2021 IEEE International Conference on Image Processing (ICIP) | 978-1-6654-4115-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICIP42928.2021.9506785

MULTI-SCALE FEATURE GUIDED LOW-LIGHT IMAGE ENHANCEMENT

Lanqing Guo[1], *Renjie Wan*[1], *Guan-Ming Su*[2], *Alex C. Kot*[1] *and Bihan Wen*[1]*

[1]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[2]Dolby Laboratories, Sunnyvale, CA, USA

ABSTRACT

Low-light image enhancement aims at enlarging the intensity of image pixels to better match human perception and to improve the performance of subsequent vision tasks. While it is relatively easy to enlighten a globally low-light image, the lighting condition of realistic scenes is usually non-uniform and complex, *e.g.*, some images may contain both bright and extremely dark regions, with or without rich features and information. Existing methods often generate abnormal light-enhancement results with over-exposure artifacts without proper guidance. To tackle this challenge, we propose a multi-scale feature guided attention mechanism in the deep generator, which can effectively perform a spatially-varying light enhancement. The attention map is fused by both the gray map and extracted feature map of the input image, to focus more on those dark and informative regions. Our baseline is an unsupervised generative adversarial network, which can be trained without any low/normal-light image pair. Experimental results demonstrate the superiority in visual quality and performance of subsequent object detection over state-of-the-art alternatives.

Index Terms— Low-Light, Generative Adversarial Network (GAN), Attention, Unsupervised Learning

1. INTRODUCTION

Low-light image is ubiquitous in practical scenes, *e.g.*, night-time surveillance, autonomous driving. However, images captured under low-light conditions often suffer from poor visibility, such as low contrast, low intensity, and high ISO noise, which challenges both human perception and the subsequent vision tasks (*e.g.*, object detection, classification, and Re-ID) [1]. Thus, restoring low-light images for better visibility is an essential problem, which has been studied in the past decades. Histogram equalization (HE) [2] first attempted to uniformly enhance the image by spreading out the most frequent intensity values. With the rapid development of the deep neural network, CNN-based methods achieved remarkable performance in low-light image enhancement [3, 4, 5]. For example, Retinex-Net [3] and KinD [5] decomposed images into illumination and reflectance in a data-driven way based on Retinex theory [6].

With the guidance of normal-light ground truth, the recent learning-based methods demonstrate more promising results comparing to the classic non-learning methods. However, most of the learning-based enhancement algorithms are supervised requiring large-scale paired training data, which limits their availability since it is difficult to obtain such training image pairs in practice. Though the synthetic images can partly alleviate the issues caused by the lack of training image pairs [4], the model trained with synthesized

* Bihan Wen (bihan.wen@ntu.edu.sg) is the corresponding author. This work was supported in part by Science and Technology Foundation of Guangzhou Huangpu Development District under Grant 2019GH16.

Fig. 1. Examples of low-light images: Globally dark (left); Partially dark (middle); Dark background without rich information (right). Note that we amplify the first image $\times 15$ for better visualization.

low-light images would impose various artifacts when transferring to real-world low-light images due to the domain gap issue. More recently, some researchers enhanced images in unsupervised or weakly-supervised way [7, 8], which can be trained without paired low/normal-light images.

Furthermore, the existing low-light image enhancement methods focus on restoring the globally dark images only as shown in the first example of Figure 1, while we argue that real images can have much more complicated lighting conditions. Figure 1 shows example images with non-uniformly low-light conditions caused by the extremely bright objects (*e.g.*, the light bulb in the second example) or dark background without much information (*e.g.*, the ocean region in the third example). Naively increasing the intensity of these images uniformly throughout the image leads to the degraded results.

To tackle these challenges, we propose a multi-scale feature guided attention map as an auxiliary module for the generator to focus more on those **dark and informative** regions. We choose an unsupervised generative adversarial network as the baseline, consisting of two networks that compete against each other. The U-Net-based generator tries to simulate realistic images to fool a discriminator by multi-level features, while the discriminator detects whether the image is fake or real distribution. The proposed self-attention map is fused by extracted feature map and the gray map, which can be flexibly plugged in any generator. To the best of our knowledge, this is the first tempt to enhance low-light images guided by features. Thanks to the guidance of the feature attention map, the over/under-exposure artifact has been effectively eliminated and has better preserved structural details on various low-light datasets. Moreover, we apply our image enhancer as pre-processing to improve the performance of object detection and experimental results show that our image enhancer outperforms other competing methods, correcting the illumination and suppressing the artifact.

2. RELATED WORK

Low-light image enhancement. Many researchers have explored the low-light image enhancement task in the past decades. Histogram equalization (HE) [2] and its follow-ups [9] spread out the most frequent intensity values to achieve uniformly contrast

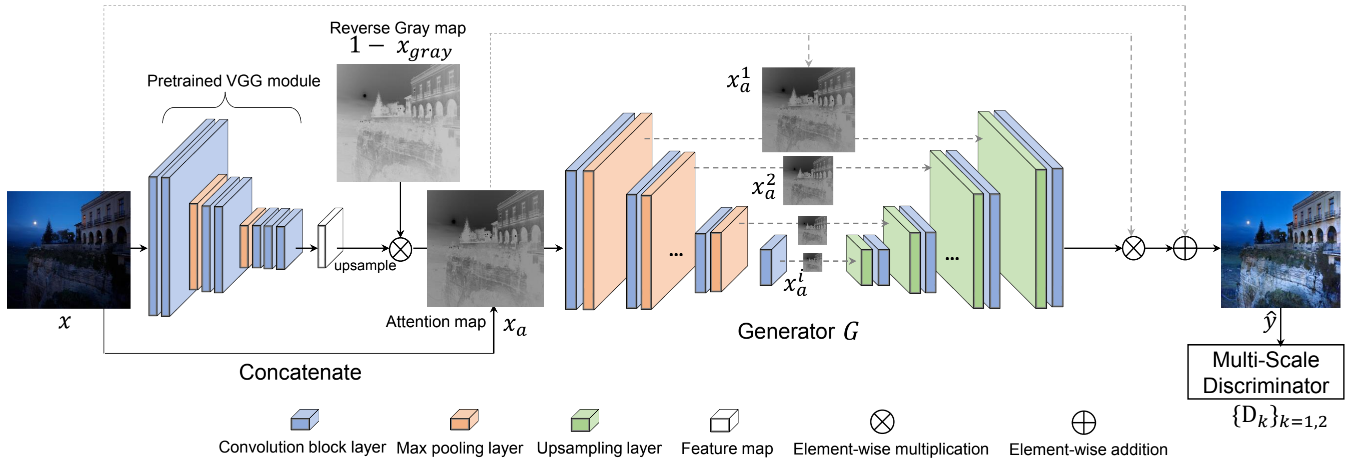

Fig. 2. The overall architecture of our model. Given the input low-light image x, first calculate self-attention map x_a by extracted feature map and gray map x_{gray}. Then feed the concatenation of input x and attention map x_a into generator G to generate fake normal-light image $\hat{y}$. The attention map x_a would be resized as $\{x_a^1, x_a^2, \ldots, x_a^i\}$ to multiply each layer of feature map from generator.

improvement. Retinex theory [6], assuming the image can be decomposed into reflectance and illumination, has been widely used in traditional illumination-based methods [10, 11]. NPE [10] jointly enhanced contrast and illumination. LIME [11] proposed a structure-aware smoothing model to estimate the illumination map. Recently, Retinex-Net [3] decomposed images into illumination and reflectance in a data-driven way based on Retinex theory. LL-Net [4] proposed a stacked auto-encoder to simultaneously conduct denoising and enhancement with synthesized low-light images. HDRNet [12] extended bilateral grid processing to the neural network and refine the color transform with the pairwise supervision. Some researchers addressed the target problem, low-light image enhancement, in an unsupervised way [7, 13, 14], which can be trained without low/normal-light image pairs. Inspired by image-image translation, EnlightenGAN [7] proposed a one-path GAN with a global-local discriminator. Zero-DCE [13] applied deep learning to simulate intensity curve adjustment. [14] targeted to adverse weather conditions and focused more on the subsequent high-level tasks rather than the quality of the generated images.

Unpaired imge-to-image translation. Image-to-image translation aims at translating images from the source domain to the corresponding target domain. For unpaired image-to-image translation, corresponding ground truth from the target domain is not available. Cycle-consistency has been widely applied to this task [15], whose key idea is obtaining the same image after transferring to the target domain and then transferring back. UNIT [16] proposed to learn a shared latent space assumption based on coupled GANs. DRIT [17] embedded images onto domain-invariant content space and domain-specific attribute space via disentanglement.

3. PROPOSED METHOD

In this section, we first introduce our generative adversarial baseline, then explain the multi-scale feature guided attention mechanism and the whole objective function. The goal of our method is to learn a mapping from low-light domain $\mathcal{X}$ to normal-light domain $\mathcal{Y}$. As shown in Figure 2, the baseline of our model is a generative adversarial network (GAN) with a generator and multi-scale discriminators. We use a multi-scale feature guided attention mechanism to guide the model to focus more on those dark and informative regions, making the enhanced image more natural and preserving image details.

U-Net based multi-scale generative adversarial model. Our generative adversarial network consists of two components that compete against each other. We apply the U-Net[18] as the backbone of generator, which extracts multi-level features and maintains rich structural information. Given the unpaired image samples $\{\mathbf{x}, \mathbf{y}\} \in \{\mathcal{X}, \mathcal{Y}\}$, the generator G tries to map the low-light image x to normal-light domain $\hat{\mathbf{y}} = G(\mathbf{x})$. We further adopt the multi-scale discriminators [19] to encourage the output to be visually similar to images from the target normal-light domain in the local and global regions, as follows,

$$\mathcal{L}_{adv} = \min_{G} \max_{D_k \in \mathcal{D}} \sum_{k=1,2} \mathcal{L}_{\mathrm{GAN}}(G, D_k) \tag{1}$$

where G denotes the generator and $\mathcal{D}$ is the set of discriminators, $D_k(k = 1, 2)$ denotes the two discriminators for different image scales and the largest one is always global discriminator, and the $\mathcal{L}_{\mathrm{GAN}}$ is defined as

$$\mathcal{L}_{\mathrm{GAN}} = E_{\mathbf{x}\sim\mathcal{X}}\left[(D_k(G(\mathbf{x})))^2\right] + E_{\mathbf{y}\sim\mathcal{Y}}\left[1 - (D_k(\mathbf{y}))^2\right] \tag{2}$$

Multi-scale feature guided self-attention. We propose a multi-scale feature guided attention mechanism for the U-Net generator to focus on regions with non-uniform illuminations. For the low-light image enhancement problem, the aim is to enhance (1) the dark region instead of the bright region, and (2) the informative region instead of smooth background. Therefore, extremely dark region without much meaningful information should not be substantially enhanced, which leads to over-exposure artifacts and damages structural details. To this end, we take the normalized gray map $\mathbf{x}_{gray} \in [0, 1]$ of input low-light image $\mathbf{x}$ and use $1 - \mathbf{x}_{gray}$ as the partial weighting map, which enables the model to attend to the low-intensity (*i.e.*, dark) region. Moreover, to focus on regions with rich structural information, we propose an attention map $\mathbf{x}_{\mathrm{a}}$ by extracting high-level features from the low-light image $\mathbf{x}$ as follows:

$$\mathbf{x}_{\mathrm{a}} = \sigma((1 - \mathbf{x}_{gray}) \circ \mathcal{U}(\phi_n(\mathbf{x}))^{\gamma}), \tag{3}$$

where ϕ_n denotes the n-th layer feature map of the pretrained VGG-16 model on ImageNet and $\circ$ designates element-wise multiplication. We choose the feature map from the first convolutional layer

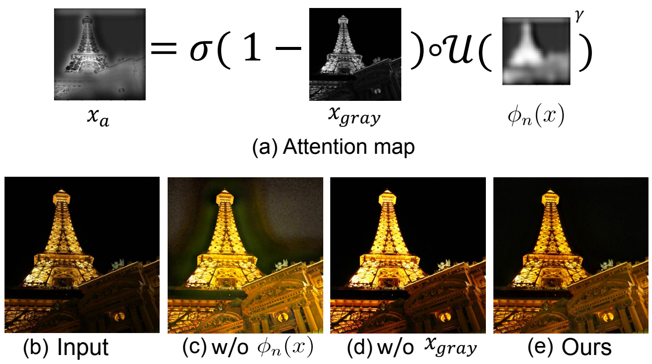

Fig. 3. (a) is visual example of our attention map. Given an input image $\mathbf{x}$, the attention map $\mathbf{x}_a$ is fused by gray map $\mathbf{x}_{gray}$ and extracted feature map $\phi_n(\mathbf{x})$. Second row is example of input low-light image (b), the result without the feature map (c), the result without the gray map (d), the result of our complete model (e).

after the fourth max-pooling layer since the high-level feature map is relatively compact and noise-free. $\gamma \geq 1$ is a parameter that controls the curvature of the feature map. $\mathcal{U}(\cdot)$ denotes up-sampling followed by sigmoid operation, which magnifies the extracted feature map to the scale of the gray map $\mathbf{x}_{gray}$. $\sigma(\cdot)$ represents sigmoid function, normalizing final attention map $\mathbf{x}_a$ to $[0, 1]$, illustrated in Figure 3(a). The results in the second row in Figure 3 show that the effectiveness of two components of self-attention map. The final estimated image without extracted feature map in attention map is with over-exposure in those extremely dark regions, and the estimated image without gray map is failing on those dark but informative regions. They facilitate performance while keeping image details and naturalness. We then downscale the attention map $\mathbf{x}_a$ to multi-scale $\{\mathbf{x}_a^1, \mathbf{x}_a^2, \ldots, \mathbf{x}_a^i\}$ (i is the number of downsampling layers in generator), matching the scale of each layer of feature map in G and multiplying it with all intermediate feature map as well as the output image, as shown in Figure 2.

Perceptual consistency loss. Furthermore, to preserve the perceptual details after generator, we employ a perceptual loss as

$$\mathcal{L}_{per} = \sum_{k=1}^{K_n} \frac{1}{K_n} \left(\phi_n\left(\mathbf{x}\right) - \phi_n\left(G(\mathbf{x})\right)\right)^2 , \tag{4}$$

which makes the generated image $\hat{\mathbf{y}} = G\left(\mathbf{x}\right)$ perceptually similar to $\mathbf{x}$ (inspired by the perceptual loss in [20]). Here ϕ_n denotes the n-th layer feature map of the pretrained VGG-16 model on ImageNet and K_n indicates the number of elements in that feature map. We also choose the high-level feature map output from the first convolutional layer after the fourth max-pooling layer.

The full objective function used for training our model is formed by combining (1) and (4) as follows,

$$\mathcal{L} = \mathcal{L}_{adv} + \alpha\mathcal{L}_{per} \ . \tag{5}$$

where α is a weighting factor balancing the influence of two terms.

4. EXPERIMENTS

4.1. Implementation Details

We have implemented the proposed model using PyTorch [21]. The U-Net generator is with 8 convolutional blocks, which consists of two 3×3 convolutional layers, followed by LeakyReLu and a batch normalization layer [22]. For the generator G, we replace the standard deconvolutional layer with one upsampling layer plus one convolutional layer, to eliminate the checkerboard artifacts [23]. For loss minimization, we adopt the ADAM optimizer, with an initial learning rate of 0.0001 and linearly decaying to zero after 100 epochs. We set the hyper-parameters $\alpha = 1$, and $\gamma = 4$. The network parameters are initialized randomly. During training, we randomly crop patches of resolution 256×256 from training data.

Table 1. Quantitative evaluation results using no-reference image quality metric NIQE $\downarrow$ on the whole testing set (All) and each subset, *i.e.*, DICM [25], LIME [11], MEF [24], and NPE [10], respectively.

	DICM	LIME	MEF	NPE	All
Input	4.255	4.438	4.265	4.319	4.280
Retinex-Net [3]	4.200	4.420	4.149	4.485	4.236
CycleGAN [15]	4.629	4.493	5.561	5.167	4.817
LIME [11]	3.846	4.155	3.720	4.268	3.889
Zero-DCE [13]	3.560	3.769	3.283	3.927	3.563
EnlightenGAN [7]	3.570	3.719	3.232	4.113	3.571
Ours	**3.450**	**3.679**	**3.137**	**3.833**	**3.450**

4.2. Comparing with state-of-the-art

Since the proposed model is based on unsupervised learning, any low/normal-light images can be used for training. We follow the same training dataset processed by EnlightenGAN [7], consisting of 914 low-light and 1016 normal-light images. We use some widely-used datasets as evaluation set to evaluate our method, *i.e.*, LIME [11], NPE [10], MEF [24], and DICM [25]. We choose the state-of-the-art hand-crafted method LIME [11], deep unsupervised methods CycleGAN [15], EnlightenGAN [7], and Zero-DCE [13], and fully supervised methods Retinex-Net [3] as the competitors. Three unsupervised methods CycleGAN [15], EnlightenGAN [7], and Zero-DCE [13] are all trained on the same dataset with us, due to no requirement of training pairs. For supervised method Retinex-Net [3], we directly run their released code and pretrained model on our evaluation set. We adopt a widely-used no-reference image quality assessment metric, *i.e.*, Natural Image Quality Evaluator (NIQE) [26], for quantitative measurement. A lower NIQE value indicates better visual quality. Table 1 shows the quantitative results. The proposed method outperforms the competing methods on all evaluation datasets.

The qualitative comparisons are shown in Figure 4. Only the results of some images in DICM [25], MEF [24], and NPE [10] datasets are shown due to the page limit. Our proposed method not only enhances the illumination but also preserves some details of the scenes. CycleGAN [15] and LIME [11] lead to color distortion for extremely dark cases and fail on some partial bright cases (*e.g.*, the second and fourth rows in Figure 4). The results of CNN-based enhancer, such as Retinex-Net, are always unnatural when generalizing to real-world low-light image cases. Though EnlightenGAN [7] makes the generated images brighter than the source low-light images, over-exposure wipes out some structural information and outcomes artifacts (*e.g.*, the second and third rows in Figure 4). The Zero-DCE [13] has less issue of over-exposure, but saturation and contrast ratio are not as good as the proposed method (*e.g.*, the fourth and fifth row in Figure 4).

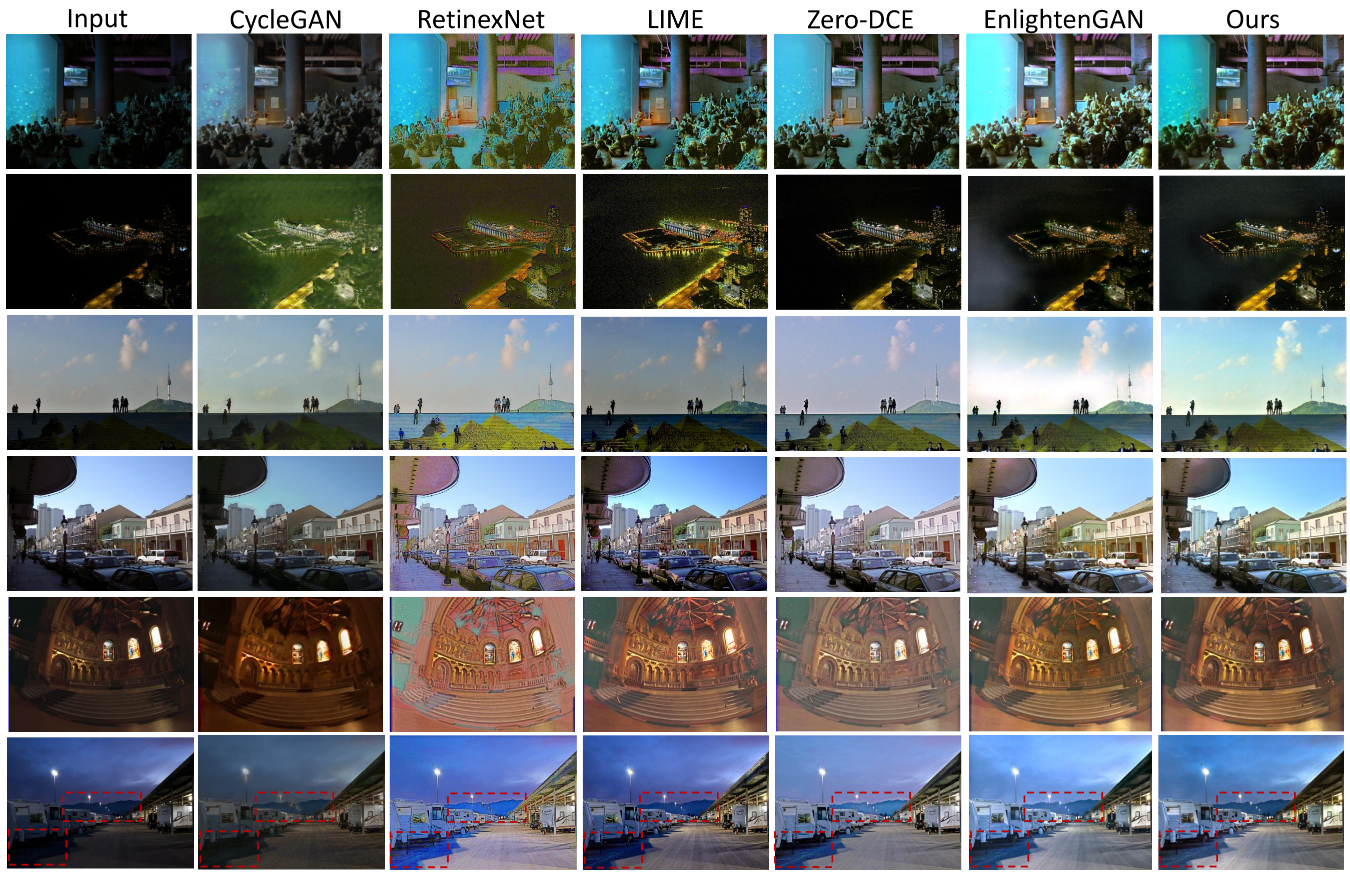

Fig. 4. Visual comparison of the results on DICM [25], MEF [24], and NPE [10] datasets, compared with CycleGAN [15], Retinex-Net [3], LIME [11], Zero-DCE [13], EnlightenGAN [7]. Red boxes indicate the over/under-exposure details where most existing methods fail.

Table 2. Quantitative evaluation using both no-reference image quality metric NIQE and detection accuracy indicator mAP on ExDark dataset [27]. EG denotes EnlightenGAN [7].

	Input	LIME	EG	CycleGAN	Zero-DCE	Ours
NIQE ↓	4.927	4.601	4.215	4.539	4.590	**4.168**
mAP ↑	55.78	55.49	54.84	51.95	55.03	**56.99**

4.3. Pre-processing for Object Detection

To further demonstrate the performance of semantic information preservation, we investigate the impact of low-light image enhancement methods on the object detection task under low-light conditions. We treat the detection results after enlightening as an indirect metric as [28]. Concretely, we use the latest extremely dark (ExDark) dataset [27], which was specifically built for the task of low-light image detection and classification. The ExDark dataset consists of 7,363 low-light images, including 3000 images in training set, 1800 images in validation set and 2563 images in testing set, annotated into 12 object classes. Due to our method is unsupervised, we use all images as dark training dataset and PASCAL VOC [29] as normal-light training dataset to adapt the domain. For the fair comparison, we also re-train those unsupervised methods, *i.e.*, CycleGAN [15], EnlightenGAN [7] and Zero-DCE [13], with the same training data as ours. We feed the results of different low-light image enhancement methods to the RFBNet detector [30], which is pre-trained on PASCAL VOC, and compare the mean average precision (mAP) on all classes in ExDark except *Cup* (since PASCAL VOC does not include *Cup* class). Table 2 shows the result of image quality NIQE and object detection mAP on the testing set of ExDark. Out method outperforms all competing methods both in the measurement of image quality and subsequent detection. Though some enhancement methods successfully enlighten input images, they damage crucial structural details and lead to unexpected artifacts. Those drop-outs and artifacts pose a great threat to the following high-level vision task since the detector is trained on natural images.

5. CONCLUSION

We propose a multi-scale feature guided image enhancer based on a vanilla unsupervised generative adversarial network to solve the image enhancement problem, which can be trained without low/normal-light image pairs. By combining the gray map and extracted feature map into an attention map, we address the difficulty of over/under-enhancement, *i.e.*, the generator effectively focuses on those dark and informative regions. The proposed multi-scale attention mechanism can be easily plugged-in any other generators. Experimental results demonstrate that our model achieves superior performance in both visual quality and subsequent high-level vision tasks. In the future, we plan to extend our feature guided attention mechanism to tackle other inverse problems in the real world, where the enhancement operation is non-uniform or spatially varying to different image local structures.

6. REFERENCES

[1] Ding Liu, Bihan Wen, Jianbo Jiao, Xianming Liu, Zhangyang Wang, and Thomas S Huang, "Connecting image denoising and high-level vision tasks via deep learning," *IEEE Transactions on Image Processing*, vol. 29, pp. 3695–3706, 2020.

[2] Mohammad Abdullah-Al-Wadud, Md Hasanul Kabir, M Ali Akber Dewan, and Oksam Chae, "A dynamic histogram equalization for image contrast enhancement," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 2, pp. 593–600, 2007.

[3] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu, "Deep retinex decomposition for low-light enhancement," *arXiv preprint arXiv:1808.04560*, 2018.

[4] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar, "Llnet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognition*, vol. 61, pp. 650–662, 2017.

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

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

[7] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang, "Enlightengan: Deep light enhancement without paired supervision," *arXiv preprint arXiv:1906.06972*, 2019.

[8] Lanqing Guo, Zhiyuan Zha, Saiprasad Ravishankar, and Bihan Wen, "Self-convolution: A highly-efficient operator for non-local image restoration," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 1860–1864.

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

[10] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3538–3548, 2013.

[11] Xiaojie Guo, Yu Li, and Haibin Ling, "Lime: Low-light image enhancement via illumination map estimation," *IEEE Transactions on image processing*, vol. 26, no. 2, pp. 982–993, 2016.

[12] Michaël Gharbi, Jiawen Chen, Jonathan T Barron, Samuel W Hasinoff, and Frédo Durand, "Deep bilateral learning for real-time image enhancement," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 4, pp. 118, 2017.

[13] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong, "Zero-reference deep curve estimation for low-light image enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1780–1789.

[14] Ziqiang Zheng, Yang Wu, Xinran Han, and Jianbo Shi, "Forkgan: Seeing into the rainy night," 2020.

[15] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[16] Ming-Yu Liu, Thomas Breuel, and Jan Kautz, "Unsupervised image-to-image translation networks," *Advances in neural information processing systems*, vol. 30, pp. 700–708, 2017.

[17] Hsin-Ying Lee, Hung-Yu Tseng, Jia-Bin Huang, Maneesh Singh, and Ming-Hsuan Yang, "Diverse image-to-image translation via disentangled representations," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 35–51.

[18] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[19] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8798–8807.

[20] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[21] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer, "Automatic differentiation in pytorch," 2017.

[22] Sergey Ioffe and Christian Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[23] Augustus Odena, Vincent Dumoulin, and Chris Olah, "Deconvolution and checkerboard artifacts," *Distill*, 2016.

[24] Kede Ma, Kai Zeng, and Zhou Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3345–3356, 2015.

[25] Chulwoo Lee, Chul Lee, and Chang-Su Kim, "Contrast enhancement based on layered difference representation," in *2012 19th IEEE International Conference on Image Processing*. IEEE, 2012, pp. 965–968.

[26] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal processing letters*, vol. 20, no. 3, pp. 209–212, 2012.

[27] Yuen Peng Loh and Chee Seng Chan, "Getting to know low-light images with the exclusively dark dataset," *Computer Vision and Image Understanding*, vol. 178, pp. 30–42, 2019.

[28] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří 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.

[29] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, June 2010.

[30] Songtao Liu, Di Huang, and andYunhong Wang, "Receptive field block net for accurate and fast object detection," in *The European Conference on Computer Vision (ECCV)*, September 2018.