GLADNet: Low-Light Enhancement Network with Global Awareness

Wenjing Wang*, Chen Wei*, Wenhan Yang, Jiaying Liu
Institute of Computer Science and Technology, Peking University

***Abstract*—In this paper, we address the problem of low-light enhancement. Our key idea is to first calculate a global illumination estimation for the low-light input, then adjust the illumination under the guidance of the estimation and supplement the details using a concatenation with the original input. Considering that, we propose a GLobal illumination-Aware and Detail-preserving Network (GLADNet). The input image is rescaled to a certain size and then put into an encoder-decoder network to generate global priori knowledge of the illumination. Based on the global prior and the original input image, a convolutional network is employed for detail reconstruction. For training GLADNet, we use a synthetic dataset generated from RAW images. Extensive experiments demonstrate the superiority of our method over other compared methods on the real low-light images captured in various conditions.**

***Keywords*-Low-light enhancement, deep learning, detail reconstruction, encoder-decoder**

I. Introduction

Insufficient illumination can severely degrade the quality of images. This is due to poor shooting environments, limited performance of photographic equipments and improper operations of photographers. It often causes insufficient saturation and contrast of images, damages visual quality and degrades the performance of many computer vision algorithms.

In the past decades, various algorithms have been proposed to improve the subjective and objective quality of low-light images. Histogram equalization (HE) [1] is a widely-used technique. By restraining the histograms of the output images to meet some constraints, HE and its variants can improve the contrast effectively. Retinex is a model of lightness and color perception of human vision. Images are assumed to be a combination of two components, reflectance and illumination. Single Scale Retinex (SSR) [2] manipulates the reflectance component and treats it as the final output. Multi-Scale Retinex with Color Restoration (MSRCR) [3] extends the single-scale enter/surround retinex to a multi-scale version. Guo *et al*. [4] tried to estimate the illumination map. Fu *et al*. [5] simultaneously estimated reflectance and illumination. De-hazing based methods [6] utilize the high similarity between low-light images and those with the dense fog. Fusion based methods [7] introduce the fusion mechanism of human visual system to help build an accurate image enhancement algorithm.

* indicates equal contributions.

978-1-5386-2335-0/18/$31.00

Deep Neural Networks has achieved excellent results on many low-level computer vision tasks. There are also efforts on utilizing deep neural networks for low light enhancement. Lore *et al*. [8] used variant of the stacked-sparse denoising auto-encoder learning from synthetically darkened and noise-added training examples for simultaneously low-light enhancement and noise reduction (LLNet).

In this paper, we propose a **GL**obal illumination-**A**ware and **D**etail-preserving **Net**work (**GLADNet**). The architecture of the proposed network can be divided into two steps. In order to obtain a global illumination prediction, the image is first down-sampled to a fixed size and passed through an encoder-decoder network, which we call the global illumination estimation step. The bottle-neck layer of the encoder-decoder has a receptive field that covers the whole image. The second step is a detail reconstruction step, which helps to supplement the details lost in the rescaling procedure. For training such a network, we synthesize a training dataset from RAW pictures captured in various conditions and use L_1 norm as the loss function. The effect of GLADNet is evaluated on real images with other state-of-the-art methods. Extensive experiments demonstrate the superiority of our method over other compared methods.

The rest of this paper is organized as follows. In Section II, the proposed framework is described in detail. In Section III, we talk about the datasets and settings for training, evaluate the performance and applications on real images. Section IV contains conclusions and discussions on further work.

II. Proposed Method

The architecture of the proposed network comprises two adjacent steps. One is for global illumination estimation and the other is for detail reconstruction.

As shown in Fig. 1, in the global illumination estimation step, inputs are down-sampled to a fixed size. Then, feature maps are passed through an encoder-decoder network. At the bottle-neck layer, the global illumination is estimated. After scaling back to the original size, an illumination prediction for the whole image is obtained. The global illumination estimation step is followed by a detail reconstruction step. Three convolutional layers adjust the illumination of the input image referring to the global-level illumination prediction, and fill in the details lost in the down-sampling and up-sampling procedure at the same time.

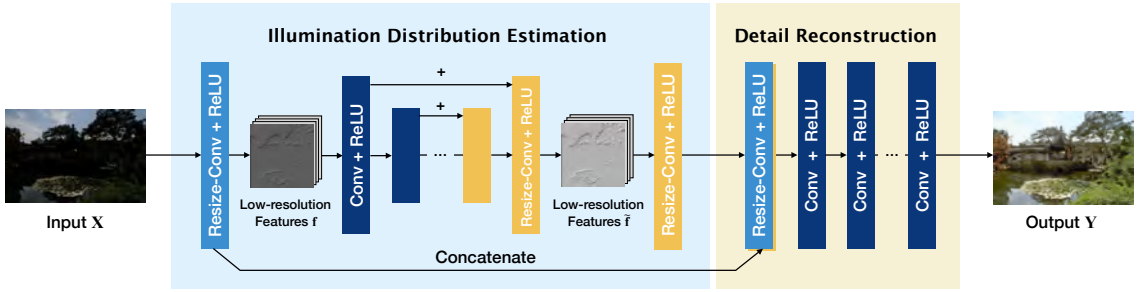

Figure 1. The architecture of GLADNet. The architecture consists of two steps, global illumination estimation step and detail reconstruction step. In the first step, the encoder-decoder network produces an illumination estimation of a fixed size (96×96 here). In the second step, a convolutional network utilizes the input image and the outputs from the previous step to compensate the details.

A. Global illumination estimation

The global illumination estimation step has three sub-steps: scaling the input image to a certain resolution, passing it through an encoder-decoder network for global illumination prediction, and rescaling it to the original resolution.

First, the input is down-sampled to a certain size $W_0 \times H_0$ by nearest-neighbor interpolation. A convolutional layer with a ReLU is followed. Then, the feature maps pass through a series of cascaded down-sampling blocks. The number of down-sampling blocks are carefully designed according to W_0 and H_0, so that the receptive field of the bottle-neck layer of the encoder-decoder network can cover the entire image. The network thus has a global awareness of the whole illumination distribution. This design can also reduce the requested storage, and increase the efficiency of the network. After a series of symmetrical up-sampling blocks, $W_0 \times H_0$ feature maps for illumination prediction is obtained. By another up-sampling block, the feature maps are rescaled to the size of the original input.

Skip connections are introduced from a down-sampling block to its corresponding mirrored up-sampling block. Outputs of the down-sampling block are passed to and summed up with the feature maps of the up-sampling block. This enforces the network to learn residuals rather than predicting the actual pixel values.

A down-sampling block consists of a convolutional layer with stride two and a ReLU. In the up-sampling block, resize-convolutional layers [9] are used to replace normal deconvolution layers. Different from normal deconvolution layers, resize-convolutional layers avoid checker-board pattern of artifacts and have no limit of the size of the input image. Resize-convolutional layer consists of a nearest-neighbor interpolation operation, a convolutional layer with stride two and a ReLU.

B. Details reconstruction

The first step is to produce an illumination estimation from a global perspective. However, details are lost due to the rescaling procedure. In order to address this issue, a detail reconstruction procedure is proposed.

The original input is considered to contain more details than the output of the encoder-decoder network, therefore can provide information for detail restoration. Concatenation is used instead of skip-connection to combine the feature maps of the last up-sampling block and the input image, so that both the original information and the illumination estimation can be completely preserved and transmitted to the next step. The concatenation layer is followed by three convolutional layers with ReLUs. It assemble the input images information with the estimated global illumination information and finally generate enhanced results with better details.

C. Loss function

The training procedure is achieved by minimizing the loss between the restored image $F(X, \Theta)$ and the the corresponding ground-truth image Y. We use L_1 norm here. L_2 norm to better remove noise and ringing artifacts in the enhanced results [10]. The loss function can be written as:

$$Loss(X, Y) = \frac{1}{N} \sum_{i=1}^{N} \|F(X_i, \Theta) - Y_i\|_1, \tag{1}$$

where N is the number of all training samples and $\|\cdot\|_1$ is L_1 norm.

Also the red, the green and the blue channel have their own weights in the loss function: (0.29891, 0.58661, 0.11448), which is the same weight for the transformation from RGB images to gray images. This helps to maintain color balance and improve the robustness of the network.

III. EXPERIMENTS

A. Dataset generation

We use synthesized pairs as training data. Different from [8][11] which synthesize pairs on 8-bit RGB images, we synthesize pairs on raw images. Calculating on 8-bit RGB images can cause loss of information with only 256 values. On raw images, all adjustments are performed in one step on the raw data, leading to more accurate results.

We collect 780 raw images from RAISE [12], 700 for generating pairs for training and 80 for validation. Adobe Photoshop Lightroom offers a series of parameters for raw

Figure 2. The results enhanced by different methods on natural images: (top-to-bottom) "House" from NPE dataset, "Chinese Garden" from MEF dataset, and "Street" from DICM dataset.

image adjustment, including exposure, vibrance and contrast. We synthesize low-light images by setting exposure parameter E to $[-5, 0]$, vibrance parameter V to $[-100, 0]$, contrast parameter C to $[-100, 0]$. In order to prevent color-bias, we add to the training dataset 700 gray-scale image pairs which are converted to color image pairs. To keep the black and the white regions the same before and after the enhancement, we add five black-to-black and five white-to-white training pairs. Finally, all images are resized to 400×600 and converted to Portable Network Graphics format.

B. Network training

The kernel size of GLADNet is set to 3×3. W_0 and H_0 are both set to 96. The encoder-decoder architecture has five down-sampling and five corresponding up-sampling blocks. This design makes the size of the bottle-neck layer in the encoder-decoder network 3×3, so that the receptive field can cover the whole image.

We initialize the weights using the initialization proposed by [13]. Adam is used as the optimization method and each mini-batch contains 8 image pairs. We start with learning rate of $1e-3$, multiply 0.96 after each 100 batch. The model is trained for 50 epochs.

C. Subjective evaluation

Although GLADNet is trained on synthetic data, we evaluate its performance on real under-exposed images. We compare GLADNet with several state-of-the-art methods, including Multi-Scale Retinex with Color Restoration (MSRCR) [3] , dehazing based method (DeHz) [6], Illumination Estimation based method (LIME) [4] and Simultaneous Reflection and Illumination Estimation (SRIE) [5]. Here we show several subjective results. Quantitative results on quality of enhanced images and the analysis of runtime can be found in our project website[1].

We evaluate our approach on real under-exposed images from public LIME-data [4], DICM [14], and MEF [15]

[1]https://daooshee.github.io/fgworkshop18Gladnet/

Figure 3. Results of Google Cloud Vision API for "Eiffel Tower" from MEF dataset. Before enhancement, Google Cloud Vision can not recognize the Eiffel Tower. After enhanced by GLADNet, the Eiffel Tower is identified and marked by a green box.

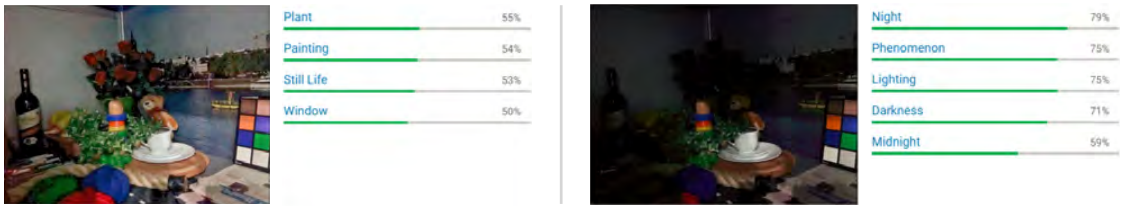

Figure 4. Results for "Room" from LIME-data dataset. Potted plant and painting in the non-enhanced version are not recognized by Google Cloud Vision.

datasets. LIME-data contains 10 low-light images used in [4]. DICM dataset is collected by [14] including 69 captured images with commercial digital cameras. MEF contains 17 image sequences with multiple exposure levels. We select 76 poor-exposed ones for evaluation.

Fig. 2 shows three sets of visual comparisons in the test dataset. As we can see, MSRCR can fully illuminate the images, but the results are kind of whitish. The results of LIME are visually pleasant, but bright regions are over enhanced and details are lost. For example, in "House" and "Chinese Garden", the sky regions behind the trees are over-exposed. The results of DeHZ have artifacts on edges, which reduces the visual aesthetics of enhanced results. On the other hand, SRIE does not sufficiently improve the brightness of low-light images and the details can not be seen clearly. In "Chinese Garden", tiles above the pavilion are still not visible.

Compared with other methods, our method produces more vivid and natural results. Since GLADNet has a global awareness of the input and adjusts the whole image at the same time, over-exposure in brighter regions and under-exposure in darker regions can be avoided. Further more, the details are still kept after enhancement, which is benefited from the detail reconstruction step.

D. Applications on Computer Vision

One of GLADNet's main applications is to help improve the performance of other computer vision tasks, such as object detection and recognition. Since most visual recognition models are based on high-quality data, poor conditions such as poor visibility, haze, and low-illumination can greatly reduce the performance of these algorithms.

To illustrate the effectiveness of our method for improving the performance of object recognition, we test several real low-light images and their corresponding enhanced results on Google Cloud Vision API[2], which can understand the content of images via machine learning models and dividing them into thousands of categories.

Fig. 3 shows one of the paired results. The original image is from MEF dataset. Due to the low-illumination, Google Cloud Vision can only label the image as "sky", "cloud" and "spire". After enhanced, the foreground Eiffel Tower is successfully detected and marked by a green box precisely, showing the effectiveness of our method.

Another example is from LIME dataset. Before enhancement, labels are settled on "night" and "phenomenon". The potted plants, paintings and other items are too dark to be detected. As shown in Fig. 4, GLADNet helps Google Cloud Vision API identify the objects in this image.

IV. CONCLUSIONS

In this paper, a global illumination-aware and detail-preserving network is proposed. The proposed architecture consists of two steps. First, an encoder-decoder network obtains an illumination prediction of a fixed size from a global perspective. Then, a convolutional network reconstructs details utilizing the illumination prediction and the original input. Results show that our method outperforms other state-of-the-art methods.

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China under contract No. 61772043. We also gratefully acknowledge the support of NVIDIA Corporation with the GPU for this research.

[2]https://cloud.google.com/vision/

REFERENCES

[1] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. T. H. Romeny, and J. B. Zimmerman, "Adaptive histogram equalization and its variations," *Computer Vision Graphics & Image Processing*, vol. 39, no. 3, pp. 355–368, 1987.

[2] 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. 451–62, 1997.

[3] Z. Jobson, Daniel Jand 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. 965–76, 1997.

[4] X. Guo, Y. Li, and H. Ling, "Lime: Low-light image enhancement via illumination map estimation," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 982–993, 2017.

[5] X. Fu, D. Zeng, Y. Huang, X. P. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," *Computer Vision and Pattern Recognition*, pp. 2782–2790, 2016.

[6] X. Dong, Y. Pang, and J. Wen, "Fast efficient algorithm for enhancement of low lighting video," *IEEE International Conference on Multimedia and Expo*, pp. 1–6, 2011.

[7] Z. Ying, G. Li, and W. Gao, "A bio-inspired multi-exposure fusion framework for low-light image enhancement," *Computing Research Repository*, vol. abs/1711.00591, 2017.

[8] K. G. Lore, A. Akintayo, and S. Sarkar, "Llnet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognition*, vol. 61, pp. 650–662, 2016.

[9] A. Odena, V. Dumoulin, and C. Olah, "Deconvolution and checkerboard artifacts," *Distill*, 2016. [Online]. Available: http://distill.pub/2016/deconv-checkerboard/

[10] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for neural networks for image processing," *Computer Science*, 2016.

[11] L. Shen, Z. Yue, F. Feng, Q. Chen, S. Liu, and J. Ma, "Msr-net: Low-light image enhancement using deep convolutional network," *Computing Research Repository*, vol. abs/1711.02488, 2017.

[12] D. T. Dang-Nguyen, C. Pasquini, V. Conotter, and G. Boato, "Raise: a raw images dataset for digital image forensics," *ACM Multimedia Systems Conference*, pp. 219–224, 2015.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," *Computing Research Repository*, vol. abs/1502.01852, 2015.

[14] C. Lee, C. Lee, and C. S. Kim, "Contrast enhancement based on layered difference representation," *IEEE International Conference on Image Processing*, pp. 965–968, 2013.

[15] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Transactions on Image Processing*, vol. 24, no. 11, p. 3345, 2015.