Self-Supervised Image Enhancement Network: Training with Only Low-Light Images

Authors: Zhang Yu, Ti Xiaoguang, Zhang Bin, Wang Chunhui, Ti Xiaoguang

Date: 2020-03-06T00:00:00+00:00

Abstract

This paper proposes a self-supervised low-light image enhancement method based on deep learning. Inspired by information entropy theory and the Retinex model, we propose a Retinex model based on maximum information entropy. Utilizing this model, a very simple network can separate the illumination map and reflectance map, and can be trained using only low-light images. To achieve self-supervised learning, we introduce a constraint in the model: the maximum channel of the reflectance map is consistent with the maximum channel of the low-light image, and its entropy is maximized. Our model is very simple, does not rely on any elaborately designed datasets (even a single low-light image is sufficient for training the network), and the network requires only minutes of training to achieve image enhancement. Experiments demonstrate that the method achieves state-of-the-art performance in both processing speed and quality.

Full Text

Preamble

Self-supervised Image Enhancement Network: Training with Low Light Images Only

Yu Zhang, Xiaoguang Di, Bin Zhang, and Chunhui Wang

Abstract– This paper proposes a self-supervised low-light image enhancement method based on deep learning. Inspired by information entropy theory and the Retinex model, we introduce a maximum entropy-based Retinex model. With this model, a very simple network can separate illumination and reflectance, and the network can be trained using only low-light images. We introduce a constraint that the maximum channel of the reflectance should conform to the maximum channel of the low-light image while achieving maximum entropy in our model to enable self-supervised learning. Our model is very simple and

does not rely on any carefully designed dataset (even a single low-light image can complete the training). The network only requires minute-level training to achieve image enhancement. Experiments demonstrate that the proposed method achieves state-of-the-art performance in terms of both processing speed and enhancement quality.

Index Terms– Low Light Image Enhancement, Self-supervised Learning, Max Entropy, Retinex.

1 Introduction

Recently, various deep learning-based algorithms have achieved remarkable results in numerous image processing and computer vision tasks, such as object detection [1], [2], [3], [4], image segmentation [4], [5], [6], etc. One important reason for the rapid development of deep learning in these tasks is the availability of large datasets with clear and unambiguous labels. Although dataset construction incurs some cost, it remains acceptable, and numerous open-source datasets can be found on the Internet to support network training for these tasks. However, in low-level image processing tasks such as low-light image enhancement, image dehazing, and image restoration, it is difficult to obtain large numbers of true input/label image pairs.

For the low-light image enhancement task, previous works have proposed solutions such as synthesizing low-light images [7] and using images captured at different exposure times [8], achieving good visual effects. However, these methods still face two fundamental problems. First, how can we ensure that a pre-trained network can be applied to images collected from different devices, scenes, and lighting conditions without building a new training dataset? Second, how can we determine whether the normal-light image used for supervision is optimal, as numerous normal-light images may correspond to a single low-light image?

Typically, dataset builders obtain normal-light images through experience or manual adjustment, which consumes substantial time and energy, and we cannot guarantee that the enhanced image reveals the information contained in the low-light image to the greatest extent using those normal-light images.

To address these two questions, this paper proposes a self-supervised low-light image enhancement network based on information entropy theory and the Retinex model, achieving state-of-the-art performance in terms of enhancement quality and efficiency. In this work, the only data we require are low-light images, without any paired or unpaired normal-light images. To our knowledge, this is the first fully self-supervised image enhancement method based on deep learning. The proposed method does not rely on a carefully designed complex network structure; with only a simple fully convolutional neural network (CNN) as shown in Fig. 2 and minute-level training, we can complete low-light image enhancement tasks.

There exist some image enhancement networks based on the Retinex model [9], [10], but they all require paired data and utilize the assumptions that images captured under different lighting conditions should have the same reflectance and that the illumination map should be smooth to decompose low-light images into corresponding reflectance and illumination maps. Similar to these works, we also use a network to decompose low-light images into reflectance and illumination. However, unlike previous works, we employ self-supervised methods to train the network. Only low-light images are required for training (even a single low-light image), enabling us to obtain reflectance with good visual effects that can be treated as an enhanced image.

We believe that the low-light image enhancement task aims to display the information contained in low-light images more intuitively rather than creating new information. Simultaneously, according to entropy theory, images with uniform histogram distributions have maximum entropy and contain the most information. Based on this analysis, we propose the assumption that the histogram distribution of the maximum channel of the enhanced image should conform to the histogram distribution of the maximum channel of the low-light image after histogram equalization. With this assumption, the loss function can be designed without normal-light images, which not only retains the authenticity of the enhanced image but also ensures that the enhanced image contains sufficient information.

The proposed method has no dependence on how low-light images are acquired, and the training process is completely self-supervised. Consequently, our method exhibits good generalization capability; even if the pre-trained network performs poorly in a new environment, retraining or fine-tuning is possible without building paired/unpaired normal-light image datasets. Our contributions include:

- We propose a new maximum entropy-based Retinex model and provide its theoretical foundation.
- Combined with deep learning, we propose a self-supervised low-light image enhancement network that can complete training with even a single low-light image.
- The proposed method only requires minute-level training and achieves good real-time performance. We verify the enhancement effect and stability of the algorithm through experiments and objective metrics.

2 Related Works

Our method primarily draws inspiration from histogram equalization, model-based methods, and deep learning-based image enhancement methods.

Histogram Equalization. In low-light image enhancement tasks, Histogram Equalization (HE) is the simplest and most widely used method. It can make the histogram of the enhanced image follow a uniform distribution to achieve maximum entropy. However, HE cannot avoid problems such as detail disappearance (over-enhancement, under-enhancement), poor color restoration, and noise amplification.

To address these issues, various improved algorithms have been proposed, such as Adaptive Histogram Equalization (AHE) [11] and Contrast-limited Adaptive Histogram Equalization (CLAHE) [12] for preserving details, Hue-preserving color image enhancement [13] for hue preservation, Brightness Bi-Histogram Equalization Method (BBHE) [14], Dualistic Sub-Image Histogram Equalization Method (DSIHE) [15] for brightness preservation, etc. References [16] and [17] propose methods considering the relationship between adjacent pixels and large gray-level differences. Although many improved methods have been proposed, numerous problems remain when applying histogram equalization directly to image enhancement.

Model-Based Image Enhancement Methods. Among model-based low-light image enhancement methods, the primary approaches are based on the dehazing model [18] and the Retinex model [19]. Methods based on the dehazing model rely on the observation that low-light images resemble haze images after inversion.

Dong et al. proposed an enhancement method that performs dehazing after inverting the low-light image and then inverts the image back [18]. Some studies have extended these works [20]. Although these methods have achieved good effects, they lack corresponding physical interpretation, which limits their application across various scenes.

According to Retinex theory, captured images can be decomposed into illumination and reflectance, but this is a highly ill-posed problem when attempting to obtain them from low-light images alone. Therefore, additional constraints must be introduced. Early research such as the single-scale Retinex [21] model and the multi-scale Retinex [22] model only use the constraint that the illumination map should be smooth to solve the problem. The captured image is smoothed using Gaussian filters with one or multiple scales to obtain illumination; however, the enhanced images often exhibit unrealistic phenomena such as over-enhancement and whitening. Reference [23] proposes a Bright-Pass Filter that preserves natural characteristics. Reference [24] performs global illumination adjustment and local contrast enhancement on the initially estimated illumination map.

Although these methods have achieved some good effects, without considering structural characteristics, information loss is prone to occur in areas rich with details. LIME [25] introduces a filter considering structural characteristics to smooth the illumination map and uses BM3D to denoise the enhanced image. Although these methods are proposed to maintain image details and naturalness, the denoising process will still cause blur or loss of details in both pre- and post-enhancement tasks.

It is difficult to add additional priors to those methods based solely on the Retinex model, leading to problems such as noise, halos, and detail preservation. Therefore, in recent years, many algorithms based on the variational

Retinex model have been proposed. Kimmel et al. [26] first proposed a variational Retinex model and used L2 regularization to obtain a smooth illumination map. Fu et al. [27] introduced the bright channel prior to the variational Retinex model to suppress the halo effect. Park et al. [28] proposed weighted L2 regularization to constrain the reflectance image, which has a slight noise suppression effect. Fu et al. [29] proposed an L2-Lp norm to constrain the illumination map and preserve more details. Although these variation-based methods have achieved good results, they are very time-consuming due to the need for multiple iterations to solve the variational equation. Even with Fast Fourier Transform (FFT), it is difficult to ensure real-time performance.

In addition, in those model-based low-light image enhancement algorithms, spatial smoothing prior [30] and its improvements are mostly used to constrain the illumination map, with no constraint on the contrast information of the reflectance. In this paper, we use maximum information entropy to constrain the reflectance image, thereby further improving its contrast information.

Learning-Based Methods. Learning-based methods have achieved good results in some low-level image processing tasks, such as image denoising [31], [32], super-resolution reconstruction [33], [34], restoration [35], [36], etc. However, most current deep learning-based algorithms are supervised, and it is difficult to obtain both degraded and normal images for those low-level image processing tasks. Some research has proposed synthesizing low-light image data with normal-light images for training. For example, LLNET [37] is the first work to use deep learning to solve image enhancement problems; it proposes training networks separately with synthetically noisy and dark images but does not consider natural image characteristics. References [38] and [39] apply gamma transformation to natural image patches to generate low-light image patches for training but do not consider other degradations of real collected low-light images such as noise and color changes. In MSR-net [40], high-quality (HQ) data are obtained through artificial selection and Photoshop, and low-light images are obtained by processing HQ data with random brightness and contrast reduction and gamma transformation. Although data obtained by these methods appear similar to low-light images, they fail to truly reflect the characteristics of low-light images, such as noise, overexposed and underexposed areas existing simultaneously in the same image.

To address this problem, some methods propose using real low-light images for training. Reference [41] establishes a large multi-exposure image database, and reference images are obtained by combining different exposure images with subjective selection. Retinex-net [9] attempts to obtain low/normal-light image pairs by adjusting exposure time and achieves good enhancement effects, but the exposure time is still artificially determined, making it difficult to choose the best exposure time to obtain a reference image. Reference [8] introduces a parameter to link two images with different exposure times and, with end-to-end training, can effectively handle the noise problem, but it can only be used for raw images. Reference [42] introduces a light adjustment network to link paired images with any different exposure time, solving the problem of acquiring normal-light images. However, in practical applications, if we want to obtain better images, we may need to choose a hyper-parameter for each low-light image.

Although these deep learning-based methods have achieved good visual effects in low-light image enhancement, they all rely on paired images, and the cost of building training data is high. Moreover, they do not solve the two problems we mentioned earlier: how to obtain an optimal reference image and how to ensure the method' s adaptability to new environments or new equipment.

3.1 Maximum Entropy Based Retinex Model

Based on the Retinex model, an image can be decomposed into reflectance and illumination map as follows:

$$S = R \circ I$$

where S represents the captured image, R represents the reflectance, and I represents the illumination map. This is a highly ill-posed problem whose solution requires additional priors. According to Bayes' formula, the problem can be expressed as:

$$p(R, I|S) \propto p(S|R, I)p(R)p(I)$$

where $p(R, I|S)$ is the posterior probability, $p(S|R, I)$ is the class conditional probability, and $p(R)$ and $p(I)$ are the prior probabilities of reflectance and illumination. Existing methods generally add prior probabilities $p(R)$ and $p(I)$ to find the maximum posterior probability and estimate the reflectance and illumination.

By calculating the negative logarithm of equation (2), the image enhancement problem can be transformed into the form of three distance terms, as shown in formula (3):

$$\mathcal{L}_{rcon} + \lambda_1 \mathcal{L}_R + \lambda_2 \mathcal{L}_I$$

where $\mathcal{L}_{rcon}$ represents reconstruction loss, $\mathcal{L}_R$ represents reflectance loss, and $\mathcal{L}_I$ represents illumination loss. λ_1 and λ_2 are weight parameters.

In this paper, we use the L1 norm to constrain all losses. We do not compare the impact of L1, L2, SSIM, and other loss functions on low-level image processing tasks, as there are related studies such as [43]. The reconstruction loss $\mathcal{L}_{rcon}$ can be expressed as:

$$\mathcal{L}_{rcon} = \|S - R \circ I\|_1$$

Regarding the reflectance loss, different from existing methods that only use $\|\nabla R\|_1$ [28], [44], we propose a new distance measurement method for reflectance loss based on the following considerations:

- For image enhancement tasks, the processed image should contain sufficient information.
- The processed image should conform to the original image information.
- Histogram equalization can greatly improve the information entropy of an image.

Based on these considerations, we propose equation (5) as the loss for the reflectance image, which also uses L1 loss:

$$\mathcal{L}_R = \left\| \max_{c\in\{R,G,B\}} R_c - \mathcal{F}\left(\max_{c\in\{R,G,B\}} S_c\right)\right\|_1 + \lambda\|\nabla R\|_1$$

where $\mathcal{F}(X)$ denotes the histogram equalization operator applied to image X. λ is a weight parameter. This loss function means that the maximum channel of the reflectance should conform to the maximum channel of the low-light image and have maximum entropy. We choose the maximum channel for constraint for three main reasons. First, for a low-light image, the maximum channel has the greatest impact on its visual effect. Second, if other channels are selected, saturation will undoubtedly occur according to the prior that the maximum channel must be greater than the other two channels. Third, if we choose one of the color channels, such as R, G, or B channel, it obviously does not align with natural image characteristics.

For the illumination loss, we adopt the structure-aware smoothness loss proposed in [9]:

$$\mathcal{L}_I = \|\nabla I \circ \exp(-\lambda_3 \nabla R)\|_1$$

Reference [9] proposes that equation (6) can make the illumination loss aware of the image structure. This loss means that the original TV function $\|\nabla I\|_1$ is weighted with the gradient of reflectance.

From equations (3) to (6), we obtain the maximum entropy-based Retinex model, as shown in equation (7):

$$\mathcal{Z} = \|S - R\circ I\|_1 + \lambda_1 \left\| \max_{c\in\{R,G,B\}} R_c - \mathcal{F}\left(\max_{c\in\{R,G,B\}} S_c\right)\right\|_1 + \lambda_2\|\nabla I \circ \exp(-\lambda_3\nabla R)\|_1 + \lambda_4\|\nabla R\|_1$$

Variational methods or FFT are generally used to solve equation (7) with L2 loss; however, they both require multiple iterations, which introduces time consumption problems, and with more constraints, the solution becomes more complicated. To enhance images in real time, we propose a solution based on deep learning. The network uses equation (7) as the loss function. We can see that in equation (7), only low-light images are present, so the network can be trained in a self-supervised manner.

The values of λ_1, λ_2, λ_3, λ_4 are 0.1, 0.1, 10, and 0.01 in this paper. The influence of λ_1 and λ_2 values is not obvious in visual effect, so we simply choose 0.1. The value of λ_3 comes from [9]. Regarding λ_4, in our experiments, we found that it can be used to control noise. When its value increases, noise decreases, but simultaneously, the image becomes more blurry. Through experiments, we choose 0.01 for λ_4; if $\lambda_4 = 0.1$, the enhanced image exhibits obvious blur.

3.2 Self-supervised Network Based Solution

If we use variational methods or FFT to solve the model proposed from equations (3) to (7), it means we need to perform the same iterative processing for each low-light image. This not only introduces time-consuming problems, but the number of iterations for each low-light image may also be uncertain, which is almost disastrous in many real applications. At the same time, this type of solution cannot take advantage of big data, as previous data processing provides no benefit to new data processing.

In previous deep learning-based research, due to the lack of models that support self-supervised training, only paired or unpaired low/normal-light images collected in advance could be used to complete network training. However, data collected in advance cannot contain all real low-light situations, such as different environments, devices, or degradation problems, which also limits the application scope of the pre-trained network. After all, it is impossible to build the dataset when we are using them.

However, based on the model proposed from equations (3) to (7), we can achieve self-supervised training, which means we can build the dataset online and avoid the applicability problem. Compared with supervised learning whose supervisor is selected through artificial methods, the model based on maximum entropy can ensure that the enhanced image has sufficient information entropy.

We only need a very simple CNN structure to achieve the decomposition of illumination and reflectance. The specific structure of the CNN we finally adopted is shown in Fig. 2. The network input is the low-light image and its maximum channel; after several convolution and concatenation layers, reflectance and illumination can be obtained with a sigmoid layer. Table 1 provides the specific information for each layer of the network.

In fact, we have experimented with different network structures, and the stacking of convolutional layers with a sigmoid layer can also produce acceptable results. However, if we add some concatenation layers, the enhancement results become clearer. As can be seen, we use down-sampling and up-sampling in the network; their primary function is to reduce noise. In some experiments, we found that adding the down-sampling layer makes the image blurry; however, it also reduces noise.

4 Experiment

We use the LOL database [9], which contains 500 low/normal-light image pairs, with 485 used for training and images sized at 400 × 600. Note that during training, we only use natural low-light images and do not use synthetic data or normal-light images. During training, our batch size is set to 16 and the patch size is set to 48 × 48. We use Adam stochastic optimization [45] to train the network, and the learning rate is set to 0.001.

Network training and testing are completed on an Nvidia GTX 2080Ti GPU and Intel Core i9-9900K CPU, and the code is based on the TensorFlow framework.

In Section 4.1, we introduce some objective evaluation metrics. In Section 4.2, we measure the influence of training epochs on loss and evaluation metrics. In Section 4.3, we measure the stability of the algorithm through repeated experiments. In Section 4.4, we compare our algorithm with some existing methods. In Section 4.5, we present some enhancement results when the network is trained with a single low-light image.

4.1 Evaluation Metrics

Many metrics, both reference-based and reference-free, can be used to evaluate enhanced image quality. However, the constraint we use in this paper does not conform to natural image characteristics, making it difficult to evaluate the enhanced image accurately with existing evaluation metrics. In this paper, we use gray entropy (GE), color entropy (CE, where color entropy is the sum of entropy across R, G, B channels), gray mean illumination (GMI), gray mean gradient (GMG), LOE [23], NIQE [46], PSNR, and SSIM to evaluate the enhanced image. It should be noted that these metrics can only reflect image quality in certain aspects, which are not completely consistent with evaluation results from the human visual system.

The LOE_low and LOE_high are calculated with low-light and high-light images, respectively.

4.2 The Influence of Training Epochs

We use 485 low-light images from the LOL dataset for training and 15 for testing. Considering that our method is self-supervised and lacks an absolute reference, and some parameters and constraints in our loss function come from individual experience, we cannot determine whether training has reached optimal performance through loss changes alone. Therefore, we train the network for 1,000 epochs, process the testing data every 20 training epochs, and use these metrics to evaluate the network’ s training results.

Fig. 3 and Fig. 4 show the changes in loss and metrics with increasing training epochs. It can be seen that the loss falls rapidly at the beginning. On our GPU, training one epoch takes less than 0.65 seconds. Fig. 5 shows enhancement results of low-light images in the testing data with different training epochs. We only display results from the first 200 epochs. It can be observed that as training progresses, metrics that reflect image clarity, such as entropy and gradient, increase; however, the gap between enhanced images and reference images also grows. This is caused by noise—although the image becomes clearer as training progresses, noise also keeps increasing. To maintain a balance between clarity and noise, we stop training after 200 epochs. In our experiments, if training epochs exceed approximately 1,000, artifacts appear in some testing images, similar to [8]. Early stopping is a reasonable method to avoid noise and artifacts.

4.3 Repeated Learning Stability

Due to the characteristics of learning-based methods, in most cases, we cannot reproduce optimal results. Therefore, we repeat the experiments many times to evaluate the method' s repeatability. In each experiment, we train the network for 200 epochs and evaluate the network on testing data using the metrics mentioned in Section 4.1.

Fig. 6 and Fig. 7 show the evaluation metrics and some enhancement results across different experiments, respectively. It can be seen that there are large fluctuations in some metrics, such as LOE, GMG, and NIQE; however, the changes in enhancement results are not obvious in most experiments. In the fifth experiment, the colors of enhanced images are lighter than in others. We believe that differences in enhancement results may come from the L1 loss functions and differences among training data in each experiment.

In each experiment, the only difference is the training patch, which seems to impact training results. Those training patches are randomly selected and cropped, and considering the training epochs and the large size difference between images and patches, they represent only a small portion of training images. At the same time, we use L1 loss for training. Compared with L2 loss, L1 loss may have multiple solutions, and its solutions are highly affected by training data. When training data changes, results may also change significantly. However, from a visual perspective, the method proposed in this paper is relatively stable.

4.4 Comparisons with Existing Algorithms

We compare our algorithm with several existing classic and state-of-the-art methods, including HE, MSR [22], LIME [25], MF [24], NPE [23], SRIE [47], Gladnet [48], Retinex-Net [9], and L2-Lp [29].

The 15 images from the LOL dataset are used for testing to obtain objective metrics. Fig. 8 and Fig. 9 show enhancement results by different methods, and Table 2 displays the objective metrics and time consumption of those methods.

The low-light image in Fig. 10 is from the LIME [25] dataset. All results of our method come from a randomly selected experiment.

SSIM is generally used to measure structural similarity between two images. NIQE is a non-reference image quality evaluation method. These two metrics show that our method maintains good structural similarity and image quality after processing.

In CE, GE, and GMG, we can see that our method is lower than some methods. Although larger values for these metrics indicate more abundant information and clearer images, we must also consider that these metrics are highly affected by noise. It can be seen that compared with most methods, our method is closer to the reference image in these metrics.

In LOE_low and LOE_high, our method does not perform well. This is probably because overexposed and underexposed areas may coexist in low-light images or reference images, and the model proposed in this paper can avoid this problem to a certain extent, leading to poor performance in these two metrics. As shown in the lower left corner of Fig. 7, there are overexposed areas in the reference image. If we use such a reference image for training, we cannot promise that training results will not be overexposed. This also explains why it is difficult to obtain the optimal reference image by adjusting exposure time.

In PSNR, although our method is lower than Gladnet [48], we must note that our method does not consider the reference image during training, making it difficult to ensure brightness similarity between the enhanced image and the reference image, which impacts PSNR.

Although our method cannot achieve the best results in all metrics, it achieves state-of-the-art performance in terms of visual effect and some important metrics. Compared with the HE method, our method has a slight denoising effect and can better maintain structural and color information. As shown in Table 2, the PSNR and SSIM of our method are higher than those of the HE method. The HE method is not suitable for heavy noise environments, and as seen in Fig. 8 and Fig. 9, dark areas remain in images after directly using the HE method, which is caused by the theory of histogram equalization itself. Compared with model-based methods, our method costs less running time. As shown in Table 2, our method is more than 6× faster than MSR [22], which has the shortest time among model-based methods.

Compared with learning-based methods, our method does not require careful dataset construction, saving substantial time and energy and providing better applicability for new environments and equipment.

4.5 Training with Single Low-Light Image

To further evaluate the performance of the method proposed in this paper, we conduct an experiment training the network with a single low-light image. The training image is one of the test images from the LOL dataset, and the test data consists of all 15 test images from the LOL dataset. In Fig. 12, we use only image 12-(a) to train the network, and 12-(b) to 12-(k) show enhancement results of image 12-(a) at different training epochs. Fig. 11 displays the evaluation metrics on the test data at different training epochs. Fig. 13 shows enhancement results on some LOL test data and other low-light images. Table 2 shows the metrics on the test data after 10,000 training epochs.

From Fig. 11 to Fig. 13, we can see that our method can be quickly applied to new environments, even if we only have one image from the new environment. It can be seen that in terms of visual effect and some metrics, the result of single-image training is worse than that of multi-image training. However, in our experiment, with increasing training epochs, single-image training does not produce artifacts. This proves that artifacts are not produced by the model proposed in this paper. As we train with a single low-light image, there is no need for the network to fit the histogram equalization stretch, which is a main reason why there are no artifacts in single low-light training. It may be difficult to fit the histogram equalization stretch when the network is trained without considering the whole image information in multi-image training. We believe that to avoid artifacts in multi-image training, we must make the network deeper or consider the whole image information; however, this will also increase time consumption.

5 Conclusion

In this paper, we propose a maximum entropy-based Retinex model and a self-supervised image enhancement network. The network can be trained with only low-light images and can slightly reduce noise during enhancement. Testing on real low-light images shows that with short training times, the network can produce good visual effects and achieve good real-time performance. It should be noted that our method is self-supervised, so it can adapt to new environments and devices. However, the enhanced image may differ from real data and appear more like nighttime in color. Future work will focus on color restoration, noise and artifact suppression, better detail preservation, etc. We believe these can be achieved through Generative Adversarial Networks or new constraints.

References

[1] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.

[2] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[3] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[4] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

[5] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[6] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[7] 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, 2015.

[8] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark,"in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3291–3300.

[9] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," *arXiv preprint arXiv:1808.04560*, 2018.

[10] S. Park, S. Yu, M. Kim, K. Park, and J. Paik, "Dual autoencoder network for retinex-based low-light image enhancement," *IEEE Access*, vol. 6, pp. 22 084–22 093, 2018.

[11] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Computer vision, graphics, and image processing*, vol. 39, no. 3, pp. 355–368, 1987.

[12] E. D. Pisano, S. Zong, B. M. Hemminger, M. DeLuca, R. E. Johnston, K. Muller, M. P. Braeuning, and S. M. Pizer, "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, p. 193, 1998.

[13] S. K. Naik and C. Murthy, "Hue-preserving color image enhancement without gamut problem," *IEEE Transactions on Image Processing*, vol. 12, no. 12, pp. 1591–1598, 2003.

[14] Y.-T. Kim, "Contrast enhancement using brightness preserving bi-histogram equalization," *IEEE transactions on Consumer Electronics*, vol. 43, no. 1, pp. 1–8, 1997.

[15] Y. Wang, Q. Chen, and B. Zhang, "Image enhancement based on equal area dualistic sub-image histogram equalization method," *IEEE Transactions on Consumer Electronics*, vol. 45, no. 1, pp. 68–75, 1999.

[16] T. Celik and T. Tjahjadi, "Contextual and variational contrast enhancement," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3431–3441, 2011.

[17] C. Lee, C. Lee, and C.-S. Kim, "Contrast enhancement based on layered difference representation of 2d histograms," *IEEE transactions on image processing*, vol. 22, no. 12, pp. 5372–5384, 2013.

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

[19] E. H. Land, "The retinex theory of color vision," *Scientific american*, vol. 237, no. 6, pp. 108–129, 1977.

[20] L. Li, R. Wang, W. Wang, and W. Gao, "A low-light image enhancement method for both denoising and contrast enlarging," in *2015 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2015, pp. 3730–3734.

[21] D. J. Jobson, Z.-u. 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, 1997.

[22] ——, "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.

[23] S. Wang, J. Zheng, H.-M. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3538–3548, 2013.

[24] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, and J. Paisley, "A fusion-based enhancing method for weakly illuminated images," *Signal Processing*, vol. 129, pp. 82–96, 2016.

[25] 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, 2016.

[26] R. Kimmel, M. Elad, D. Shaked, R. Keshet, and I. Sobel, "A variational framework for retinex," *International Journal of computer vision*, vol. 52, no. 1, pp. 7–23, 2003.

[27] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A variational framework for single low light image enhancement using bright channel prior," in *2013 IEEE Global Conference on Signal and Information Processing*. IEEE, 2013, pp. 1085–1088.

[28] S. Park, S. Yu, B. Moon, S. Ko, and J. Paik, "Low-light image enhancement using variational optimization-based retinex model," *IEEE Transactions on Consumer Electronics*, vol. 63, no. 2, pp. 178–184, 2017.

[29] G. Fu, L. Duan, and C. Xiao, "A hybrid l2-lp variational model for single low-light image enhancement with bright channel prior," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 1925–1929.

[30] R. L. Lagendijk and J. Biemond, *Iterative identification and restoration of images*. Springer Science & Business Media, 2012, vol. 118.

[31] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[32] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, "Toward convolutional blind denoising of real photographs," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1712–1722.

[33] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[34] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, and W. Wu, "Feedback network for image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3867–3876.

[35] S. Nah, T. Hyun Kim, and K. Mu Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3883–3891.

[36] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "Deblurgan: Blind motion deblurring using conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8183–8192.

[37] 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, 2017.

[38] C. Li, J. Guo, F. Porikli, and Y. Pang, "Lightennet: A convolutional neural network for weakly illuminated image enhancement,"*Pattern Recognition Letters*, vol. 104, pp. 15–22, 2018.

[39] J. Yang, X. Jiang, C. Pan, and C.-L. Liu, "Enhancement of low light level images with coupled dictionary learning," in *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE, 2016, pp. 751–756.

[40] L. Shen, Z. Yue, F. Feng, Q. Chen, S. Liu, and J. Ma, "Msr-net: Low-light image enhancement using deep convolutional network," *arXiv preprint arXiv:1711.02488*, 2017.

[41] J. Cai, S. Gu, and L. Zhang, "Learning a deep single image contrast enhancer from multi-exposure images," *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 2049–2062, 2018.

[42] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp. 1632–1640.

[43] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on computational imaging*, vol. 3, no. 1, pp. 47–57, 2016.

[44] X. Fu, Y. Liao, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4965–4977, 2015.

[45] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[46] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a" completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2012.

[47] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2782–2790.

[48] W. Wang, C. Wei, W. Yang, and J. Liu, "Gladnet: Low-light enhancement network with global awareness," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 2018, pp. 751–755.

Yu Zhang was born in Hebei, China. He received the B.E. degree and M.S. degree in Control Science and Engineering from Harbin Institute of Technology, China, in 2016 and 2018, respectively. He is currently pursuing a Ph.D. degree in Electronic Science and Technology at the National Key Laboratory of Tunable Laser Technology, Harbin Institute of Technology. His research interests include image restoration and enhancement, and SLAM.

Xiaoguang Di was born in Heilongjiang, China. He received the M.S. and Ph.D. degrees in navigation, guidance and control from Northwestern Polytechnical University, China, in 1999 and 2004, respectively. He is currently an associate professor with the Control and Simulation Center, Harbin Institute of Technology, where he teaches courses in digital image processing and computer vision. His current research interests include real-time image restoration and enhancement, 3D object detection and recognition, and SLAM. Prof. Di is a member of the China Simulation Federation and Chinese Society of Astronautics.

Bin Zhang was born in Gansu, China. He received the B.Sc. degree in applied physics and the M.Sc. degree in physics from China University of Petroleum, in 2010 and 2013, respectively. He is currently pursuing a Ph.D. degree at the National Key Laboratory of Tunable Laser Technology, Harbin Institute of Technology. His research interests include laser imaging and laser transmission.

Chunhui Wang received the BS, MS, and PhD degrees from Harbin Institute of Technology in 1987, 1991, and 2005, respectively. He is currently a professor and

Deputy Director of the Institute of Optoelectronic Technology, Harbin Institute of Technology. His current major research interests are laser remote sensing, lidar, laser detection and recognition.

Note: Figure translations are in progress. See original paper for figures.