BI-DIRECTIONAL NORMALIZATION AND COLOR ATTENTION-GUIDED GENERATIVE ADVERSARIAL NETWORK FOR IMAGE ENHANCEMENT

Shan Liu *Guoqiang Xiao* *Xiaohui Xu* *Song Wu**

College of Computer and Information Science, Southwest University, Chongqing, China

ABSTRACT

Most existing image enhancement methods require paired images, and rarely consider the aesthetic quality. This paper proposes a bi-directional normalization and color attention-guided generative adversarial network (BNCAGAN) for unsupervised image enhancement. An auxiliary attention classifier (AAC) and a bi-directional normalization residual (BNR) module are designed to assist the generator in flexibly controlling the local details with the constraint from both the low/high-quality domain. Moreover, a color attention module (CAM) is proposed to preserve the color fidelity in the discriminator. The qualitative and quantitative experimental results demonstrate that our BNCAGAN is superior to the existing methods with distinctively improved authenticity and naturalness of the enhanced images. The source code is available at https://github.com/SWU-CS-MediaLab/BNCAGAN.

***Index Terms*—** Image enhancement, Generative adversarial network, Unsupervised learning

1. INTRODUCTION

The rapid development of digital media technology makes it easy to capture pictures of our daily lives and upload them on social networks. However, the limitations of digital devices and the noise from the environment result in an unsatisfactory quality of the captured pictures. Thus robust image enhancement systems are required to improve the visual effect and the aesthetics of the captured low-quality images.

Relying on the robust feature abstraction capability of deep convolutional neural networks (CNNs) [1] [2] [3] [4] [5], many image enhancement methods are designed under a deep end-to-end framework [6] [7]. However, it is challenging to simultaneously capture large-scale paired low/high-quality images for deep model training. The characteristic of generative adversarial networks (GANs) [8] [9] makes it is possible to transfer the low-quality image into the high-quality domain without paired training data. A two-path GAN (two generators and two discriminators) with a cycle consistency is usually designed to train the image enhancement model in an unsupervised manner [10] [11]. However, the training process of a two-path GAN suffers from considerably more expensive computational cost. The one-path GAN methods such as EnlightenGAN [12] and UEGAN [13], are proposed to alleviate the inefficiency of training. However, existing methods only use the high-quality domain as guidance in the generator, whereas the potential benefits of also treating low-quality domain as supervise signal is ignored such as the stabilization of training process. Moreover, how to improve the authenticity and naturalness of the enhanced images is still a challenge for GAN-based methods.

In this paper, a bi-directional normalization and color attention-guided generative adversarial network is designed for high-quality image enhancement (BNCAGAN). As shown in Fig. 1, in the generator, an auxiliary attention classifier (AAC) module is designed with the motivation of improving the discriminative power of the backbone in distinguishing low-quality and high-quality domains. Inspired by [14] which shows the normalization function has a significant impact on the quality of the enhanced images in details, a bi-directional normalization residual (BNR) module is devised to achieve high-quality image enhancement with supervised signals from both low-quality and high-quality domains. Furthermore, a multi-scale fusion backbone network with a color attention module (CAM) is used in the discriminator to achieve more desirable visual effect.

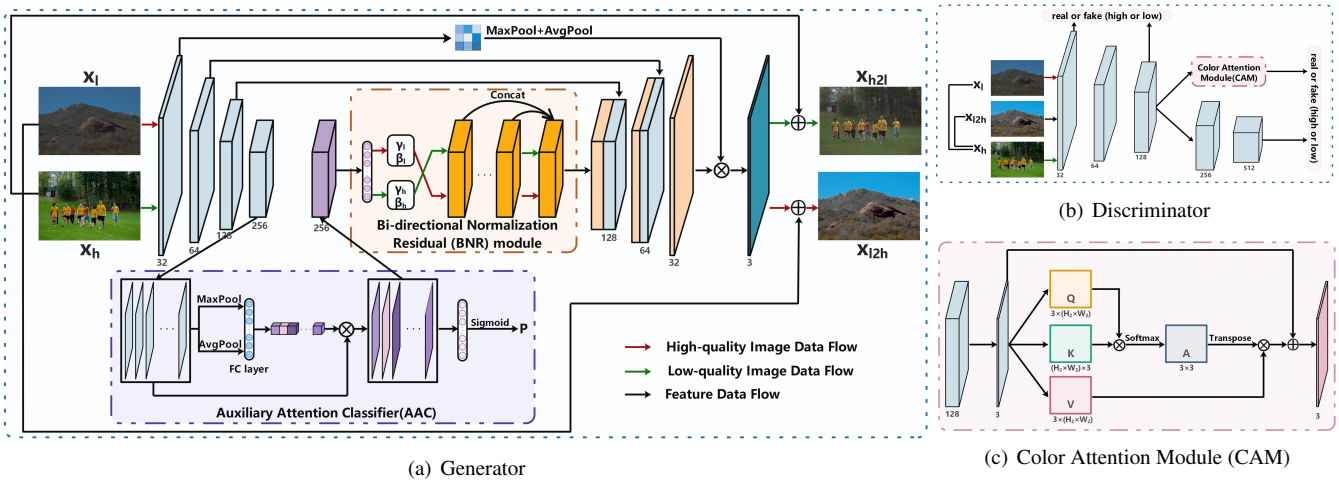
The main contributions of our proposed BNCAGAN are summarized as follows:

- Under a multi-task learning framework of the generator, the AAC module can effectively emphasis the significant features while suppressing unimportant parts.
- The proposed BNR module can learn the information from both high-quality and low-quality domains to supervised the image enhancement training while adaptively retain content information.
- In the discriminator, the CAM module can effectively strengthen the color authenticity by focusing on global and local color details in a self-attention manner.
- The experimental results on two popular datasets quantitatively and qualitatively demonstrated the superiority of our BNCAGAN.

2. THE PROPOSED BNCAGAN

In this section, we first introduce the AAC and BNR modules in the generator, and then analyze the discriminator structure.

*Corresponding author.



(a) Generator (b) Discriminator (c) Color Attention Module (CAM)

Fig. 1. Overview of BNCAGAN. (a) is the generator with the auxiliary attention classifier (AAC) and the bi-directional normalization residual (BNR) modules. The features of inputs are extracted through the backbone network and then fed into the AAC to calculate the importance of each channel. The image enhancement is realized through the BNR. (b) is the discriminator with the color attention module (CAM). (c) is the detailed implementation of CAM.

2.1. Generator

As shown in Fig. 1(a), $x \in \{X_l, X_h\}$ represent an unpaired sample from the low-quality and the high-quality domain, respectively. Our goal is to train an image enhancement generator G, which can be expressed as:

$$x_{l2h} = G\left(x_l\right), \tag{1}$$

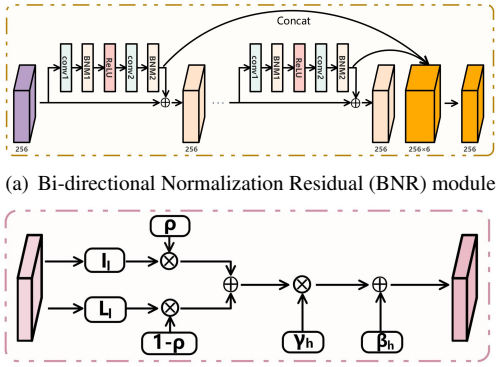
where x_l is the image to be enhanced, x_{l2h} is the result we expect to obtain. Similarly, an image degradation process $x_{h2l} = G\left(x_h\right)$ can also be obtained in the training process.

In order to make the enhanced results closer to human perception, it is necessary to preserve the semantic content, only modify the global quality domain, and adjust the details in a targeted manner. Therefore, the generator adopts a symmetrical encoder-decoder structure with skip connections similar to U-Net [15], which effectively keeps the high-level semantics consistent before and after enhancement. Besides taking the global information into account, the auxiliary attention classifier (AAC) and the bi-directional normalization residual (BNR) module are designed for local detail optimization through channel attention and the supervision from the characteristics of both low-quality and high-quality domains.

Auxiliary Attention Classifier. For each input image, a series of convolution operations are performed to obtain the intermediate features. For the specific output feature maps $u \in \mathbb{R}^{C\times H\times W}$ from the last layer of the encoder, the attention mask u_{mask}, which calculated the significance of each channel, is expressed as:

$$u_{mask} = \sigma\left(F_{fc}\left(F_{gp}(u)\right)\right), \tag{2}$$

where $F_{gp}(\cdot)$ is the concatenation of the output from global average pooling and global max pooling operations, $F_{fc}(\cdot)$ represents the fully connection operation. The updated $\hat{u}$ is obtained by multiplying the attention mask u_{mask} and u to increase the importance of the effective channels and reduce the weight of the invalid channels. The learning of the attention mask is under the supervision of the binary classification loss and the generator loss. Thus, it can effectively assist the generator in obtaining the distribution of high-quality domains in a multi-task learning manner.



(a) Bi-directional Normalization Residual (BNR) module

(b) Bi-directional Normalization Module (BNM)

Fig. 2. (a) shows the details of bi-directional normalization residual (BNR) module. The details of the Bi-directional normalization module (BNM) are illustrated in (b).

Bi-directional Normalization Residual Module. A bi-directional normalization function is designed to consider the global quality domain and flexibly control the local details of the enhanced images by the learned parameters from the normalization functions. As shown in Fig. 2, the residual blocks are applied in the proposed bi-directional normalization module (BNM), where the parameters are learned dynamically to enable the generator to control local details flexibly with the transfer constraint from low-quality to high-quality and high-quality to low-quality domains during the training process. The BNM is calculated as:

$$\begin{aligned} \mathrm{BNM}(IN, LN, \gamma, \beta) &= \gamma \cdot (\rho \cdot IN + (1-\rho)\cdot LN) + \beta, \\ \rho &\leftarrow \mathrm{clip}_{[0,1]}(\rho - \tau\Delta\rho), \end{aligned} \tag{3}$$

where $\gamma = \{\gamma_h, \gamma_l\}$ and $\beta = \{\beta_h, \beta_l\}$ are calculated by a fully connected layer on $\hat{u}_h$ and $\hat{u}_l$. The IN and LN are

instance normalization [16] and layer normalization [17] operated on both $\hat{u}_h$ and $\hat{u}_l$. $IN = \frac{\hat{u}_{h/l} - \mu_{IN}}{\sqrt{\sigma_{IN}^2 + \epsilon}}$, $LN = \frac{\hat{u}_{h/l} - \mu_{LN}}{\sqrt{\sigma_{LN}^2 + \epsilon}}$, where μ_{IN}, μ_{LN} and σ_{IN}, σ_{LN} are the mean and variance on channel and layer, respectively. τ represents the learning rate, and $\Delta\rho$ means the parameter update vector (e.g., the gradient) determined by the optimizer to adjust the weights for two normalization methods.

2.2. Discriminator

The discriminator is shown in Fig. 1(b). A multi-scale fusion framework is employed in the discriminator. Thus, it can distinguish low/high-quality domains from a global perspective and consider local details from receptive fields with different sizes. In addition, we observe that the details of color information significantly influence the authenticity and naturalness of enhanced images. A color attention module (CAM) designed by self-attention mechanism [18] is used to strengthen the effective color information. As shown in Fig. 1(c), $I \in \mathbb{R}^{3\times H_2\times W_2}$ is firstly transformed into feature spaces to generate Q, $V \in \mathbb{R}^{3\times(H_2\times W_2)}$ and $K \in \mathbb{R}^{(H_2\times W_2)\times 3}$, where 3 means the number of color channels. The transpose of K is multiplied with Q to obtain the color attention matrix $A \in \mathbb{R}^{3\times 3}$ which represents the correlation among the three-color channels. The color attention feature maps can be obtained by multiplying the transpose of A and V.

3. LOSS FUNCTION

Adversarial loss. In our BNCAGAN, there are two adversarial relationships, $\{x_l, x_h\}$ and $\{x_g, x_h\}$. During the process of adversarial learning, an image with a similar domain distribution to the real distribution of a high-quality domain can be generated by the generator. Specifically, inspired by RaHingeGAN [19], in order to significantly improve the quality and stability of GANs, the adversarial loss in our BNCAGAN is calculated as follows:

$$\begin{aligned} L_D =& \mathbb{E}_{x_h\sim P_h}\left[\max\left(0, 1-\left(D\left(x_h\right)-\mathbb{E}_{x_l\sim P_l}D\left(x_l\right)\right)\right)\right] \\ &+\mathbb{E}_{x_l\sim P_l}\left[\max\left(0, 1+\left(D\left(x_l\right)-\mathbb{E}_{x_h\sim P_h}D\left(x_h\right)\right)\right)\right] \\ &+\mathbb{E}_{x_h\sim P_h}\left[\max\left(0, 1-\left(D\left(x_h\right)-\mathbb{E}_{x_l\sim P_l}D\left(G\left(x_l\right)\right)\right)\right)\right] \\ &+\mathbb{E}_{x_l\sim P_l}\left[\max\left(0, 1+\left(D\left(G\left(x_l\right)\right)-\mathbb{E}_{x_h\sim P_h}D\left(x_h\right)\right)\right)\right], \end{aligned} \tag{4}$$

$$\begin{aligned} L_{adv}^G =& \mathbb{E}_{x_l\sim P_l}\left[\max\left(0, 1-\left(D\left(G\left(x_l\right)\right)-\mathbb{E}_{x_h\sim P_h}D\left(x_h\right)\right)\right)\right] \\ &+\mathbb{E}_{x_h\sim P_h}\left[\max\left(0, 1+\left(D\left(x_h\right)-\mathbb{E}_{x_l\sim P_l}D\left(G\left(x_l\right)\right)\right)\right)\right], \end{aligned} \tag{5}$$

Perceptual loss. Adversarial loss is not sufficient to guarantee semantic consistency. In order to maintain the perceptual similarity of the images before and after the enhancement process, a perceptual loss [20] is employed to constrain the image content at the high semantic level:

$$\begin{aligned} L_{per} =& \sum_{j=1}^{J}\mathbb{E}_{x_l\sim P_l}\left[\left\|\varphi_j\left(x_l\right)-\varphi_j\left(G\left(x_l\right)\right)\right\|_2\right] \\ &+\sum_{j=1}^{J}\mathbb{E}_{x_h\sim P_h}\left[\left\|\varphi_j\left(x_h\right)-\varphi_j\left(G\left(x_h\right)\right)\right\|_2\right], \end{aligned} \tag{6}$$

where $\varphi_j(\cdot)$ represents the features extracted by the j_{th} layer of the pre-trained backbone network.

Identity loss. Identity loss encourages the enhanced images to have a similar color and contrast to high-quality images. It guides the learning of both low and high-quality domains:

$$\begin{aligned} L_{idt} =& \mathbb{E}_{x_h\sim P_h, x_l\sim P_l}\left[\left\|x_h - G\left(x_l\right)\right\|_1\right] \\ &+\mathbb{E}_{x_h\sim P_h, x_l\sim P_l}\left[\left\|x_l - G\left(x_h\right)\right\|_1\right]. \end{aligned} \tag{7}$$

Auxiliary classification loss. By classifying the real or fake images based on the auxiliary classification loss, the generator pays more attention to crucial regions of the generated images that are difficult to distinguish.

$$L_{cla} = -\left\{\mathbb{E}_{x_h\sim P_h}\left[log\left(C\left(x_h\right)\right)\right] + \mathbb{E}_{x_l\sim P_l}\left[log\left(1-C\left(x_l\right)\right)\right]\right\}, \tag{8}$$

where $C(\cdot)$ means the process of binary classification.

Total loss. The total loss for can be defined as:

$$L_G = \lambda_{adv}L_{adv}^G + \lambda_{per}L_{per} + \lambda_{idt}L_{idt} + \lambda_{cla}L_{cla}, \tag{9}$$

where λ_{adv}, λ_{per}, λ_{idt}, λ_{cla} are the weights to balance the multiple objectives.

4. EXPERIMENTS

4.1. Dataset and Implementation Details

MIT-Adobe FiveK dataset [21]: it is randomly divided into three parts: (1) 2250 original images and 2250 modified images are used for training; (2) 400 pair-images are used for testing; (3) 100 pair-images are used for validation. We select the images processed by photographer C as labels.

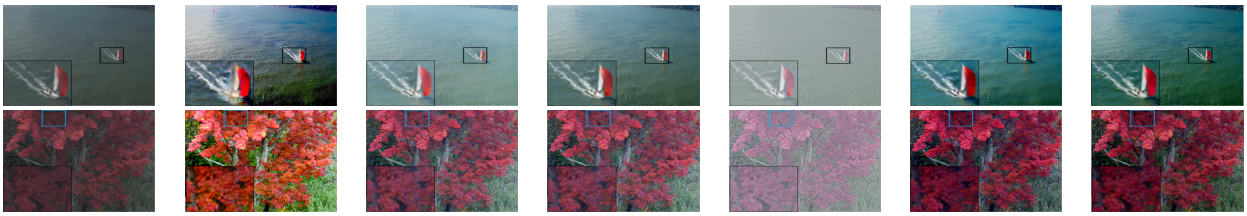
DPED dataset [6]: the image are taken by three smartphones and one DSLR camera. The images taken by iPhone are trained and tested as low-quality images in our experiments.

We implement our network in Pytorch. The first 75 epochs are trained with the learning rate of 0.0001, and it decays to zero in another 75 epochs. The λ_{adv}, λ_{per}, λ_{idt}, λ_{cla} are set as 0.1, 1, 0.1, 0.1 in the experiment.

4.2. Comparison to State-of-the-art

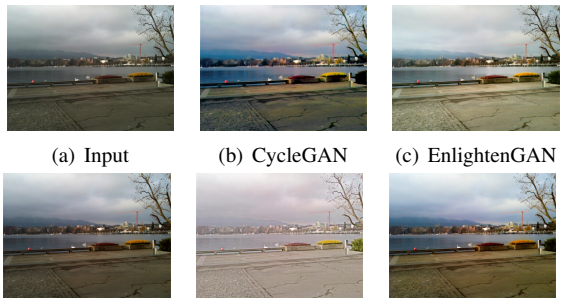
To verify the effectiveness of our BNCAGAN, we compare it with several state-of-the-art image enhancement methods.

Qualitative Comparison. As shown in Fig. 3 and Fig. 4, the input images are characterized by low brightness, low saturation, and unclear texture. The results of CycleGAN have changed the contrast excessively, and the shape of the object cannot be preserved. It obtained the most unsatisfactory visual effect among all the methods. EnlightenGAN results in simply being brighter overall, while also suffering from a lack of color. The results of UEGAN look good, but there are weird textures that do not belong to inputs, especially on dense objects. And Zero-DCE can only brighten images. In contrast, our BNCAGAN has bright colors, precise details, and textures for optimum performance.



(a) Input (b) CycleGAN (c) EnlightenGAN (d) UEGAN (e) Zero-DCE (f) BNCAGAN (g) Label

Fig. 3. From left to right, the comparison results of MIT-Adobe FiveK are displayed. Besides qualitative comparison with existing state-of-the-art methods in global, for the detailed textures, we have magnified them to better reflect the image quality.



(a) Input (b) CycleGAN (c) EnlightenGAN (d) UEGAN (e) Zero-DCE (f) BNCAGAN

Fig. 4. The comparison results of DPED dataset.

Quantitative Comparison. For the quantitative evaluation, besides the commonly used PSNR and SSIM, NIMA [22] is also used to evaluate aesthetic quality. From Table. 1, the score of CycleGAN is relatively low. The reason may be the significant change in contrast and the lack of details. The results of EnlightenGAN and Zero-DCE are not ideal because the overall change in contrast resulting the images becoming dull. The generated redundant textures make the effects of UEGAN worse. Our BNCAGAN shows superiority in quantitative evaluation.

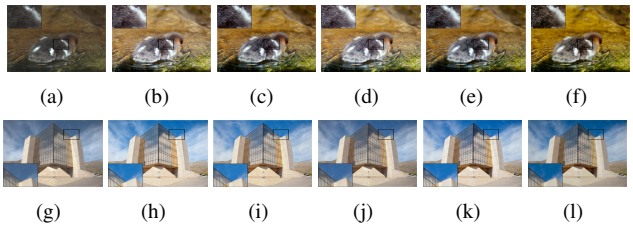
Table 1. Quantitative comparison of PSNR, SSIM, and NIMA on MIT-Adobe FiveK dataset.

Method	PSNR	SSIM	NIMA
Input	18.131	0.810	4.472
CycleGAN [10]	21.708	0.815	4.620
EnlightenGAN [12]	20.142	0.867	4.648
UEGAN [13]	23.936	0.890	4.734
Zero-DCE [23]	12.509	0.727	4.200
BNCAGAN	**24.332**	**0.910**	**4.810**

4.3. Ablation Study

Loss Analysis. In this part, we show the influence of L_{per}, L_{idt} and L_{cla}. From the first row of Fig. 5, we observe that L_{idt} can suppress excessive enhancement and keep the authenticity of colors, and L_{cla} pays more attention to color saturation, which is related to distinguishing two domains in color. From Table. 2, we can find that L_{cla} leads to a relatively large improvement on PSNR, and L_{idt} can further improve the effect.

Architecture Analysis. In order to verify the effectiveness of AAC, BNM, and CAM, we remove these three modules for objective comparison. As shown in the second row in Fig. 5, we find that the AAC module has little effect on the vision, the BNM module helps in the local color conversion of the image, and the CAM will affect the overall visual effect. From Table. 2, the combination of AAC, BNM, and CAM can achieve a balanced optimization result.



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k) (l)

Fig. 5. The results of ablation experiments on MIT-Adobe FiveK. (a) Input$_1$, (b) w/o L_{idt}, L_{cla}, (c) w/o L_{idt}, (d) w/o L_{cla}, (e) BNCAGAN$_1$, (f)Label$_1$, (g) Input$_2$, (h) w/o AAC, (i) w/o BNM, (j) w/o CAM, (k) BNCAGAN$_2$, (l) Label$_2$.

Table 2. Quantitative comparison of our method with different loss and modules on MIT-Adobe FiveK.

Method	PSNR	SSIM	NIMA
BNCAGAN w/o L_{idt}, w/o L_{cla}	23.910	0.905	4.772
BNCAGAN w/o L_{idt}	24.140	0.909	4.804
BNCAGAN w/o L_{cla}	24.004	0.904	4.790
BNCAGAN w/o AAC	24.124	0.907	**4.820**
BNCAGAN w/o BNM	24.201	**0.910**	4.800
BNCAGAN w/o CAM	23.730	0.900	4.778
BNCAGAN	**24.332**	**0.910**	4.810

5. CONCLUSION

This paper proposes a novel bi-directional normalization and color attention-guided generative adversarial network (BNCAGAN) for high-quality image enhancement. The BNCAGAN consists of an auxiliary attention classifier (AAC), a bi-directional normalization residual (BNR) module, and a color attention module (CAM). The AAC module can effectively focus on significant feature channels. The BNR module can supervise the enhancement from both high-quality and low-quality domains by the learned normalization parameters. The CAM can strengthen extraction of detailed color information to achieve more robust visual effects. We qualitatively and quantitatively demonstrate that our method is superior to the existing state-of-the-art methods.

6. REFERENCES

[1] J. Liu, W. Zhang, Y. Tang, J. Tang, and G. Wu, "Residual feature aggregation network for image super-resolution," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[2] Song Wu, Ard Oerlemans, Erwin M Bakker, and Michael S Lew, "Deep binary codes for large scale image retrieval," *Neurocomputing*, vol. 257, pp. 5–15, 2017.

[3] Xinzhi Wang, Xitao Zou, Erwin M Bakker, and Song Wu, "Self-constraining and attention-based hashing network for bit-scalable cross-modal retrieval," *Neurocomputing*, vol. 400, pp. 255–271, 2020.

[4] S. Chen, S. Wu, and L. Wang, "Hierarchical semantic interaction-based deep hashing network for cross-modal retrieval," *PeerJ Computer Science*, vol. 7, no. 2, pp. e552, 2021.

[5] Xitao Zou, Xinzhi Wang, Erwin M Bakker, and Song Wu, "Multi-label semantics preserving based deep cross-modal hashing," *Signal Processing: Image Communication*, vol. 93, pp. 116131, 2021.

[6] A. Ignatov, N. Kobyshev, R. Timofte, K. Vanhoey, and L. V. Gool, "Dslr-quality photos on mobile devices with deep convolutional networks," *IEEE Computer Society*, pp. 3297–3305, 2017.

[7] W. Ren, S. Liu, L. Ma, Q. Xu, X. Xu, X. Cao, J. Du, and M. H. Yang, "Low-light image enhancement via a deep hybrid network," *IEEE Transactions on Image Processing*, pp. 1–1, 2019.

[8] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Advances in Neural Information Processing Systems*, vol. 3, pp. 2672–2680, 2014.

[9] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[10] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *IEEE*, 2017.

[11] S. C. Yu, Y. C. Wang, M. H. Kao, and Y. Y. Chuang, "Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans," in *IEEE International Conference on Computer Vision and Pattern Recognition*, 2018.

[12] 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," *IEEE Transactions on Image Processing*, vol. 30, pp. 2340–2349, 2021.

[13] Z. Ni, W. Yang, S. Wang, L. Ma, and S. Kwong, "Towards unsupervised deep image enhancement with generative adversarial network," *IEEE Transactions on Image Processing*, vol. PP, 2020.

[14] Junho Kim, Minjae Kim, Hyeonwoo Kang, and Kwanghee Lee, "U-gat-it: unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation," *arXiv preprint arXiv:1907.10830*, 2019.

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015.

[16] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *arXiv preprint arXiv:1607.08022*, 2016.

[17] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton, "Layer normalization," *arXiv preprint arXiv:1607.06450*, 2016.

[18] Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena, "Self-attention generative adversarial networks," in *International conference on machine learning*. PMLR, 2019, pp. 7354–7363.

[19] Alexia Jolicoeur-Martineau, "The relativistic discriminator: a key element missing from standard gan," *arXiv preprint arXiv:1807.00734*, 2018.

[20] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," *Springer, Cham*, 2016.

[21] V. Bychkovsky, S. Paris, E. Chan, and F. Durand, "Learning photographic global tonal adjustment with a database of input/output image pairs," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 2011.

[22] Hossein Talebi and Peyman Milanfar, "Nima: neural image assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 3998–4011, 2018.

[23] C. Guo, C. Li, J Guo, C. C. Loy, J Hou, S. Kwong, and R. Cong, "Zero-reference deep curve estimation for low-light image enhancement," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.