LLCNN: A convolutional neural network for low-light image enhancement

Conference Paper · December 2017

DOI: 10.1109/VCIP.2017.8305143

Li Tao
The University of Tokyo

Chuang Zhu
Beijing University of Posts and Telecommunications

Yuan Li
Peking University

Huizhu Jia
Peking University

LLCNN: A Convolutional Neural Network for Low-light Image Enhancement

Li Tao[1], Chuang Zhu[1,*], Guoqing Xiang[1], Yuan Li[1], Huizhu Jia[1,2], Xiaodong Xie[1,2,*]

[1] National Engineering Laboratory for Video Technology, Peking University, Beijing, China

[2] Cooperative Medianet Innovation Center and Beida (Binhai) Information Research, Tianjin, China

***Abstract*—In this paper, we propose a CNN based method to perform low-light image enhancement. We design a special module to utilize multiscale feature maps, which can avoid gradient vanishing problem as well. In order to preserve image textures as much as possible, we use SSIM loss to train our model. The contrast of low-light images can be adaptively enhanced using our method. Results demonstrate that our CNN based method outperforms other contrast enhancement methods.**

***Keywords*—low-light image, contrast enhancement, deep learning, CNN, SSIM loss**

I. INTRODUCTION

Good quality images and videos are important for many tasks. However, not all images are in good qualities because they are captured in various light conditions. When an image is captured under insufficient light condition, the pixel values are in a low-dynamic range, which will cause image quality to descend apparently. Since the whole image appears very dark, it's hard to identify objects or textures clearly. Thus, it is necessary to improve the quality of low-light images.

To enhance image contrast and improve image brightness, several algorithms were proposed. Histogram equalization (HE) methods [1], [2] rearrange pixel values to make them obey uniform distribution. Retinex-theory-based methods [3]–[7] utilize a model which assumes an image is an interaction of illumination and reflectance. Methods based on dehaze model [8] fix pixel values to make them obey a natural distribution. We refer to all these methods as traditional methods.

Recently, convolutional neural network (CNN) achieves impressive progress in several computer vision applications. As to low-level image processing applications, CNN makes several breakthroughs in super resolution [9], image denoising [10], etc. However, to the best of our knowledge, there are no existing works using CNN to conduct low-light image enhancement.

In this paper, we apply CNN to enhance low-light images. We refer to the proposed network as Low-light Convolutional Neural Network (LLCNN). It learns to filter low-light images with different kernels and then combine multiscale feature maps together to generate enhanced images, which seem to be captured under normal light conditions, and reserve original features and textures. Besides, SSIM loss is integrated into our LLCNN to reconstruct more accurate image textures. We compare our results against other methods and the results demonstrate that our method achieves the best performance among common low-light image enhancement methods.

II. RELATED WORK

A. Low-light Image Enhancement Methods

Generally, low-light image enhancement methods can be divided into three categories. HE methods keep relative relationship among pixel values and try to make them obey uniform distribution. Dynamic histogram equalization (DHE) [1] divides the histogram into several parts and performs HE processing in each sub-histogram. Contrast limited adaptive histogram equalization (CLAHE) [2] adaptively limits the extent of contrast enhancement result of HE. For these HE methods, serious color cast problem will be raised and details in darken areas are not enhanced properly in many cases.

Retinex-theory-based methods calculate the illumination of an image, and by removing them to realize the image enhancement. Single scale retinex (SSR) [3], multiscale retinex (MSR) [4] and multiscale retinex with color restoration (MSRCR) [5] are the typical works of this kind of method. Recently, some new methods (SRIE [6], LIME [7]) are proposed to estimate illumination maps and reflectance to enhance low-light images. These methods may suffer severe color distortion in many cases.

Dehaze-model-based methods invert low-light images and apply dehaze method on them. In [8], this method is used to enhance low-light images. However, the images are usually over-enhanced and the saturation of enhanced images is usually exaggerated.

B. Related Deep Learning Methods for Image Processing

Driven by large datasets and the improvement of calculation capability, deep learning-based methods have shown great success in low-level image processing applications.

For super resolution, VDSR [9] utilizes VGG filters and uses twenty convolutional layers to get impressive results. When referring to image denoising, DnCNN [10] uses similar network to VDSR and adds batch normalization layers after convolutional layers, which achieves higher PSNR than traditional image denoising methods.

As far as we know, LLNet [11] is the only method using deep neural networks to enhance low-light images. The network is a variant of the stacked-sparse denoising autoencoder, and it does not use convolutional layers. Natural images are darkened using nonlinear method to simulate low light conditions. These images are set as training data. After training, the network can enhance low-light images.

* The corresponding author. E-mail: {czhu, donxie}@pku.edu.cn

This work is partially supported by the National Science Foundation of China (61421062, 61502013 and 61602011), the Major National Scientific Instrument and Equipment Development Project of China under contract No. 2013YQ030967, National Key Research and Development Program of China (No. 2016YFB0401904, No. 2016YFB0402001).

978-1-5386-0462-5/17/$31.00 ©2017 IEEE VCIP 2017, Dec. 10 – 13, 2017, St Petersburg, U.S.A.

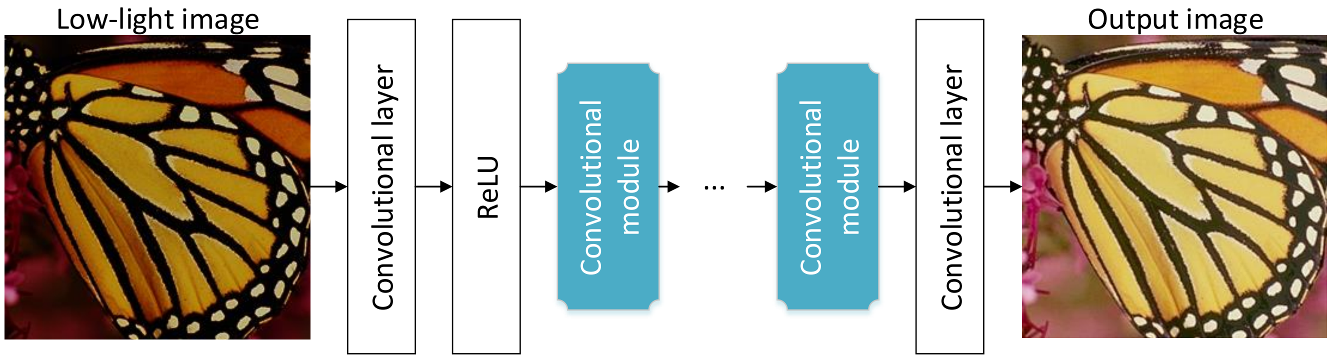

Fig. 1. The architecture of LLCNN. One convolutional layer is used to generate uniform input data of our convolutional module and another is used to generate output image. The number of convolutional module is flexible.

C. Inception Modules and Residual Learning

In many computer vision tasks, deeper networks have better performance than those non-deep ones. However, when stacked convolutional layers go deeper, the network will meet a serious problem of vanishing gradients, which will hamper convergence during training.

In inception modules [12], several layers of different sizes are connected to a previous layer, and the output data of these layers are concatenated into a single output vector. By using inception modules, GoogLeNet achieved the state-of-the-art for classification and detection in ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC14).

Residual learning [13] is also proposed to combat this problem. Shortcut connections are used to make deep residual nets easier to optimize and enjoy accuracy gains from the increased depth. Residual nets with 152 layers can be trained easily and won the first places on several tasks in ILSVRC15.

III. PROPOSED METHOD

Though low-light image enhancement belongs to low-level image processing tasks, it differs a lot from super resolution and image denoising. For these two tasks, pixel values in degraded images are around the true values, and the average pixel values almost doesn't change, which is different in our task. Therefore, we design a different but effective CNN architecture to enhance low-light images.

A. Network Architecture

The network architecture is shown in Fig. 1.We design a special convolutional module to form our network. Since deeper networks usually yield better performance, we tend to design deeper networks. Thus, we have to cope with vanishing gradient problem. The design of our convolutional module is inspired by inception module and residual learning.

The module is illustrated in Fig. 2 and can be divided into two stages. In the first stage, data are processed in two separate ways. One way is a 1×1 convolutional layer and the other way is two 3×3 convolutional layers. We combine them together to form the input of the second stage. The first stage is similar to the inception module. We do not concatenate the results but add them instead. In the second stage, there are also two ways. The first way is to process data using two 3×3 convolutional layers while in the second way, the input data is bypassed directly, which is the shortcut connection used in residual learning.

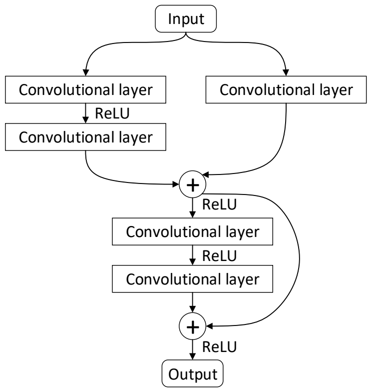

Fig. 2. The special-designed convolutional module used in our model.

As to VDSR and DnCNN, the network will generate a residual image and the final image is calculated by adding the residual image to the original one. This works because for super resolution and image denoising, the difference between the ground truth and the degraded image is not very big. For low-light image enhancement, it seems that it will be more difficult for a network to learn a residual image than enhance the image little by little. Therefore, we do not use this kind of architecture in VDSR or DnCNN. Another reason is that we have already utilized residual learning in our module.

The architecture of LLCNN is described as follows: one convolutional layer is used to do pre-processing to produce uniform input, and another convolutional layer is used to generate enhanced image, several special-designed convolutional modules are placed between those two layers. For each filter, we use 64 filters except the last one. The number of the filters used in the last layer depend on the number of color channels.

The network takes low-light images as input and do processing to make the output image appear to be captured in normal light conditions. Similar to VDSR and DnCNN, we cut the original image into patches. The patch size is set as 41×41. All input images are generated using nonlinear method to simulate low light condition.

B. SSIM Loss

In low-light enhancement, we do not need the enhanced image to be in specific light conditions. For example, given a natural image captured in daylight, it's fine to change all pixel values by adding or subtracting a small number, in which case, the structure almost does not change while PSNR will report a big difference. Therefore, the default loss function, Euclidean loss, may not be suitable for our task. For low-light image enhancement, texture preserving is more important and the brightness is allowed to be fluctuant around the ground truth. Therefore, SSIM loss should be more suitable for our task.

The SSIM value for pixel p is calculated by the definition

$$SSIM(p) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}. \qquad (1)$$

The SSIM value is in range (0, 1]. Value 1 means that the two images are totally the same. Thus, we use $1 - SSIM(p)$ to calculate the loss for pixel p. The loss function for SSIM is defined as follows

$$\ell_{ssim} = \frac{1}{N}\sum_{p \in P} 1 - SSIM(p) \qquad (2)$$

C. Training

How to collect a set of training data for low-light image enhancement is a big problem. Unlike other image processing tasks, the ground truth is unknown. We follow [11] and set natural images as ground truth and low-light images are produced by a degradation method. Random gamma adjustment is used to simulate low-light images. The parameter gamma is set randomly in range (2, 5), which will enable the network to enhance images adaptively.

We test two kinds of different depth of our network. One is named LLCNN, which has three modules so that it has 17 convolutional layers. The other one, LLCNN-s, uses two modules and the number of convolutional layers is 12. We set 64 filters in each layer except for the last one.

When training, the momentum is set to 0.9 and weight decay is 0.0001. The base learning rate is 0.01. The learning rate will change during training. The parameters for SSIM loss layer are also provided here. The kernel size and C_1, C_2 are set to 8, 0.0001 and 0.001 respectively. In the process of our training, 40,000 iterations means 1 epoch.

IV. EXPERIMENTS

In this section, we first investigate the impact of different depths of our model. Next, we explore the effectiveness using SSIM loss compared with the Euclidean loss. At last, a CNN model which has achieved success in super resolution and image denoising will be trained as a CNN baseline for low-light image enhancement task. We compare our proposed model with other traditional contrast enhancement method, as well as deep learning-based methods such as LLNet and the CNN baseline. The CNN baseline is trained using the same model architecture as VDSR but our training data.

A. Choices for the Depth of the Network

Models of two depth are tested to evaluate the influence. We test two models in this part, LLCNN and LLCNN-s.

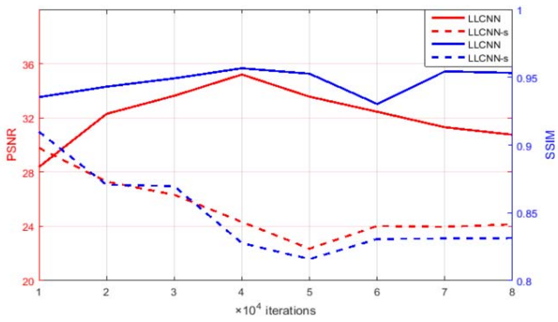

Fig. 3. Performance of LLCNN and LLCNN-s in PSNR and SSIM.

We test the networks using Set-5 and Set-14 together. The performances are shown in Fig. 3. As we can see from the figure, 17-layer network has obvious advantages over 12-layer network after about 10,000 iterations (2.5 epoch), which means that the generalization ability of the deeper network is much better. We use 17-layer network to do further study.

B. Gains from SSIM Loss Function

Two loss functions are tested in this part. Euclidean loss is the most widely used loss function in CNN-based image processing methods. SSIM loss function is used in our LLCNN.

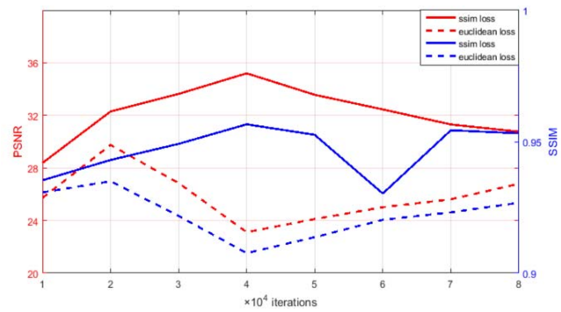

Fig. 4. Performance of different loss functions when using the same network architecture in PSNR and SSIM.

The performances are shown in the Fig. 4. The result shows that the network benefits a lot from SSIM loss and outperforms the same network using Euclidean loss.

C. Comparisons to Other Methods

We compare our LLCNN with other contrast enhancement methods for low-light images. These methods include retinex-theory-based methods, SSR [3], MSRCR [5], SRIE [6], LIME [7]; histogram equalization methods, DHE [1], CLAHE [2]; dehaze-model-based method, Li's method [8] and deep learning-based method, LLNet [11]. We also use the same network structure as VDSR [9] and train it using our training data.

TABLE 1 COMPARISON WITH LLNET

PSNR/SSIM	Test	LLNet	LLCNN
Bird	12.27/0.18	18.43/0.60	**29.95/0.87**
Girl	9.50/0.51	22.45/0.80	**36.29/0.99**
House	12.12/0.32	21.10/0.64	**29.15/0.85**
Pepper	10.45/0.37	21.33/0.78	**32.67/0.93**
Town	10.17/0.36	22.47/0.81	**33.90/0.96**

For the comparison with LLNet, we compare the results reported in [11]. Test images are degraded using gamma adjustment with $\gamma = 3$. Results are shown in the Table 1. It can be seen that LLCNN performs much better than LLNet.

For other methods, we use a combination of Set-5 and Set-14 to evaluate. In addition to PSNR and SSIM, we also use lightness-order-error (LOE) [14] and structure natural measure (SNM) [15] to assess the performance. Results are shown in Table 2 and an example is illustrated in Fig. 5. Gamma is set 3 to generate dark images.

TABLE 2 COMPARISON WITH OTHER METHOD IN SET-5 AND SET-14

	PSNR	SSIM	LOE	SNM
Dark	12.70	0.476	50.82	0.118
SSR	22.33	0.908	25.93	0.571
MSRCR	25.97	0.894	65.93	**0.612**
SRIE	14.83	0.605	56.82	0.200
LIME	19.17	0.719	92.70	0.565
DHE	13.35	0.546	58.30	0.195
CLAHE	16.54	0.686	75.73	0.346
Li's	18.63	0.754	27.98	0.381
CNN baseline	26.08	0.890	31.68	0.545
LLCNN	**35.20**	**0.957**	**10.27**	0.598

Fig. 5. An example of the comparison with other contrast enhancement method for low-light images.

Table 2 shows that our LLCNN achieves state-of-the-art performance among these contrast enhancement methods. And Fig. 5 demonstrates that by using LLCNN, the enhanced image appears more naturally than other methods. We also test our method on natural dark images and an example is shown in Fig. 6, which shows our method is effective on natural dark images.

V. CONCLUSIONS

We propose a CNN-based method to enhance low-light images. The proposed approach, LLCNN, learns to adaptively enhance image contrast and increase image brightness. In LLCNN, a special module is designed to help training and improve the performance. We also find that SSIM loss suits better for low-light image enhancement task. Experimental results demonstrate that LLCNN can achieve superior performance over other contrast enhancement methods.

Fig. 6. An example of LLCNN on natural dark images.

REFERENCES

[1] Abdullah-Al-Wadud, M, et al. "A Dynamic Histogram Equalization for Image Contrast Enhancement," *IEEE Transactions on Consumer Electronics*, vol. 53, pp. 593-600, 2007.

[2] Pisano E D, Zong S, Hemminger B M, et al. "Contrast Limited Adaptive Histogram Equalization image processing to improve the detection of simulated spiculations in dense mammograms," *Journal of Digital Imaging*, vol. 11, no. 4, pp. 193-200, Nov. 1998.

[3] D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Transactions on Image Processing*, vol. 6, no. 3, pp. 451–462, Mar. 1996.

[4] Rahman, Z., D. J. Jobson, and G. A. Woodell, "Multi-scale retinex for color image enhancement," In *Image Processing, 1996. Proceedings., International Conference*, 1996, vol. 3, pp. 1003-1006.

[5] Jobson, D. J., Z. Rahman, and G. A. Woodell. "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image processing*, vol. 6, no. 7, pp. 965-976, 1997.

[6] Fu X, Zeng D, Huang Y, et al. "A weighted variational model for simultaneous reflectance and illumination estimation," *IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 2782-2790.

[7] Guo X, Li Y, Ling H. "LIME: Low-Light Image Enhancement via Illumination Map Estimation," *IEEE Transactions on Image Processing*, vol. 26, no.2, pp. 982-993, 2017

[8] Li, Lin, et al. "A low-light image enhancement method for both denoising and contrast enlarging," *IEEE International Conference on Image Processing IEEE (ICIP)*, 2015, pp. 3730-3734.

[9] Jiwon Kim, Jung Kwon Lee and Kyoung Mu Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1646-1654.

[10] Zhang K, Zuo W, Chen Y, et al. "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising", *IEEE Transactions on Image Processing*, 2017.

[11] Lore K G, Akintayo A, Sarkar S. "LLNet: A deep autoencoder approach to natural low-light image enhancement", *Pattern Recognition*, vol. 61, pp. 650-662, 2017.

[12] Szegedy C, Liu W, Jia Y, et al. "Going deeper with convolutions," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1-9.

[13] He K, Zhang X, Ren S, et al. "Deep residual learning for image recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778.

[14] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error measurement to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 1, Jan. 2004.

[15] Yeganeh, Hojatollah, and Zhou Wang. "Objective quality assessment of tone-mapped images." *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 657-667, 2013.