Adversarial Reconstruction CNN for Illumination-Robust Frontal Face Image Recovery and Recognition

Liping Yang, Key Laboratory of Optoelectronic Technology and Systems, MOE, Chongqing University, Chongqing, China
https://orcid.org/0000-0001-9194-7248

Bin Yang, Key Laboratory of Optoelectronic Technology and Systems, MOE, Chongqing University, Chongqing, China

Xiaohua Gu, School of Electrical Engineering, Chongqing University of Science and Technology, Chongqing, China

ABSTRACT

This article proposes an adversarial reconstruction convolution neural network (ARCNN) for non-uniform illumination frontal face image recovery and recognition. The proposed ARCNN includes a reconstruction network and a discriminative network. The authors employ GAN framework to learn the reconstruction network in an adversarial manner. This article integrates gradient loss and perceptual loss terms, which are able to preserve the detailed and spatial structure image information, into the overall reconstruction loss function to constraint the reconstruction procedure. Experiments are conducted on the typical illumination-sensitive dataset, extended YaleB dataset. The reconstructed results demonstrate that the proposed ARCNN approach can remove the illumination and shadow information and recover natural uniform illuminated face image from non-uniform illuminated ones. Face recognition results on the extended YaleB dataset demonstrate that the proposed ARCNN reconstruction procedure can also preserve the discriminative information of face image for classification task.

KEYWORDS

Adversarial Learning, Convolution Neural Network, Deep Learning, Face Recognition, Illumination-Robust, Image Recovery, Machine Learning, Pattern Recognition

1. INTRODUCTION

High quality images are important for improving the performance of image analysis systems and the visualization of human beings. The actual systems such as face analysis (Punnappurath, 2015; Zohra, 2017), retinal scanning (Adal, 2018; Anitha, 2010), intelligent transportation (Bulan, 2017; Xu, 2014), underwater target recognition (Hou, 2018), human-computer interaction (Zhou, 2017) all want the input images to be of high quality. However, images that captured in unconstrained conditions or using not ideal imaging devices, usually exhibit non-uniform illumination distribution and low contrast, which cause detail loss in dark and overexposure regions. This paper considers the problem of recovering the detailed information of an object under standard illumination from a non-uniform illuminated image.

Over the years, researchers present a surge of qualitative and quantitative studies on 2D and 3D non-uniform illumination processing (Gao, 2018; Xu, 2018). Since the authors only focus on
the recovery methods of 2D non-uniform illumination images, this paper gives a brief survey of 2D illumination processing methods. Readers interested in non-uniform illumination processing of 3D objects can refer to (Xu, 2014, 2018). In 2D non-uniform illumination processing domain, removing illumination and re-mapping illumination are common traditional strategies (Gao, 2018; Shin, 2015), which are able to factorize an image into its intrinsic components, i.e. illumination, shape and reflectance components (Georgoulis, 2018). Representative approaches in this domain can roughly divide into two categories according to their diverse theoretical backgrounds, including illumination compensation/normalization methods and illumination-invariant representation methods. Illumination normalization methods attempt to redistribute the intensities of an input image in a more appropriate representation, which is less sensitive to lighting changes by applying a simple gray-scale intensity adjustment. Histogram equalization (HE), block histogram equalization and logarithmic transform are representative methods of this category (Arici, 2009). Notwithstanding their ease of implementation and the visualization improvement on lighting normalization, these methods can hardly handle shadow or highlight effects of non-uniform illumination images. In contrast, illumination-invariant representation methods try to estimate and remove unwanted illumination. Retinex-based image enhancement methods, which can recall the visual content of dark regions as well as keep the visual realism, are the mainstream methods (Gao, 2018; Shin, 2015; Chen, 2006). According to the assumption that illumination corresponds to low frequency information, Retinex-based methods estimate illumination information using low-pass Gaussian filter (Shin, 2015; Jobson, 1997) or total variation (TV) normalization model (Chen, 2006; Ng, 2011). However, due to the problems of low-pass Gaussian filter in edge preserving and TV normalization model in side effect, halo artifacts and details loss may occur in the recovered images. Recently, several optimization strategies, which globally locate important gradients of images, are proposed to estimate illumination information of images (Fu, 2015, 2016; Guo, 2017). Nevertheless, these methods may cause the damage of illumination edges and result in light source confusion and artifacts. In summary, although the state-of-the-art non-uniform illumination recovery methods have achieved great success, it still requires study to improve the recovery quality of images from non-uniform illumination condition, especially from extreme non-uniform illumination situations.

Recovering detailed information of an object from a non-uniform illuminated image is a special image-to-image transformation task. In recent years, both convolution neural networks (CNNs) and generative adversarial networks (GANs) are trained for various image-to-image transformation tasks (Wang, 2018), such as image de-noising (Zhang, 2017), image super resolution (Dong, 2016; Ledig, 2017), cross-domain image translation (Yi, 2017), etc. CNNs, which discover the optimal mapping from an input image to the transformed image by minimizing the discrepancy between the output image and ground-truth image (Dong, 2016), are very efficient for image feature representation. GANs, which estimate generate models via an adversarial training process alternating between identifying and faking (Goodfellow, 2014), are beneficial for generating realistic images. In (Wang, 2018), a perceptual adversarial network (PAN), whose adversarial architecture is composed of two CNNs, is proposed for image-to-image transformation. Experimental results show that PAN has a great capability of accomplishing image-to-image transformations. To deal with illumination recovery issue, several deep learning based approaches are proposed. In (Lore, 2017), a deep auto-encoder approach, i.e. LLNet, which is composed by three de-noising auto-encoder layers, is presented to enhance the quality of natural low-light images. Experimental results demonstrate that deep auto-encoders are effective tools to learn underlying signal characteristics and noise structures from low-light images without handwriting. In (Li, 2018), a trainable CNN method for weakly illuminated image enhancement is proposed. This method estimate the illumination map of an input image using CNN and enhance the image based on Retinex model. In (Georgoulis, 2018), authors present a two-step deep learning approach that estimates reflectance and illumination information from a single image. CNNs are employed to estimate a reflectance map and decompose the reflectance map into reflectance parameters and an illumination map.
While CNNs and GANs have been widely studied for image-to-image transformation tasks, the authors are not aware of any existing work integrating CNNs and GANs together for image recovery from a non-uniform illuminated image. In this paper, the authors propose an adversarial reconstruction convolution neural networks (ARCNN) for non-uniform illumination frontal face image recovery. The proposed ARCNN includes a reconstruction network and a discriminative network. Both the reconstruction and discriminative networks are convolution neural networks. The reconstruction network functionally consists of four steps, including feature decomposition, multi-scale feature mapping and fusion, nonlinear mapping and image reconstruction. As that the gradient information is much stable to illumination variant and the perceptual reconstruction can preserve image content and overall spatial structure, the authors incorporate gradient and perceptual constraints into the overall loss function of the reconstructed network. The authors employ the GAN framework for training the reconstruction network and the discriminative system in an adversarial manner to prompt the quality of reconstructed images. The experimental results both for face image recovery and face recognition on Extended YaleB dataset show that the proposed ARCNN is efficient for removing non-uniform illumination information and preserving individual classification information, which is useful for face image visualization and classification tasks. In addition, this paper is an extension of the authors’ previous conference paper published in the proceeding of 17th IEEE International Conference on Cognitive Informatics & Cognitive Computing (Yang, 2018). The extension are two folds:

- This article employs the GAN framework to prompt the reconstruction performances of the illumination-robust deep reconstruction CNN proposed in (Yang, 2018).
- A perceptual reconstruction loss term is included in the overall loss function of reconstruction network to preserve image content and overall spatial structure, to make the reconstructed images present naturally.

The paper is organized as follows: Section 2 introduces the proposed adversarial reconstruction convolution neural networks, and section 3 provides experimental results on two non-uniform illumination face image datasets. Section 4 concludes the paper.

2. ADVERSARIAL RECONSTRUCTION CONVOLUTION NEURAL NETWORKS

This section describes the proposed adversarial reconstruction convolution neural networks in detail. The authors first revisit the GAN framework and then depict the architecture and learning procedure of the proposed ARCNN, shown in Figure 1.

2.1 GAN Framework

GAN framework consists of two networks, a generator $G$ and a discriminator $D$, trained by competing with each other. The generator $G$ is trained to generate samples $G(z)$ from some noise vectors $z$, such that $G(z)$ are difficult to discriminate from real samples $x$ using the discriminator $D$. On the other hand, the discriminator $D$ is trained to distinguish the generated samples $G(z)$ from the real samples. In this competing way, the performances of both generator $G$ and discriminator $D$ can be improved. Therefore, the learning procedure of GAN is a two-player min-max game with the following objective function:

$$\min_{G} \max_{D} \mathbb{E}_{x}[\log D(x)] + \mathbb{E}_{z}[\log(1 - D(G(z)))]$$
2.2 The Proposed ARCNN

The proposed ARCNN employs the structure of GAN, such that it contains a reconstruction network and a discriminative network that are simultaneously trained using the min-max formulation of GAN shown in Eq. 1. The reconstruction network of ARCNN tries to recover a uniform illuminated face image using a non-uniform illuminated face image as input. The discriminative network tries to distinguish the recovery face images from the reference uniform illuminated ones.

2.2.1 The Reconstruction Network Architecture of ARCNN

To recover a uniform illuminated face image from a non-uniform illuminated one, a common strategy is to separate illumination component and reflection component of a face image firstly, and then try to remove unwanted illumination information. Due to its flexibility and superiority for image feature representation, convolution neural network is used to build the reconstruction network of ARCNN for illumination-robust face image recovery. The reconstruction network, shown in Figure 1(a), consists of seven convolution layers, which include four stages, namely feature decomposition, multi-scale feature mapping and fusion, nonlinear mapping and reconstruction. In order to get multi-scale features of the input image, we set convolutional kernels size of each layer in feature decomposition and multi-scale feature mapping and fusion stage in a pyramid way from 7×7 to 3×3. At the nonlinear mapping stage, 1×1 convolutional kernels are used to save the nonlinear mapping coefficients. Finally, at the reconstruction stage, a relative large 5×5 convolutional kernel size is employed to make the reconstruction face image much smooth and natural. Table 1 lists the hyper-parameters of the reconstruction network. Figure 2 illustrates an example of the outputs at different stages of the reconstruction network.

The feature decomposition stage, which includes the first three convolution layers, \( l_1 \), \( l_2 \) and \( l_3 \), with different kernel sizes, is used for decomposing information of the input non-uniform face image into different channels. As that a face image contains both illumination and reflection components, which respectively correspond to the low and high frequency components of the image, the convolution operations at this stage may be able to divide and reserve the reflection components into different channels. However, the boundaries between the bright and dark regions in a non-uniform illuminated image are difficult to remove in a single scale feature maps. Therefore, the reconstruction network cascades a multi-scale feature mapping and fusion stage after the feature decomposition stage. The multi-scale feature mapping and fusion stage consists of two convolution layers, \( l_4^{(i)} \) and \( l_5^{(i)} \) (Figure 1a. The architecture of ARCNN: the reconstruction network)
Each of which is composed of three parallel convolution operation with different kernel size. The two-layer multi-scale convolution and fusion operations on the feature maps can suppress the boundaries of the bright and dark regions and reserve the high-frequency reflection components in the non-uniform face images. As shown in Figure 2, the first two stages produce a $3n_3$-dimensional feature map, which contains only high-frequency reflection components, for each input image. To obtain a uniform-illuminated image, the reconstruction network employs a nonlinear mapping, which transforms the $3n_3$ dimensional feature map to a $n_6$-dimensional feature map, at the third stage to mix uniform illumination information into each channel of the output. The ReLU activation function is applied on the output of each convolution layer of these three stages for non-linear mapping (Nair, 2010). Therefore, the operations $F_i^{(k)}$ of these convolution layers can be expressed as:

$$F_i^{(k)}(X) = \max \left( 0, W_i^{(k)} * X_t + B_i^{(k)} \right)$$

where $W_i^{(k)}$ and $B_i^{(k)}$ represent the $k$-th scale convolution kernel and bias of layer $l$, $X_t$ is the input of layer $l$ ($k = 1$ for $l = 1, 2, 3, 6$ and $k = 1, 2, 3$ for $l = 4, 5$). At each layer, $W_i^{(k)}$ corresponds to $n_l$ convolution kernels of support $f_i^{(k)} \times f_i^{(k)}$. Finally, the $n_6$-dimensional feature map is linearly combined at the reconstruction stage of our architecture to produce the uniform illumination face image. The operation is:

$$F_i(X) = W_i * X + B_i$$

Here, $W_i$ is an $n_i \times f_i \times f_i$ convolution kernel and $B_i$ is a scalar.

### 2.2.2 Learning of the Reconstruction Network

In order to recover uniform-illuminated face image from a non-uniform illuminated one, the output of the reconstruction network should be as similar as possible to the uniform illuminated image. Let us denote a non-uniform illumination face image as $X$ and its corresponding uniform illumination face image (ground truth) as $Y$. Suppose that the mapping of the reconstruction network is $F$, which can recover from $X$ an image $F(X, \Theta)$. Learning of the reconstruction network is to estimate the optimal parameters $\Theta$ of the network.

Given a set of non-uniform illuminated face images $\{X_i\}$ and their corresponding uniform-illuminated face images $\{Y_i\}$ ($i = 1, 2, \cdots, N$), traditional convolution neural network uses mean squared error (MSE) as the loss function to estimate the parameters $\Theta$, that is:

$$L_{MSE}(\Theta) = \frac{1}{N} \sum_{i=1}^{N} \|F(X_i, \Theta) - Y_i\|^2$$
Using MSE as the loss function is predominant for image super-resolution. However, it may cause the reconstructed face image to be too smooth. To overcome this drawback, the authors incorporate a gradient loss into the reconstruction loss in (Yang, 2018):

$$L_{\text{Grad}}(\Theta) = \frac{1}{N} \sum_{i=1}^{N} \left( \|F(X_i, \Theta) - Y_i\|^2 + \beta \|\nabla F(X_i, \Theta) - \nabla Y_i\|^2 \right)$$

| layer | # channels ($n_i$) | kernel size ($f_i \times f_i$) | stride |
|-------|-----------------|-------------------------------|--------|
| $I_1$ | 64              | 7×7                           | 1      |
| $I_2$ | 48              | 5×5                           | 1      |
| $I_3$ | 32              | 3×3                           | 1      |
| $I_4^{(1)}$ | 32          | 7×7                           | 1      |
| $I_4^{(2)}$ | 32          | 5×5                           | 1      |
| $I_4^{(3)}$ | 32          | 3×3                           | 1      |
| $I_5^{(1)}$ | 32          | 7×7                           | 1      |
| $I_5^{(2)}$ | 32          | 5×5                           | 1      |
| $I_5^{(3)}$ | 32          | 3×3                           | 1      |
| $I_6$ | 32              | 1×1                           | 1      |
| $I_7$ | 1               | 5×5                           | 1      |

Figure 2. An example illustrates the outputs at different stages of the reconstruction network

Using MSE as the loss function is predominant for image super-resolution. However, it may cause the reconstructed face image to be too smooth. To overcome this drawback, the authors incorporate a gradient loss into the reconstruction loss in (Yang, 2018):
where \( \nabla F(X_i, \Theta) \) and \( \nabla Y_i \) are the gradient images of reconstructed and ground truth images, \( \beta \) is a hyper-parameter to balance the MSE and gradient loss. However, due to the influence of the MSE and gradient losses, which are per-pixel losses, the reconstruction results of (Yang, 2018) introduce new high-frequency noise while retaining the detailed information, making the reconstruction results look unnatural.

In this paper, the authors introduce a perceptual reconstruction loss, which depends on high-level features extracting from a pre-trained classification network (the loss network) \( \Phi \). It has been demonstrated that the perceptual reconstruction loss can preserve image content and overall spatial structure (Johnson, 2016), which is valid for removing the high-frequency noise. To compute the perceptual reconstruction loss \( L_p \), the pre-trained VGG-19 classification network is employed, shown in Figure 3. Both the reconstructed image \( F(X_i, \Theta) \) and the corresponding ground truth \( Y_i \) are served into the VGG-19 network. The perceptual reconstruction loss is then calculated using the Euclidian distance between output feature maps of \( F(X_i, \Theta) \) and \( Y_i \) on the last convolution layer of VGG-19:

\[
L_p(\Theta) = \frac{1}{N} \sum_{i=1}^{N} \| \phi(F(X_i, \Theta)) - \phi(Y_i) \|^2
\]

where \( \phi(\bullet) \) represents the output feature map of VGG-19.

The overall loss function of training the reconstruction network is:

\[
L_R(\Theta) = \frac{1}{N} \sum_{i=1}^{N} \left( \| F(X_i, \Theta) - Y_i \|^2 + \beta \| \nabla F(X_i, \Theta) - \nabla Y_i \|^2 + \gamma \| \phi(F(X_i, \Theta)) - \phi(Y_i) \|^2 \right)
\]

2.2.3 The Discriminative Network of ARCNN

The discriminative network of ARCNN aims to discriminate recovered face images from the reference uniform illuminated ones. The authors employ a CNN classification network, which consists of 1 convolution layer, 4 residual blocks, 2 fully connection layers (shown in Figure 1(b)), to construct the discriminative network. Batch normalization is used in each residual block.

The discriminative network and the reconstruction network are trained in an adversarial manner to promote the performance of the reconstruction network. Therefore, the loss function of training the discriminative network is:

\[
L_D(\theta) = \sum_{i=1}^{N} \log D_\theta(Y_i) + \sum_{i=1}^{N} - \log(D_\theta(F(X_i, \Theta)))
\]
where \( N \) is the number of images, \( D_\theta(Y) \) represents the probability of classifying a reference uniform illuminated face image into a reconstructed one, \( D_\theta(F(X, \Theta)) \) represents the probability of classifying a reconstructed face image into a reference one.

3. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed ARCNN approach, experiments for non-uniform illumination face image reconstruction and face recognition were carried out on extended YaleB (Lee, 2005) face dataset. The extended YaleB face dataset contains 16,128 images of 38 individuals under 9 poses and 64 illumination conditions. Here we restrict using frontal face views across illumination variations in the experiments.

3.1 Experiments on Non-Uniform Illumination Face Image Reconstruction

To evaluate performance of the non-uniform illuminated face image reconstruction of the proposed ARCNN, we randomly selected all the frontal face images of 30 individuals as training set, the face images of the rest 8 individuals are used as test set. In the training set, face images with uniform illumination are used as training labels of the reconstruction network. All the other 63 non-uniform illuminated face images of each individual are used as inputs of the reconstruction network. As that the number of samples in the dataset is quite limited, each face image is overlapped segmented into 12 patches to enlarge the size of the training set. In addition, all the images in the training set are mirrored for data augmentation.

The authors compare the reconstruction results of the proposed ARCNN with that of histogram equalization (HE), multi-scale Retinex (MSR) and IRDRCNN in (Yang, 2018). Figure 4 illustrates an example of the reconstructed images using different reconstruction methods. Comparing with HE and MSR, it is clear that the CNN-based methods, i.e. the proposed ARCNN and IRDRCNN, can efficiently remove the non-uniform illumination information. The reconstructed results of the proposed ARCNN are more natural than that of IRDRCNN, which demonstrates that employing perceptual reconstruction loss into the reconstruction loss function is efficient for suppressing high-frequency noise and preserving overall image content.

To measure the reconstruction performance of ARCNN, the authors introduce signal-to-noise ratio (SNR) here. Face image with uniform illumination distribution is regard as a reference signal. SNRs of all the other illumination images and reconstructed images are calculated. Figure 5 illustrates the peak SNRs of all the images of one individual. It is clear that the reconstruction images of ARCNN have higher SNR values than that of HE, MSR and IRDRCNN in most cases. Therefore, we can infer that the reconstruction results of non-uniform illumination face images using ARCNN are much closer to the uniform illuminated ones. Table 2 shows the mean and variance of peak SNRs for different reconstruction methods. The highest mean SNR and smallest variance of ARCNN indicate that the reconstructed face images of ARCNN separate out much unwanted illumination information and meanwhile preserves the detailed features, which is useful for face recognition task.

3.2 Experiments on Illumination-Robust Face Recognition

To verify that the reconstructed face images of ARCNN have superiority in maintaining identity discriminant information, face recognition experiments on extended YaleB dataset using original images and reconstructed images are conducted. Three traditional feature descriptors, namely LBP (Heikkilä, 2009), POEM (Vu, 2012) and RGHF (Yang, 2016)[9] are used to represent facial features. The nearest neighbour classifier is used for classification.

We feed all images of the extended YaleB dataset into the proposed ARCNN and IRDRCNN to generate two reconstructed image datasets, then face recognition using LBP, POEM and RGHF features are conducted in the original and two reconstructed image datasets separately. For each
dataset, we randomly select $i$ ($i = 1, 2, \cdots, 10$) samples of each individual to train the classifier and the rest samples for testing. All the experiments are conducted 10 times. Figure 6 shows the recognition rates of LBP, IRDRCNN+LBP, ARCNN+LBP, POEM, IRDRCNN+POEM, ARCNN+POEM, RGHF, IRDRCNN+RGHF and ARCNN+RGHF versus the number of training samples.

As illustrated in Figure 6(a)~(c), the recognition rates on two reconstructed image datasets significantly outperform that on the original image dataset. The results demonstrate that convolution neural networks are efficient for recovering the detailed information of frontal face images from non-uniform illuminated images. The recovered face images preserve much discriminative information and meanwhile suppress the unwanted illumination noise. Comparing with IRDRCNN, the proposed ARCNN performs much better, especially when the number of training samples is small. This may benefit from the employing of perceptual reconstruction loss and adversarial learning in ARCNN.

To further evaluate the illumination-robustness of the proposed ARCNN, we compare the recognition performance of ARCNN with that of several existing illumination-robust methods, including logarithmic total variation (LTV) (Chen, 2006), logarithmic wavelet transform (LWT) (Zhang, 2009), logarithmic nonsubsampled contourlet transform (LNSCT) (Xie, 2010), convolution neural network (CNN) (Ramaian, 2015) and IRDRCNN, on five subsets of extended YaleB dataset. The five subsets are divided according to the lighting angle of face images shown in Table 3. In the

---

Figure 4. An example of the reconstruction results using different reconstruction methods

![Reconstruction Results](image)
Figure 5. The peak SNRs of all the images of one individual in extended YaleB dataset

Table 2. Mean and variance values of peak SNRs for different reconstruction methods

|               | Original | HE   | MSR   | IRDRCNN | ARCNN |
|---------------|----------|------|-------|---------|-------|
| mean (dB)     | 12.8     | 10.2 | 9.8   | 22.1    | 24.0  |
| variance      | 27.10    | 0.92 | 1.45  | 2.01    | 1.38  |

experiment, we only use the images in Set 1 as training samples, face recognition are conducted on all five subsets.

Table 4 shows the recognition performances of LTV, LWT, LNSCT, CNN, IRDRCNN and ARCNN. The results show that all of these five methods perform quite well on Set 1 and Set 2. With the increase of lighting angle, the recognition rates of LTV, LWT and LNSCT decrease drastically, which indicates that these three methods may not deal non-uniform illumination distribution very well. Due to the extreme illumination condition, the CNN method performs not well in Set 5. Comparing with other four methods, ARCNN and IRDRCNN perform very well on Set 4 and Set 5, whose images are non-uniform illuminated. The results indicate that the proposed ARCNN is illumination-robust.
This article presented an adversarial reconstruction convolution neural network framework for non-uniform illuminated face image recovery and recognition. The proposed framework includes a reconstruction network and a discriminative network. The reconstruction network consists of four stages, including feature decomposition, multi-scale feature mapping and fusion, nonlinear mapping and reconstruction. The authors employ GAN to train the reconstruction network and the discriminative network in an adversarial manner. To preserve image content and overall spatial structure while removing non-uniform illumination information, this article integrate the gradient loss and perceptual reconstruction loss terms into the overall loss function of reconstruction network. Experimental results on the extended YaleB dataset demonstrate that the proposed ARCNN achieves superior reconstruction performance than the state-of-the-art illumination-robust methods.

This article considers the reconstruction of non-uniform illuminated frontal face images. However, pose variance and image corruption may occur together with non-uniform illumination in real application, e.g. video surveillance. Research on non-uniform illuminated face reconstruction in general case will be a future direction of this work.

Figure 6a. Recognition rates of different feature descriptors versus the number of training samples

4. CONCLUSION

This article presented an adversarial reconstruction convolution neural network framework for non-uniform illuminated face image recovery and recognition. The proposed framework includes a reconstruction network and a discriminative network. The reconstruction network consists of four stages, including feature decomposition, multi-scale feature mapping and fusion, nonlinear mapping and reconstruction. The authors employ GAN to train the reconstruction network and the discriminative network in an adversarial manner. To preserve image content and overall spatial structure while removing non-uniform illumination information, this article integrate the gradient loss and perceptual reconstruction loss terms into the overall loss function of reconstruction network. Experimental results on the extended YaleB dataset demonstrate that the proposed ARCNN achieves superior reconstruction performance than the state-of-the-art illumination-robust methods.

This article considers the reconstruction of non-uniform illuminated frontal face images. However, pose variance and image corruption may occur together with non-uniform illumination in real application, e.g. video surveillance. Research on non-uniform illuminated face reconstruction in general case will be a future direction of this work.
ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China under Grant No. 61903054 and the Fundamental Research Funds for the Central Universities under Grant No. 2020CDCGTM002. The authors would like to thank the anonymous reviewers for their constructive advice.
Figure 6c. Recognition rates of different feature descriptors versus the number of training samples

Table 3. Five subsets of extended YaleB dataset

| Set ID | 1  | 2  | 3  | 4  | 5  |
|--------|----|----|----|----|----|
| Lighting angle (deg) | 0~12 | 13~25 | 26~50 | 51~77 | >77 |

Table 4. Recognition performances of existing illumination-robust methods

|       | Set 1 (%) | Set 2 (%) | Set 3 (%) | Set 4 (%) | Set 5 (%) | Avg. (%) |
|-------|-----------|-----------|-----------|-----------|-----------|----------|
| LTV   | 100       | 99.8      | 78.5      | 75.7      | 82.4      | 87.3     |
| LWT   | 100       | 100       | 82.0      | 81.9      | 70.8      | 86.9     |
| LNSCT | 100       | 100       | 83.3      | 88.0      | 84.3      | 91.1     |
| CNN   | 99.6      | 98.8      | 99.0      | 95.1      | 77.1      | 93.9     |
| IRDRCNN | 100     | 99.9      | 99.7      | 98.7      | 99.1      | 99.5     |
| ARCNN | 100       | 100       | 99.8      | 99.9      | 99.9      | 99.9     |
REFERENCES

Adal, K. M., van Etten, P. G., Martinez, J. P., Rouwen, K. W., Vermeer, K. A., & van Vliet, L. J. (2018). An automated system for the detection and classification of retinal changes due to red lesions in longitudinal fundus images. IEEE Transactions on Biomedical Engineering, 65(6), 1382–1390. PMID:28922110

Anitha, J., Vijila, C., & Hemanth, D. (2010). A hybrid genetic algorithm based fuzzy approach for abnormal retinal image classification. International Journal of Cognitive Informatics and Natural Intelligence, 4(3), 29–43.

Arici, T., Dikbas, S., & Altunbasak, Y. (2009). A histogram modification framework and its application for image contrast enhancement. IEEE Transactions on Image Processing, 18(9), 1921–1935. PMID:19403363

Bulan, O., Kozitsky, V., Ramesh, P., & Shreve, M. (2017). Segmentation and annotation-free license plate recognition with deep localization and failure identification. IEEE Transactions on Intelligent Transportation Systems, 18(9), 2351–2363.

Chen, T., Yin, W., Zhou, X. S., Comaniciu, D., & Huang, T. S. (2006). Total variation models for variable lighting face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(9), 1519–1524. PMID:16929737

Dong, C., Loy, C. C., He, K., & Tang, X. (2016). Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2), 295–307. PMID:26761735

Fu, X., Liao, Y., Zeng, D., Huang, Y., Zhang, X., & Ding, X. (2015). A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation. IEEE Transactions on Image Processing, 24(12), 4965–4977. PMID:26336125

Fu, X., Zeng, D., Huang, Y., Zhang, X., & Ding, X. (2016). A weighted variational model for simultaneous reflectance and illumination estimation. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (pp. 2782-2790). Las Vegas, NV: IEEE.

Gao, Y., Hu, H., Li, B., & Guo, Q. (2018). Naturalness preserved nonuniform illumination estimation for image enhancement based on retinex. IEEE Transactions on Multimedia, 20(2), 335–344.

Georgoulis, S., Rematas, K., Ritschel, T., Gavves, E., Fritz, M., Gool, L. V., & Tuytelaars, T. (2018). Reflectance and natural illumination from single-material specular objects using deep learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(8), 1932–1947. PMID:28841552

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. In Proceedings of 27th International Conference on Neural Information Processing Systems (pp. 2672-2680). Montreal, Canada: Academic Press.

Guo, X., Li, Y., & Ling, H. (2017). LIME: Low-light image enhancement via illumination map estimation. IEEE Transactions on Image Processing, 26(2), 982–993. PMID:28113318

Heikkilä, M., Pietikäinen, M., & Schmid, C. (2009). Description of interest regions with local binary patterns. Pattern Recognition, 42(3), 425–436.

Hou, G., Pan, Z., Huang, B., Wang, G., & Luan, X. (2018). Hue preserving-bassed approach for underwater colour image enhancement. IET Image Processing, 12(2), 292–298.

Jobson, D. J., Rahman, Z., & Woodell, G. A. (1997). A multiscale Retinex for bridging the gap between color images and the human observation of scenes. IEEE Transactions on Image Processing, 6(7), 965–976. PMID:18282987

Johnson, J., Alahi, A., & Li, F. F. (2016). Perceptual losses for real-time style transfer and super-resolution. In Proceedings of European Conference on Computer Vision (LNCS vol. 9906). doi:10.1007/978-3-319-46475-6_43

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (pp. 105-114). Honolulu, HI: IEEE.

Lee, K., Ho, J., & Kriegman, D. (2005). Acquiring linear subspaces for face recognition under variable lighting. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(5), 684–698. PMID:15875791
Li, C., Guo, J., Porikli, F., & Pang, Y. (2018). LightenNet: A convolutional neural network for weakly illuminated image enhancement. *Pattern Recognition Letters*, 104, 15–22.

Lore, K. G., Akintayo, A., & Sarkar, S. (2017). LLNet: A deep auto-encoder approach to natural low-light image enhancement. *Pattern Recognition*, 61, 650–662.

Nair, V., & Hinton, G. (2010). Rectified linear units improve restricted Boltzmann machines. In *Proceedings of International Conference on Machine Learning* (pp. 807-814). Haifa, Israel: Academic Press.

Ng, M. K., & Wang, W. (2011). A total variation model for Retinex. *SIAM Journal on Imaging Sciences*, 4(1), 345–365.

Punnappurath, A., Rajagopalan, A. N., Taheri, S., Chellappa, R., & Seetharaman, G. (2015). Face recognition across non-uniform motion blur, illumination, and pose. *IEEE Transactions on Image Processing*, 24(7), 2067–2082. PMID:25775493

Ramaian, N., Ijjina, E., & Mohan, C. (2015). Illumination invariant face recognition using convolutional neural networks. In *Proceeding of International Conference on Signal Processing, Informatics, Communication and Energy Systems*. Kozhikode, India: Academic Press.

Shin, Y., Jeong, S., & Lee, S. (2015). Efficient naturalness restoration for non-uniform illumination images. *IET Image Processing*, 9(8), 662–671.

Vu, N. S., Dee, H. M., & Caplier, A. (2012). Face recognition using the POEM descriptor. *Pattern Recognition*, 45(7), 2478–2488.

Wang, C., Xu, C., Wang, C., & Tao, D. (2018). Perceptual adversarial networks for image-to-image transformation. *IEEE Transactions on Image Processing*, 27(8), 4066–4079. PMID:29993743

Xie, X., Lai, J., & Zheng, W. (2010). Extraction of illumination invariant facial features from a single image using nonsubsampled contourlet transform. *Pattern Recognition*, 42(12), 4177–4189.

Xu, D., Duan, Q., Zheng, J., Zhang, J., Cai, J., & Cham, T. J. (2014). Recovering surface details under general unknown illumination using shading and coarse multi-view stereo. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (pp.1526-1533). Columbus, OH: IEEE.

Xu, D., Duan, Q., Zheng, J., Zhang, J., Cai, J., & Cham, T. J. (2018). Shading-based surface detail recovery under general unknown illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(2), 423–436. PMID:28221993

Xu, Z., Zhi, F., Liang, C., Mei, L., & Luo, X. (2014). Generating semantic annotation of video for organizing and searching traffic resources. *International Journal of Cognitive Informatics and Natural Intelligence*, 8(1), 51–66.

Yang, L., & Li, W. (2016). Low-Rank relative gradient histogram features for illumination-robust face recognition [in Chinese]. *Tien Tzu Hsueh Pao*, 44(8), 1940–1946.

Yang, L., Yang, B., & Gu, X. (2018). A deep reconstruction CNN for illumination-robust face image recovery and recognition. In *Proceeding of IEEE 17th International Conference on Cognitive Informatics & Cognitive Computing* (pp. 417-422). Berkeley, CA: IEEE.

Yi, Z., Zhang, H., Tan, P., & Gong, M. (2017). DaulGAN: unsupervised daul learning for image-to-image translation. In *Proceedings of IEEE International Conference on Computer Vision* (pp. 2849-2857). Venice, Italy: IEEE.

Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. (2017). Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7), 3142–3155. PMID:28166495

Zhang, T., Fang, B., Yuan, Y., Tang, Y., Shang, Z., & Li, D. (2009). Multiscale facial structure representation for face recognition under varying illumination. *Pattern Recognition*, 42(2), 251–258.

Zhou, X., Liu, X., Yang, C., Jiang, A., & Yan, B. (2017). Multi-channel features spatio-temporal context learning for visual tracking. *IEEE Access: Practical Innovations, Open Solutions*, 5, 12856–12864.

Zohra, F. T., Gavrilov, A. D., Zatarain, O. A., & Gavrilova, M. L. (2017). Regression-based automated facial image quality model. *International Journal of Cognitive Informatics and Natural Intelligence*, 11(4), 22–40.
Liping Yang received his Ph.D. degree in Instrument Science and Technology from Chongqing University, Chongqing, China, in 2008. He joined the College of Opto-electronic Engineering, Chongqing University, Chongqing, China, in 2009. He is currently working as an Associate Professor at Chongqing University, Chongqing, China. His research interests include signal processing, pattern recognition and machine learning.

Xiaohua Gu received her M.S. and Ph.D. degrees from Chongqing University in 2005 and 2011, respectively. She is currently an associate professor in Chongqing University of Science and Technology. Her research interests include pattern recognition, intelligent oil-and-gas field, modeling and optimization of complex systems.