthalmology; (010.1080) Active or adaptive optics.

References and links

1. J. Liang, D. R. Williams, and D. T. Miller, “Supernormal vision and high-resolution retinal imaging through adaptive optics,” J. Opt. Soc. Am. A 14(11), 2884–2892 (1997).
2. D. T. Miller and A. Roorda, Adaptive Optics in Retinal Microscopy and Vision, in Handbook of Optics (2009).
3. H. Bao, C. Rao, Y. Zhang, Y. Dai, X. Yao, and Y. Fan, “Hybrid filtering and enhancement of high-resolution adaptive-optics retinal images,” Opt. Lett. 34(22), 3484–3486 (2009).
4. J. Arines, “Partially compensated deconvolution from wavefront sensing images of the eye fundus,” Opt. Commun. 284, 1548–1552 (2011).
5. C. Rao, Y. Tian, and H. Bao, AO-Based High Resolution Image Post-Processing (InTech, 2012).
6. J. C. Christou, A. Roorda, and D. R. Williams, “Deconvolution of adaptive optics retinal images,” J. Opt. Soc. Am. A 21(8), 1393–1401 (2004).
7. L. Blau and L. M. Mugnier, “Marginal blind deconvolution of adaptive optics retinal images,” Opt. Express 19(23), 23227–23239 (2011).
8. H. Li, J. Lu, G. Shi, and Y. Zhang, “Real-time blind deconvolution of retinal images in adaptive optics scanning laser ophthalmoscopy,” Opt. Commun. 284, 3258–3263 (2011).
9. G. Chenegros, L. M. Mugnier, F. Lacombe, and M. Glanc, “3D phase diversity: a myopic deconvolution method for short-exposure images: application to retinal imaging,” J. Opt. Soc. Am. A 24(5), 1349–1357 (2007).
10. A. Lazareva, M. Asad, and G. Slabaugh, “Learning to Deblur Adaptive Optics Retinal Images,” International Conference Image Analysis and Recognition (2017).
11. T. F. Chan and C. K. Wong, “Total variation blind deconvolution,” IEEE Trans. Image Process. 7(3), 370–375 (1998).
12. W. Ouyang, C. C. Loy, X. Tang, X. Wang, X. Zeng, S. Qiu, P. Luo, Y. Tian, H. Li, and S. Yang, “DeepID-Net: Deformable deep convolutional neural networks for object detection,” IEEE Transactions on Pattern Analysis & Machine Intelligence PP, 1–1 (2016).
13. M. Hradis, J. Kotera, P. Zemčík, and F. Sroubek, “Convolutional Neural Networks for Direct Text Deblurring,” in British Machine Vision Conference, 2015.
14. C. Dong, C. C. Loy, K. He, and X. Tang, “Image Super-Resolution Using Deep Convolutional Networks,” IEEE Trans. Pattern Anal. Mach. Intell. 38(2), 295–307 (2016).
15. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in International Conference on Neural Information Processing Systems, (2012), 1097–1105.
1. **Introduction**

Early detection of retinal pathologies such as glaucoma, age related macula degeneration (AMD) or retinitis pigmentosa (RP) is crucial in dealing with these conditions and calls for in vivo eye fundus imaging with a cellular level resolution. However, direct observation of the retina suffers from various optical aberrations of the eye and the imaging resolution is severely limited. To compensate for ocular aberrations, adaptive optics (AO) technology was introduced and near diffraction-limited retinal images were achieved in 1997 [1]. Since then, AO has been successfully integrated into a variety of retina imaging modalities, including conventional fundus camera, confocal scanning laser ophthalmoscope (SLO), and optical coherence tomography [2]. However, because of the limited closed-loop bandwidth of an AO imaging system, and its limited aberration measurement and correction accuracy, residual aberration still blurs the retinal images. Furthermore, intraocular scatter and the uncontrolled physiological vibration of the eye accompanied by cardiopulmonary movement also blur retinal images. Moreover, the pollution caused by various sources of noise, such as stray light and CCD readout noise, further degrades the quality of retinal images [3]. Interpretation of such blurred retinal images is difficult for researchers or doctors and an appropriate image post-processing method is indispensable to make up for the limitation of AO retinal imaging procedure.

In an AO retinal imaging system, the blurring process can be described as a convolution between a real retinal image and the point spread function (PSF) which is contributed by ocular aberrations and optical system’s aberrations. Additional improvements of the retinal image’s quality can be obtained by using image restoration techniques such as image deconvolution. The non-blind deconvolution algorithms based on PSF measurement have been proposed to compensate for ocular residual aberrations left by AO [4]. In these algorithms, the PSFs are measured by wavefront sensor and employed in the deconvolution process. These non-blind deconvolution algorithms require less computation and can be used in near real-time, but their performance depends on the accuracy of PSF measurement. However, the PSF measured by wavefront sensor is not accurate due to various kinds of noises in the AO imaging system. Therefore, ordinary non-blind deconvolution algorithm is not a suitable method for post-processing of AO images [5]. Thus, a more generalized technique named blind deconvolution has been proposed to restore the AO retinal images [6–9]. This type of image deconvolution allows for recovery of the retinal images and the PSF.
distributions simultaneously by physical constraints about the target and the PSF. However, this blind deconvolution algorithm has a drawback of getting trapped in local minima that makes it hard to find a unique solution, especially when there is only a single blurred image to be restored [5]. Recently, a learning-based blind deconvolution method has been introduced to deblur AO retinal images [10]. This method uses a Random Forest to learn the mapping of retinal images onto the space of blur kernels expressed in terms of Zernike coefficients. For a blurred image, the optimal vector of Zernike coefficients is predicted by the trained Random Forest and then the corresponding PSF is reconstructed. For this method, the proposed deconvolution approach is specifically designed for deconvolution of retinal images acquired with a commercially available flood-illuminated AO instrument (rtx1, Imagine Eyes, Orsay, France) and the blur kernel is modeled by the physics of this AO system. For another AO system, the approach should be redesigned. Besides, the learning process neglects the noise term and assumes that images are corrupted only by convolution blur kernel. Furthermore, the learning-based approach is developed as one of the stages in the image processing framework and a non-blind deconvolution algorithm is still needed to restore the blurred image. Therefore, the final performance of this method depends on the effect of all the steps.

In this paper, to our knowledge, we proposed a deep learning method to restore the degraded AO retinal images for the first time. Unlike previous methods, this method does not need to predict the PSF of the imaging system but directly learns an end-to-end mapping between the blurred and restored retinal images. The mapping is represented as a deep convolutional neural network that is trained to output high-quality images directly from blurry inputs without any preprocessing. The proposed method is data-driven and the advantage of this is that it is relatively easy to model forward imaging process and generate data pairs as opposed to hard task of inverting the imaging process by hand. Once a specific complex imaging process (e.g. space-variant blur, non-Gaussian noise, or saturation) is modeled and the training data is generated, the method can naturally handle this complex process which may be problematic for traditional deconvolution. This network was validated on synthetically generated retinal images as well as real AO retinal images. The assessment of the restored retinal images demonstrated that the image quality had been significantly improved.

2. Method

2.1 Image model

Generally, the blind deconvolution process of AO retinal images is defined as the task of finding an original retinal image $x$, and possibly a residual PSF $k$, from an observed blurred image $y$ which is created as

$$ y = x * k + n, $$

where ‘$*$’ denotes two-dimensional convolution and $n$ is an independent additive noise [5]. This is an ill-posed problem with an infinite number of solutions. A possible solution to the problem is a minimization of a suitable cost function, e.g.

$$ \hat{x} = \arg \min_x \| y - k * x \|_2^2 + r(x). $$

Here the data term $\| y - k * x \|_2^2$ forces the solution $x$ to be a plausible source of the observed image. The regularizing prior term $r(x)$ expresses the knowledge of the image statistics (e.g. sparsity priors, edge priors) and forces the solution to be a nice (meets the prior knowledge without obvious faults such as spurious oscillations) image [11]. The key idea of this solution is to address the ill-posedness by a suitable choice of prior, for example by using natural image statistics, or by otherwise modifying either the minimized function or the optimization
process, but doing so increases computational complexity and is often too difficult to find a unique solution.

In our work, an alternative approach was used to deconvolve AO retinal images by directly modeling the restoration process as a general function

\[ \hat{x} = F(y, \theta) \]  

(3)

with parameter \( \theta \), which can be learned from training data. The function \( F(y, \theta) \) implicitly incorporates both the data term and the prior term from Eq. (2), and directly maps blurred images to restored images. This data-driven approach has been successfully used in many computer vision fields, such as object detection [12], text deblurring [13], and image super-resolution [14]. The advantage of this approach is that it is relatively easy to model the forward imaging process and generate \((x, y)\) pairs. Given that the ground truth is provided in the form of training data \( T \) composed of real image and blurred image pairs \((x_i, y_i)\), the parameter \( \theta \) can be learned by minimizing a suitable loss function, e.g.

\[
L(\theta) = \frac{1}{N} \sum_{i=1}^{N} \| F(y_i; \theta) - x_i \|^2,
\]

(4)

where \( N \) is the number of training samples. Once the parameter \( \theta \) is learned, restoration of a blurred image \( y_i \) can be straightforward by using Eq. (3). To efficiently learn the parameter, a deep convolutional neural network architecture is introduced into our work which we will discuss in the next section.

2.2 Convolutional neural networks for retinal image deblurring

The convolutional neural networks (CNNs) date back to decades ago and deep CNNs have recently shown an explosive popularity partially due to its success in image classification [15]. CNNs are the state-of-art function approximators for computer vision problems and are well suited for the deconvolution task. They can naturally incorporate inverse filters and have enough computational capacity to model the blind deconvolution process [13]. We directly restore clean and sharp retinal images from corrupted images by deep CNNs. The proposed network is composed of five layers which combine convolutions with element-wise rectified linear units (ReLU). An overview of the network is depicted in Fig. 1. The network conceptually consists of three operations: feature extraction and representation, non-linear mapping and reconstruction. The definition and interpretation of each operation are as follows.
Consider a single blurred image $y$, our first layer is expressed as an operation $F_1$: 

$$F_1(y) = \max(0, W_1 \ast y + B_1),$$

where $W_1$ and $B_1$ represent the filters and biases respectively, and $\ast$ denotes convolution. In our network, $W_1$ corresponds to $n_1$ filters of support $f_1 \times f_1$. Intuitively, $W_1$ applies $n_1$ convolutions on the image, and each convolution has a $f_1 \times f_1$ kernel size. The output is composed of $n_1$ feature maps. $B_1$ is an $n_1$-dimensional vector, and each element of the vector is associated with a filter. We apply the ReLU (max(0, $x$)) on the filter responses. In a word, the $F_1$ operation extracts features from the blurred image $y$ and represents them as feature maps. In Fig. 1, some sample feature maps are drawn for visualization. The next three layers can be expressed as:

$$F_l(y) = \max(0, W_l \ast F_{l-1}(y) + B_l), l = 2, 3, 4,$$

where $W_l$ contains $n_l$ filters of size $n_1 \times f_2 \times f_2$, and $B_l$ is $n_2$-dimensional. Through these layers, the original $n_1$ feature maps extracted by the first layer are nonlinearly mapped to $n_2$ new feature maps as shown in Fig. 1. It is possible to add more convolutional layers to increase the non-linearity, but this will increase the complexity of the model. Considering the performance and complexity, three non-linear mapping layers are used in our network. The last layer of the network is designed to reconstruct the final deblurred image as:

$$F_5(y) = W_5 \ast F_4(y) + B_5.$$

Here $W_5$ corresponds to one filter of size $n_2 \times f_3 \times f_3$, and $B_5$ is a scalar. For the network, its performance is affected by filter numbers ($n_1, n_2$) and filter sizes ($f_1, f_2, f_3$). In our work, the network’s performance was investigated with different filter numbers and filter sizes. The result shows that a reasonably larger filter number and filter size can result in better...
performance at the price of increased complexity. But if the filter number and filter size are too large, the performance of the network will improve marginally or even degrade as shown in Table 1. Therefore, the choice of the network scale should always be a trade-off between performance and speed. For our situation, the network parameters are finally set as $n_1 = 64$, $n_2 = 64$, $f_1 = 11$, $f_2 = 3$, $f_3 = 5$.

| Network settings (n1,n2,f1,f2,f3) | Complexity (Parameter number) | Average PSNR (dB) |
|------------------------------------|-------------------------------|-------------------|
| (16,16, 9,1,3)                     | 2273                          | 23.78             |
| (32,32, 9,1,3)                     | 6081                          | 23.92             |
| (64,64, 9,1,3)                     | 18305                         | 24.21             |
| (128,128, 9,1,3)                   | 61185                         | 24.23             |
| (64,64, 7,1,3)                     | 16257                         | 24.12             |
| (64,64, 11,3,3)                    | 119169                        | 24.39             |
| (64,64, 11,3,5)                    | 120193                        | 24.41             |
| (64,64, 11,5,7)                    | 318337                        |                   |

Table 1. The Performance and Complexity of the Network with Different Filter Numbers and Filter Sizes

Training and evaluation are performed on the synthetically generated data sets which are described in the next section.

2.3 Generation of training data set

The training data set is very important for the network. On the one hand, the training data set needs to be large enough to make the CNN work well, and much effort is needed to acquire these real retinal image pairs recorded with and without AO. On the other hand, the images recorded with AO are still corrupted by residual wavefront aberrations and noises. Considering these reasons, our CNN was trained on synthetically generated retinal images. To generate the training data set, the following four steps were employed. Firstly, a set of ideal retinal images with eccentricities ranging from 0.5 mm to 1.5 mm from the foveal center were created and normalized by using the algorithm described in reference [16]. Secondly, for each ideal image, a set of PSFs were generated to simulate the residual optical aberrations of the eye. The PSFs are computed as:

$$\text{PSF} = \left| \text{FFT} \left( P(x, y) \exp\left(-\frac{2\pi}{\lambda} \phi(x, y) \right) \right) \right|^2,$$

where $P$ is the pupil function, $\lambda$ is the wavelength of imaging beam and $\phi(x, y)$ is the wavefront phase error which is defined as:

$$\phi(x, y) = \sum_{m=0}^{44} a_m Z_m(x, y),$$

where $a_m$ and $Z_m$ are Zernike coefficients and Zernike polynomial, respectively. Each $\phi(x, y)$ represents an ocular aberration and the values of the Zernike coefficients are generated from a statistical model of the wavefront aberrations in healthy eyes reported in reference [17]. For a retinal imaging AO system, the quantity of residual aberrations after correction is usually small [2]. To simulate the correction result of AO, the low-order aberrations were set to zero and the values of Zernike coefficients from $a_0$ to $a_{20}$ were scaled by 0.1. Thirdly, for each ideal image, a set of synthetic blurred retinal images were generated by convolving the ideal images with these generated PSFs. Finally, Gaussian noise was added to the blurred images and all images were normalized from the original range of intensities to $[0, 1]$.

In our work, 100 ideal retinal images were firstly created and then 5000 PSFs were generated for each of them to generate the training data set. Thus, the CNN was trained on a data set containing 500,000 image pairs.
2.4 Training

Learning the end-to-end mapping function $F$ requires the estimation of network parameters $\theta = \{W_i, B_i\}$. As in previous work [13], this is achieved by minimizing the loss between the restored images $F(y_i, \theta)$ and the corresponding ground truth images $x_i$:

$$\arg\min_{\theta, W} \frac{1}{2 N} \sum_{i=1}^{N} \| F(y_i) - x_i \|^2 + 0.0005 \| W \|^2. \quad (10)$$

The weight decay term $0.0005 \| W \|^2$ is used to improve convergence [15]. To avoid border effects during training, all convolutional layers have no padding, and the network produces a smaller output. Therefore, the loss function is calculated only by the difference between the central pixels of $x_i$ and the network output. Although a fixed image size is used for training, the CNN can be applied on images of arbitrary sizes during testing.

The optimization method we used was Stochastic Gradient Descent with momentum [18]. The learning rate and the momentum were set to be 0.001 and 0.95 respectively. We initialized weights from uniform distribution with variance equal to $1/n_{in}$, and $n_{in}$ was the size of the respective convolutional filter. We implemented this network using the Caffe deep learning framework [19].

3. Experimental results

The training process was done offline using the data set and method as described above. Training took roughly one hour on GPU GTX 1080 Ti. Figure 2 shows examples of learned first-layer filters. As shown in Fig. 2, each learned filter had its specific functionality. For example, the filters a and b were similar to edge detectors at different directions, the filters c and d were similar to Laplacian/Gaussian filters, and the filter e was similar to a texture extractor. After completing training, the network can be directly used to restore blurred retinal images. For a blurred image with a size of 256x256, it took about 1s to restore using ordinary CPU. The network’s performance was validated on both synthetically generated retinal images and real AO retinal images.

3.1 Experimental validation using synthetic data

To test the image restoration performance of the network, a testing data set was produced according to the generation of training data set as described above. 30 ideal retinal images were generated so as to reproduce different retinal eccentricities (range from 0.5 mm to 1.5 mm from the foveal center) and each of them was convolved with 100 PSFs. Gaussian noise with standard deviation uniformly sampled from [0, 0.03] was also added to the blurred images. Thus, the testing data set was composed of 3000 image pairs representing the unseen data. This data set was used to evaluate the generalization of the network.

For this synthetic data set, the peak signal to noise ratio (PSNR) and structural similarity index measurement (SSIM) of the ideal images, blurred images and the restored images were respectively computed to evaluate the network’s performance. The PSNR and SSIM are widely-used metrics for quantitatively evaluating image restoration quality, and are partially related to the perceptual quality. As it is important to automated identification of
photoreceptor cells for retinal images, the cone numbers are obtained by using a cone counting algorithm developed by Li and Roorda [20] and the error rate of cone counting is computed. The computation process is as follows. Firstly, the cone numbers are obtained for each image by using the mentioned algorithm. Secondly, the error rate of cone counting \((ER)\) for each image is computed by comparing the obtained cone numbers \((coneNo)\) with ground truth \((truthNo)\) according to Eq. (11). Finally, the statistical results of the error rates of all the images in the testing data set are computed.

\[
ER = \left| \frac{coneNo - truthNo}{truthNo} \right| \times 100\%
\]  

(11)

These three parameters, including PSNR, SSIM and ER were used to evaluate the network’s performance, simultaneously. We compared these three parameters obtained from our method with those from the blind deconvolution method developed by Sroubek [21]. In this method, blind deconvolution is represented as a \(l1\)-regularized optimization problem, where a solution is found by alternately optimizing with respect to the image and kernel blurs. For a faster convergence, minimization is addressed with an augmented Lagrangian method (ALM). The implementation is from the publicly available code provided by the author.

To evaluate the performance of the proposed method, quantitative assessment was performed on the 3000 synthetically blurred retinal images. Figure 3 represents typical examples of the original images, corresponding blurred images and restored images. The corresponding PSNR and SSIM values are also shown in the figure. The statistical results are shown in Table 2, where we have the following findings. (i) Quantitative comparison in terms of PSNR and SSIM indicated that the image quality had been significantly improved for the entire data set. (ii) After processing using the proposed method, the error rate of cone counting decreased significantly. (iii) The proposed method outperformed the ALM method in terms of PSNR, SSIM and error rate of cone counting.
Fig. 3. Deconvolution of three representative synthetic images imitating at different eccentricities.

Table 2. Quality Assessment of Blurred and Processed Synthetic Retinal Images

|                        | Blurred images | Images restored with the proposed method | Images restored with the ALM method |
|------------------------|----------------|----------------------------------------|------------------------------------|
| PSNR                   | 16.01 ± 1.39   | 24.48 ± 2.76                           | 17.14 ± 1.73                       |
| SSIM                   | 0.41 ± 0.13    | 0.83 ± 0.07                            | 0.58 ± 0.15                        |
| Error rate of cone counting,% | 37.79 ± 15.07 | 5.13 ± 2.32                            | 11.30 ± 4.13                       |

Data are expressed as means ± standard deviation.

3.2 Experimental validation using real AO retinal images

The trained CNN was used to restore real AO retinal images. In the case of real retinal images, there was no ground truth data for evaluation of the results and a No-Reference Image Quality Assessment algorithm (NR-IQA) was needed. In our work, the Blind Image Quality Index (BIQI) method [22] was used to evaluate the network’s performance. This method is widely-used for quantitatively evaluating the quality of No-reference images. Given an image, a quality score between 0 and 100 is computed by the method and smaller score value represents better image quality. Besides, two automatic cone counting algorithms were used to detect the photoreceptor cells for each retinal image. The first algorithm is developed by Li and Roorda [20] and the second one is a CNN based method which has been developed recently [26]. The counting results were compared with manual cone counts and the error rates were computed according to Eq. (11) for both algorithms. Three observers from our lab were asked to carry out manual counting on the real high-resolution retinal images. For each image, every observer counted five times and the mean value was computed for each observer, then the final manual cone number (truthNo) was determined by averaging the three mean values. In the following statement, \( Error rate_1 \) represents the error rate of cone counting of the first automatic detection algorithm and \( Error rate_2 \) represents the second
The performance of the proposed method was also compared with the ALM blind image deconvolution method.

To test the CNN’s generalization, 20 high-resolution AO retinal images captured from different AO systems at various eccentricities (range from 0.3 mm to 1.8 mm from the foveal center) had been restored. A few examples are shown in Fig. 4 to Fig. 6. For these real retinal images, the inter observer variability varies with the retinal image. Generally, the image with better quality has smaller inter observer variability and better performance of the automatic methods. For example, the coefficients of variation for the original retinal images shown in Fig. 4 and Fig. 6 are 3.3% and 2.7%, and Error rate1 are 11.15% and 10.01%, and Error rate2 are 6.45% and 5.81%, respectively. The inter observer counts demonstrated an average coefficient of variation of 3.1% for the 20 high-resolution retinal images.

The original retinal image shown in Fig. 4 was acquired with our flood-illuminated AO system [25] and its BIQI score was 62.08. The manual cone count for this image was 314. After applying the proposed restoration process, the BIQI score decreased to 55.90 and the Error rate1 decreased from 11.15% to 3.82% and the Error rate2 decreased from 6.45% to 3.13%.

The original retinal images shown in Fig. 5 and Fig. 6 were captured with AOSLO systems by other research group as described in [23, 24]. As shown in the figures, both original images were severely blurred by the residual wavefront error and the photoreceptor cells can hardly be resolved. To further demonstrate the performance of the restoration methods, the average image power spectra was also computed and showed in the figures as well as the BIQI score and error rate of cone counting. It can be seen that, after applying the proposed restoration process, the image quality had been improved. The values of the reasonable photoreceptor cell spatial frequencies ranging from 50 to 100 cycles/degree [5] were increased and the values of frequencies higher than 150 cycles/degree which correspond to noise were decreased. It showed that the visibility of photoreceptors was improved in the restored images. The decreased error rates of cone counting also indicated this conclusion.

To evaluate the performance of the proposed method, quantitative assessment was performed on the 20 high-resolution AO retinal images. The statistical results are shown in Table 3, where we have the following findings. (i) The proposed method can improve the quality of retinal images efficiently and can facilitate better distinction of photoreceptor cells. (ii) The proposed method outperformed the ALM method in terms of BIQI score and error rate of cone counting. The results of experimental validation using real AO retinal images showed that the proposed method was well generalized and could be used to deblur retinal images acquired with different AO systems.

Fig. 4. Result of cone detection on a representative retinal image obtained by the algorithm of Li and Roorda. (a). original image (1.2 mm eccentricity from the foveal center); (b). restored by the proposed method; (c). restored by the ALM method. The scale bar is 50 µm.
Fig. 5. Deconvolution result of a retinal image captured by AOSLO system. (a) original image (0.9 mm eccentricity from the foveal center); (b) restored by the proposed method; (c) restored by the ALM method; (d) the corresponding normalized image power spectra. The scale bar is 50 µm.
Table 3. Quality Assessment of Blurred and Restored Real AO Retinal Images

|                    | Blurred images | Images restored with the proposed method | Images restored with the ALM method |
|--------------------|----------------|------------------------------------------|------------------------------------|
| BIQI score         | 61.98 ± 4.33   | 51.62 ± 4.02                             | 56.03 ± 3.27                       |
| Error rate1, %     | 10.08 ± 4.28   | 4.31 ± 1.76                              | 7.87 ± 3.64                        |
| Error rate2, %     | 6.12 ± 2.73    | 3.68 ± 1.04                              | 5.22 ± 2.16                        |

Data are expressed as means ± standard deviation.

4. Conclusion and discussion

In this paper, we have demonstrated that deep convolutional neural networks are able to directly perform blind deconvolution of AO retinal images, and this deblurring method achieves good restoration quality. To our knowledge, this is the first time to restore AO retinal images using CNN. Unlike previous methods, this method does not need to predict the PSF of the imaging system but directly learns an end-to-end mapping between the blurred and restored retinal images. The proposed method is data-driven and the advantage of this is that it is relatively easy to model forward imaging process and generate data pairs as opposed to hard task of inverting the imaging process by hand. Once a specific complex imaging process (e.g. space-variant blur, non-Gaussian noise, or saturation) is modeled and the training data is generated, the method can naturally handle this complex process which may be problematic for traditional deconvolution. The method was validated on synthetically generated images as
well as real AO retinal images. The obtained results demonstrate that the method is well
generalized and can restore blurred retinal images efficiently and this facilitates better
distinction of photoreceptor cells.

It should be noted that there is space for improvement in the proposed method. The
limitations of our method are as follows. (i) The network was trained on space invariant PSFs
with Gaussian noise, more complex imaging process such as space-variant blur, non-Gaussian
noise could be further modeled and trained with the network. (ii) Compared with some
deconvolution algorithms [4, 8], the proposed method takes more time to deconvolve an
image. For now, our network was composed by convolutional layers and element-wise
rectified linear units and the optimization method we used was Stochastic Gradient Descent
with momentum. Regularization methods such as dropout and batch normalization layers
could be added to further improve the network’s performance. We intend to improve the
proposed method and combine the deblurring process with cell counting into a single network
in the future work.

**Funding**

National Natural Science Foundation of China (61378064); National High Technology
Research and Development Program of China (2015AA020510).

**Disclosures**

The authors declare that there are no conflicts of interest related to this article.