UMLE: Unsupervised Multi-discriminator Network for Low Light Enhancement*

Yangyang Qu[1,2,3,4], Kai Chen[1,2,3,4], Chao Liu[1,3,4] and Yongsheng Ou[1,3,4]

Abstract— Low-light image enhancement is a complex and vital task including, recovering color and texture details from low-light images. For automated driving, low-light scenarios will have severe implications for vision-based applications. To address this problem, we propose a real-time unsupervised generative adversarial network (GAN) with multiple discriminators. It includes a multi-scale discriminator, a texture discriminator, and a color discriminator to evaluate images from different perspectives. Furthermore, considering the uneven illumination distribution of images and the different information contained in the channels, we propose a feature fusion attention module to combine channel attention with pixel attention to extract image features. Experiments indicate that our method outperforms state-of-the-art methods in qualitative and quantitative evaluation and provides significant improvements in localization and detection results for automated driving.

I. Introduction

Enhancing the low-light image is a complex task with critical applications in numerous fields. For autopilot, taking images in low-light environments severely affects the visual effects and causes numerous difficulties, such as high noise and loss of image details. For simultaneous localization and mapping (SLAM), the localization and mapping processes are seriously affected by low-light images. Besides, many advanced tasks will be severely influenced such as image instance segmentation and depth estimation. These problems will severely affect the driving safety of automated vehicles.

Over the last few decades, researchers have developed various theories to obtain low-light enhancement within three groups. The first group is image histogram equalization [1] and correlation algorithms [2], [3]. This technique is useful when both the foreground and background are both too dark or bright. However, its disadvantage is that it may increase the contrast of background noise and reduce the contrast of valuable areas. The second group is retinex theory [4] and its variants [5]–[8]. It adaptively improves various types of images. However, the processing results are greatly affected by the chromatic aberration variation and noise. The third group is convolutional neural networks and correlation algorithms [9]–[13]. Using neural network techniques eases image enhancement practice, however, a substantial number of dark-bright paired images is required for training.

(a) (b)

Fig. 1. Results of ORB matching after enhancement of the low-light images by our proposed real-time low-light enhancement method UMLE.

This paper presents a multi-discriminator generative adversarial network (UMLE) to restore the images' color and texture features. The discriminator consists of three branches. The first branch is a multi-scale discriminator for which features of different scales can be examined adequately. The second branch is a color discriminator that represents the authentic color of the generated images. The third branch is a texture discriminator evaluating the sharpness and clarity of edges in the generated images. In addition, this paper proposes an effective channel and pixel attention module which can focus on both channel information and pixel information in an effective way. Moreover, we added a network sharing the encoder with the generator and the discriminator which cuold reduce the number of model parameters while accelerating the training. The results of extensive experimental studies indicate that our method outperforms the state-of-the-art methods in terms of visual effects and metrics.

The contributions of this paper are summarized as follows:

- We proposed a low-light enhancement GAN. The model was independently trained of the paired training data. Hence, images with various scenes and

*This work was jointly supported by National Key Research and Development Program of China under Grant 2018AAA0103001, National Natural Science Foundation of China (Grants No. U1813208, 62063006), Guangdong Special Support Program (2017TX04X265) and Shenzhen Fundamental Research Program (JCYJ20200109115610172).

[1]The authors are with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China.

[2]University of Chinese Academy of Sciences, Beijing 100049, China.

[3]Guangdong-Hong Kong-Macao Joint Laboratory of Human-Machine Intelligence-Synergy Systems (#2019B121205007).

[4]CAS Key Laboratory of Human-Machine Intelligence Synergic Systems, Shenzhen Institutes of Advanced Technology, Shenzhen 518055, China. Yongsheng Ou is the corresponding author. Email: ys.ou@siat.ac.cn.

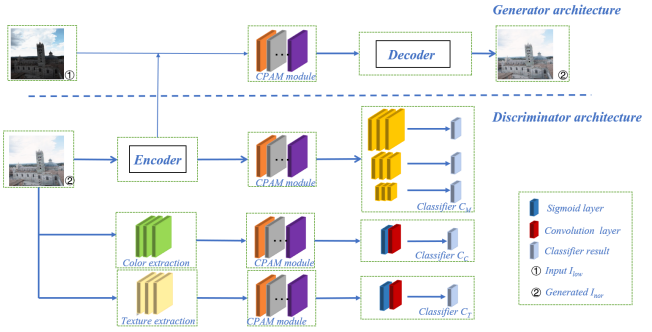


Fig. 2. The architecture of our network. The presented network consisting of two branches. One branch has the original image, and the second improves the original image after edge detection. Specific details are described in Sec.II.

different illumination levels are used to effectively improve the generalization ability of the model.

- We designed a multi-branch discriminator to comprehensively assess the image from color, texture, and global information. Besides, we put forward a novel attention module to combine channel attention and pixel attention. It also helps to further focus more on the low-light areas.
- The proposed method achieves an improvement in automatic driving tasks such as SLAM repositioning and driveable area detection under severe light changes.

II. Proposed Model

A. Overview

As seen in Fig. 2, we designed an unsupervised GAN structure not requiring paired training data to improve the images. Let L and N as the low-light domain and the normal-light domain. $x \in X_L, X_N$ represents x from the low-light domain and normal-light domain. Our model studies a transformation from the L domain to the N domain.

Our model includes two generators ($G_{L\to N}$ transforms low-light image to normal-light image and $G_{N\to L}$ transforms normal-light image to low-light image). Moreover, two discriminators are included (D_N distinguishes the generated low-light images from the real low-light images, D_L distinguishes the generated normal-light images from the real normal-light images). A generator typically involves an encoder and a decoder, and a discriminator contains an encoder and a classifier. Our generators and discriminators share the same encoder, and using this approach results in a significant improvement in training stability and model size [14].

Since most of the image's details are hidden in the dark, there will be texture error, color deviation, and other problems when generating the image. This paper presents a multi-branch discriminator for solving this problem including a texture discriminator, a color discriminator, and a multi-scaled discriminator.

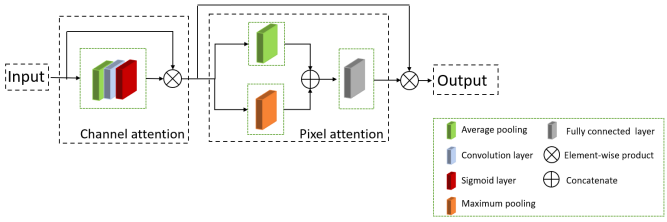


Fig. 3. The architecture of the CPA module. It contains a channle attention module and an attention module. Specific details are described in Sec. Channel and Pixel Attention Module.

B. Channel and Pixel Attention Module (CPA)

A previous study [15] revealed that channel attention could effectively enhance convolutional neural networks' performance. Wang et al. [16] demonstrate the avoiding descending and appropriate cross-channel interactions is important for learning effective channel attention. In this paper, we presented an attention module combining channel attention and pixel attention, which is shown in Fig. 3.

The channel attention can achieve channel attention without a dimensionality reduction. First, the module changes the input feature graph from $C \times H \times W$ to $C \times 1 \times 1$ through a channel-wise global average pooling, where C, W, H mean channel dimension, width, and height. Suppose that $\mathcal{X} \in \mathbb{R}^{H\times W\times C}$, the global average pooling can be obtained by:

$$g_c(\mathcal{X}) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \mathcal{X}_{ij}. \tag{1}$$

Then, the module multiplies the input by the weights of the different channels to obtain the channel attention result:

$$c_r = \sigma\left(\text{conv}\left(g_c\right)\right) \otimes \mathcal{X}. \tag{2}$$

where σ represents a sigmoid function, and $conv$ is a convolution function.

The pixel attention module can yield the pixel attention adaptively. We splice the global average pooling and global max pooling together to achieve the pixel attention feature:

$$p_r = cat(\sigma(c_r), \gamma(c_r)), \tag{3}$$

where γ represents the fully connected function.

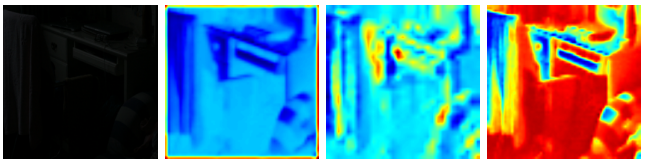
Then, the output is the result of the CPA module:

$$cp_r = \gamma(p_r) \otimes c_r, \tag{4}$$

By adding a residual module in conjunction with CPA, the proposed feature extraction module (CPA) can dynamically yield the features and better obtain the pixel and channel distribution.

C. Multi-branch Discriminator

The discriminator includes two parts, encoder E_D and classifier C_D. Utilizing only a single discriminator normally causes problems with chromatic aberration,



(a) (b) (c) (d)

Fig. 4. The feature extraction results of the three discriminators. (a) is the original image, (b) is the visualization of the color discriminator, (c) is the visualization of the texture discriminator, (d) is the visualization of the multi-scale discriminator. It can be found that each discriminator concentrates on different key information.

blurring of edges, and abnormal exposure in locally varying lighting. To counter these problems, we put forward a multi-branch discriminator D_M, consisting of three parts.

1) Color discriminator: The first part is a color discriminator D_c, for which the principal goal is to discriminate the image through the color characteristics. It is supposed to learn the color differences and contrast between the original image I_{ori} and the generated image I_{gen}, while avoiding texture comparisons. Inspired by [6], we adopted the idea of frequency separation. For this part, we applied a Gaussian low-pass filter:

$$G_{x,y} = \lambda \exp\left(-\frac{(x-\mu_x)^2}{2\sigma_x} - \frac{(y-\mu_y)^2}{2\sigma_y}\right), \quad (5)$$

where $\lambda = 0.053, \mu_{x,y} = 0$, and $\sigma_{x,y} = 3$. Then, the features are extracted through the encoder module and the CPA module to dynamically discriminate the image.

2) Texture discriminator: The second part is the texture feature discriminator D_T, discriminating the texture features of the image. Previous attempts proposed a texture classifier emulating the process of extracting SIFT descriptors [17]. They convolved it with the two filters to achieve the same x, y gradients for the discriminator in a differentiable manner. In contrast, we used a Gaussian high-pass filter in our technique. Through this filter, we can extract the image's texture information while avoiding the influence of color and other information. Then, features are extracted by the encoding module and CPA module to better discriminate the images from the texture perspective.

3) Multi-scale discriminator: For the third part, we adopted a multi-scale discriminator D_S. Previous experiments revealed that a multi-scale classifier is helpful to enhance the effect of discriminators. Different from EnlightenGAN [18] using a global-local discriminator, a three-scale classifier is trained in our model to judge whether the image is real or not. It is often essential to enhance details in different scales such as the lightened area. These regions are not considered in a global classifier, for example, to enhance some details present in different scales, such as the lightened area. Focusing another two scales in the classifier on different sizes of regions, making the resulting image more realistic. The proposed discriminator contains a local discriminator D_{SL} which is a $10 * 10$ pixel image patch.

To intuitively illustrate the function of the three discriminators, we visualized their results after the CPA module in Fig. 4. It is observed that different discriminators concentrate on different information of the image. The texture discriminator mainly considers edge regions and texture details, while the color discriminator concentrates more on the color of the whole image. However, multi-scale discriminators take into account the overall features of the image.

D. Generator

For the generator, we used a shared structure, using the same encoder as the discriminator. For the decoder, we also utilized CPA to extract features. Then, the model can obtain the final result image.

E. Loss function

To train the UMLE model, the following five losses are adopted:

1) Adversarial loss: We adopted the LSGAN [19] loss as the adversarial loss. This loss matches the distribution of translated images and target images.

$$\begin{aligned}\mathcal{L}_{adv} = &\mathbb{E}_{x\sim\mathcal{X}_N}\left[\log D_N(x)\right] \\ &+ \mathbb{E}_{x\sim\mathcal{X}_L}\left[\log\left(1 - D_N(G_{L\to N}(x))\right)\right].\end{aligned} \quad (6)$$

2) Cycle loss: This loss is adopted to make the image as similar as possible to the original image after two transformations.

$$\mathcal{L}_{cyc} = \mathbb{E}_{x\sim\mathcal{X}_L}\left[\|x - G_{N\to L}(G_{L\to N}(x))\|_1\right]. \quad (7)$$

3) Color loss: Color loss forces improved the images to have similar color distributions as the target high-quality images.

$$\mathcal{L}_{\text{color}} = \mathbb{E}_{x\sim\mathcal{X}_L}[\log(D_N\left(G_{L\to N}(x)\right))]. \quad (8)$$

4) Preserving Loss: The function of preserving loss is to limit the vgg characteristic distance between the input low-light level and its enhanced normal-light output.

$$\mathcal{L}_{pre} = \frac{1}{W_i H_i}\sum_{x=1}^{W_i}\sum_{y=1}^{H_i}\left(\phi_i\left(x\right) - \phi_i\left(G_{L\to N}(x)\right)\right)^2, \quad (9)$$

where ϕ_i represents the i-th feature map extracted from a pre-trained VGG-16 model; and H_i and W_i denote the height and width of the feature maps, respectively.

5) Reconstruction loss: When the real sample of the source domain is provided as the input of the source domain generator, the transformed image should be the same as the real sample.

$$\mathcal{L}_{idt} = \mathbb{E}_{x\sim\mathcal{X}_N}\left[|x - G_{L\to N}(x)|_1\right]. \quad (10)$$

The entire loss function of our model is as follows:

$$\mathcal{L}_{all} = \omega_1\mathcal{L}_{adv} + \omega_2\mathcal{L}_{cyc} + \omega_3\mathcal{L}_{color} + \omega_4\mathcal{L}_{pre} + \omega_5\mathcal{L}_{idt}. \quad (11)$$

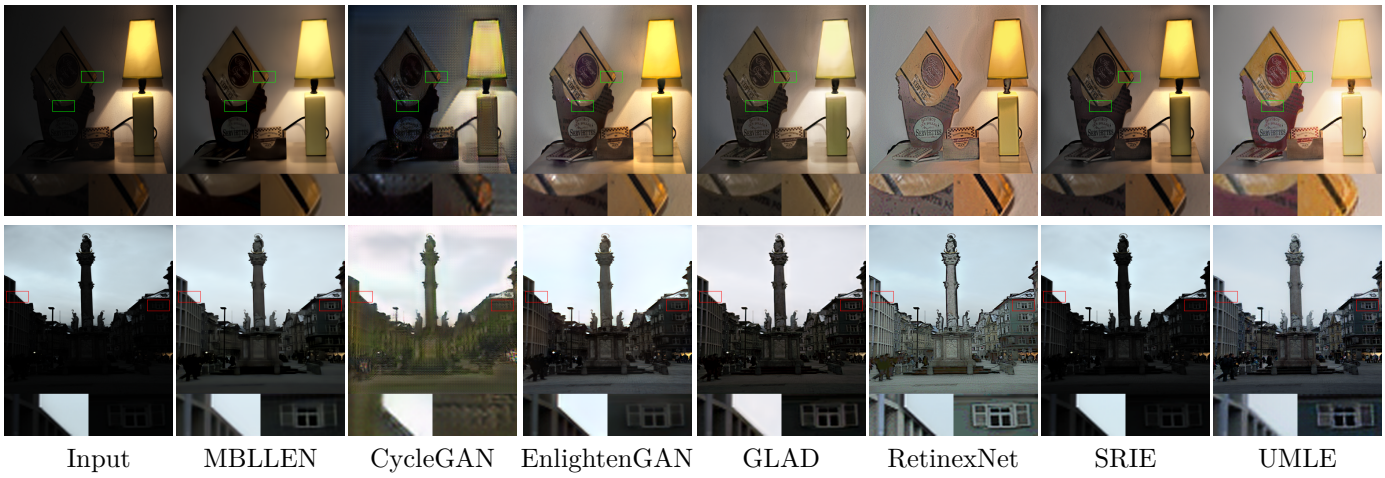


Fig. 5. The visual comparison with state-of-the-art low-light image enhancement approaches. Boxes indicate the obvious differences. It is observed that our results contain little noise while obtaining the most natural color of the image.

In the experiments, we empirically set ω_1=1, ω_2=100, ω_3=0.005, ω_4=0.005, ω_1=10.

III. Experiments

In this section, we make a comparison between the performance of our method with state-of-the-art image enhancement methods to demonstrate the validity of our proposed method including experiments (A)−(F). Besides, we demonstrate the utility of our approach by applying our method to SLAM relocation and drivable area detection, as part of the experiments, including (G), (H).

A. Dataset

To fully assess our approach, we tested various scenarios, including indoor and outdoor scenes. Since our model does not need paired datasets for training, we randomly selected images with different scenes and different illumination levels from the available datasets [10], [20], and [18] as our datasets. The dataset includes 750 normal-light images and 923 low-light images.

For the relocalization test, we used the dataset published by ETH [21].

B. Implementation details

We used ReLU [22] as the activation function. The generator and discriminator adopt AdamGC [23] optimizer. We also used a weight decay at the rate of 0.0001. For the normalization layer, the generator uses Adalin [24]. Due to memory constraints, we set the batch size to 1 for our experiments. The training process of the model was carried out on an Nvidia P40 while using only 50K iterations.

C. Qualitative Comparison

We conduct extensive qualitative evaluations on typical low-light images with several state-of-art methods. The results are illustrated in Fig. 5. with SRIE [7], Retinex-Net [10], EnlightenGAN [18], MBLLEN [25], Cyclegan [26], GLAD [27]. In terms of color and brightness, Enlightengan [18] is prone to color deviation and the result of SRIE [7] is likely to be darker. For texture, LIME [6] creates more noise while enhancing. The problem of fuzzy details exists in the results of cyclegan [26]. Retinex-Net [10] produces a kind of effect similar to crayon drawing, influencing the quality of generation. In contrast, our method performs well based on texture and color.

TABLE I
Quantitative evaluation of low-light image enhancement algorithms.

Model	NIQE(↓)	ENTROPY(↑)
Retinex-Net	7.539	6.923
MBLLEN	4.481	7.050
GLADNet	3.254	7.340
CycleGAN	4.260	7.092
EnlightenGAN	3.572	7.350
SRIE	4.659	6.874
UMLE	3.485	7.506

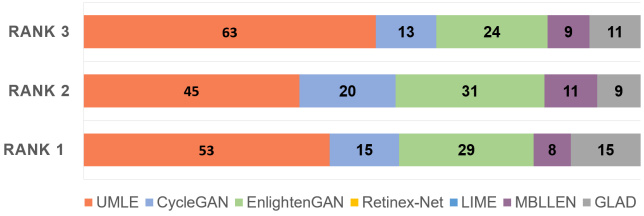


Fig. 6. The user study score on the test dataset, the number on the picture means the distribution of the images considered as the best by the user for each rank.

TABLE II

The number of parameters for generator, training iterations and testing FPS of each model.

Module / Model	Parameters	Training iterations	FPS
EnlightenGAN	22.30M	327k	52
CycleGAN	11.37M	327k	45
ULEN(Our)	16.19M	50k	65

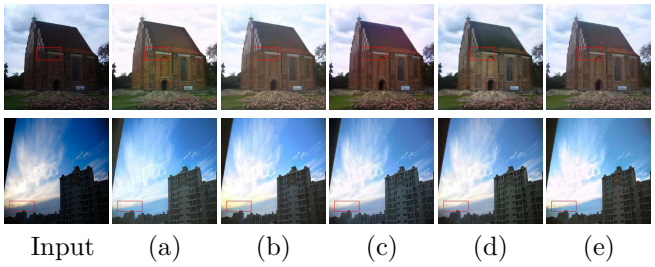


Fig. 7. Visual comparison with ablation results, boxes indicate the obvious differences. (a) is the result without color discriminator, (b) is the result without texture discriminator, (c) is the result without multi-scale discriminator, (d) is the result without CPAF module, (e) is the result of UMLE.

D. User study

To compare our method's performance and other methods we adopted a human subjective evaluation [18], a human subjective study. This approach is similar to the double stimulus continuous scale. It assesses the image quality from the following three indicators:

Rank1: How much visible noise the images contain;

Rank2: How much over or underexposure artifacts the images contain;

Rank3: Whether there are unrealistic textures or colors in the image.

To assess the results we randomly selected 120 volunteers. For each metric, the distribution of the best performing images in each algorithm is represented in Fig. 6. The number of images performed by each method best for each category is included in the figure.

Analysis of these results indicates that our methods performed best in all three metrics.

TABLE III

The result of the ablation study.

Condition	NIQE(↓)	ENTROPY(↑)
with D_C, w/o D_T, w/o D_M	5.416	7.083
with D_T, w/o D_C, w/o D_M	5.656	6.910
with D_M, w/o D_T, w/o D_C	4.840	6.743
with D_T, with D_M, w/o D_C	4.589	7.132
with D_C, with D_M, w/o D_T	3.789	7.291
with D_T, with D_C, w/o D_M	3.929	6.992
with D_M, D_T, D_C, w/o $CPAF$	6.625	7.067
default configuration	3.485	7.506

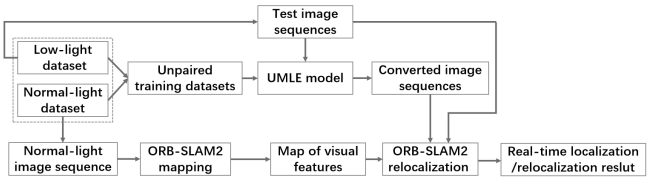


Fig. 8. Flowchart of our experiment. We use the normal illumination image for mapping while training the UMLE model using the unpaired flashlight illumination image and the normal illumination image. We then test the relocalization on the normal illumination build map using the enhanced image and the original flashlight illumination image, respectively.

E. No-reference evaluation

To evaluate the perceptual quality of the state-of-art methods, we employed the naturalness image quality evaluator (NIQE) [28] and entropy [29]. The NIQE is based on the construction of a series of features utilized to measure image quality, and lower NIQE [28] values reflect higher quality. The entropy reflects the amount of information carried by the image. The greater the information entropy of the image means the better the quality. Tab. I represents the calculated results of our NIQE [28] and entropy [29] indices. The red result reveals the best, blue shows the second best, and bold black represents the third-best result. According to the table we can observe that the proposed UMLE shows good performance in terms of NIQE [28] and entropy [29], and it further demonstrates the advantages of our model in generating normal-light images.

Besides, we compared current unsupervised models from the number of parameters, model training perspective and runtime. Tab. II represents the number of training iterations, generator parameters, and frames per second (FPS) in testing for EnlightenGAN [18], CylceGAN [26], and our method UMLE. All parameters were adopted based on the original article. Simultaneously, we converted the epoch provided to iteration, and parameters mean generator parameters. Iterations indicate the training iterations and frames per second. Bold black shows the best results. It is observed that our method has significant advantages in training time, the total number of parameters, and testing FPS.

F. Ablation study

To verify the effectiveness of each component proposed in Sec.II, we performed several ablation studies.

We tested the results separately in the absence of each discriminator. In Tab. III, we calculated the NIQE [28] values in each case and verified each component's importance experimentally.The results reveal that each module substantially contributes to enhancing the image quality.

In Fig. 7, the results demonstrate the importance of each component with images, and the data in the table reflect each condition's performance, where w means without. Through the analysis of the image, it is

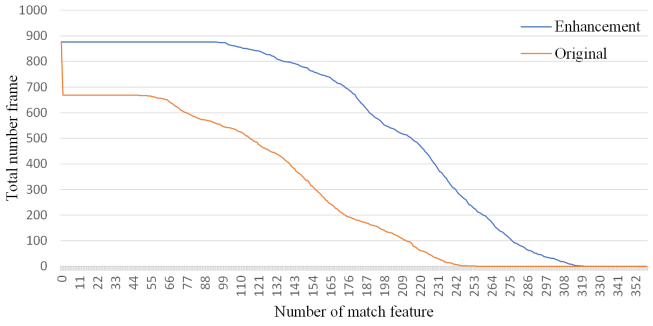


Fig. 9. The comparison of the number of successfully matched feature points before and after image enhancement. The horizontal coordinate represents the number of successful matches, and the vertical coordinate shows the number of images with more features than the corresponding horizontal coordinate. From the image, it is observed that the number of features matched in the enhanced image is increased significantly.

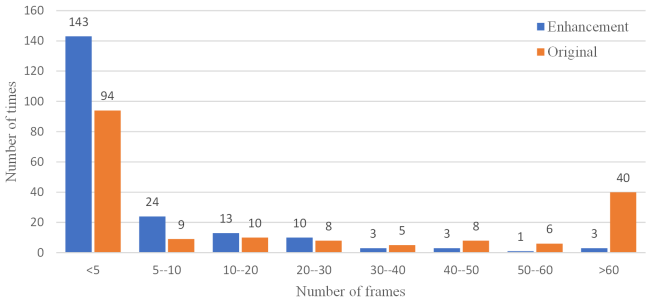


Fig. 10. The number of images required for successful repositioning. The meaning of the horizontal coordinate in the image is the number of images required for successful localization and the meaning of the vertical coordinate is the number of groups of tested sequences in this interval. When the number of images required for successful localization is greater than 60, we considered that the localization failed. According to the figure, the image sequences enhanced by UMLE have a great improvement in both localization success and speed.

indicated that the lack of a texture discriminator and a color discriminator has a great effect on the color and texture. A multi-scale discriminator has a great influence on the whole image generation result. The complete model performs well in visual effect, moreover, the importance of each component is demonstrated by the results

G. Localization application

Furthermore, we utilized the proposed algorithm in a visual map localization application in scenarios with drastic illumination changes (day→night or light on→off). In the experimental process, we use a publicly available SLAM dataset [21], which is generated by the simulation environment and contains two sets of data for normal and flashlight illumination in the same scene.

To verify that our method can still re-localize under the drastic illumination changes, we designed the next experiments, the flowchart of which is represented in Fig. 8. First, we trained the UMLE network with two kinds of images, normal-light and flashlight. Meanwhile, continuous image sequences under normal illumination are extracted to construct ORB feature maps by using ORB-SLAM2 [30] .

For the pre-constructed feature maps under normal light in the localization process, we extracted flashlight image sequences at five different locations and tried to re-localize them. Our localization method is run utilizing the relocalization module in ORBSLAM2 [30].

Fig. 9 reflects the number of feature points before and after image enhancement. The number of images is the same. They all start from the same starting point at first. There is a steep drop at the beginning since the number of successful feature points for many of the original image matches is 0. It is obverted that the number of features matched in the enhanced image is significantly higher compared to the original image.

We randomly selected 200 image sequences, each containing 60 images, from the flashlight dataset and the enhancement dataset to test the speed and success of relocalization of these images. The speed of image repositioning is reflected in Fig. 10. We considered a successful match within five frames as immediate relocation and a failed relocation above 60 frames.

According to Fig. 10, the success rate of immediate relocalization of image sequences was enhanced with UMLE increased from 47% to 71.5%, moreover, the relocalization failure rate decreased from 20% to 0.15%. This experiment verifies that the success rate and relocalization speed of enhanced images are improved significantly under our model.

Based on the experimental results in the images, our proposed method can improve significantly the success rate of image matching for the same scene in the case of drastic lighting changes. Our method presents an effective solution to the visual localization problem in scenes with significant lighting changes and has important practical value.

IV. Conclusion

In this paper, we proposed an unsupervised enhancement model for real-time low-light image enhancement. The model's training is independent of paired training data. Hence, it can use images of different scenes and different illuminations. Besides, we presented a multi-branch discriminator to comprehensively assess the image from color, texture, and global information. We also introduced a novel attention module integrating channel attention and pixel attention to focus more on the low-light areas. Both quantitative and qualitative experimental evaluations reveal that our network achieves good results in terms of visual effects and noise control. Moreover, we experimentally validated that our method is significantly more effective for tasks such as SLAM repositioning drivable area detection.

References

[1] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, and K. Zuiderveld, "Adaptive histogram equalization and its variations," Computer Vision Graphics & Image Processing, vol. 39, no. 3, pp. 355–368, 1987.

[2] H. Ibrahim and N. S. P. Kong, "Brightness preserving dynamic histogram equalization for image contrast enhancement," IEEE Transactions on Consumer Electronics, vol. 53, no. 4, pp. 1752–1758, 2008.

[3] C. Lee, C. Lee, and C. Kim, "Contrast enhancement based on layered difference representation of 2d histograms," IEEE Trans. Image Process., vol. 22, no. 12, pp. 5372–5384, 2013.

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

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

[6] X. Guo, "LIME: A method for low-light image enhancement," in Proceedings of the 2016 ACM Conference on Multimedia Conference, A. Hanjalic, C. Snoek, M. Worring, D. C. A. Bulterman, B. Huet, A. Kelliher, Y. Kompatsiaris, and J. Li, Eds., 2016, pp. 87–91.

[7] K. Nakai, Y. Hoshi, and A. Taguchi, "Color image contrast enhacement method based on differential intensity/saturation gray-levels histograms," in International Symposium on Intelligent Signal Processing and Communication Systems, 2013, pp. 445–449.

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

[9] 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.

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

[11] 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.

[12] 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.

[13] Y. Qu, Y. Ou, and R. Xiong, "Low illumination enhancement for object detection in self-driving," in 2019 IEEE International Conference on Robotics and Biomimetics (ROBIO). IEEE, 2019, pp. 1738–1743.

[14] R. Chen, W. Huang, B. Huang, F. Sun, and B. Fang, "Reusing discriminators for encoding: Towards unsupervised image-to-image translation," in 2020 IEEE/CVF Conference on

[19] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in IEEE International Conference on Computer Vision, ICCV. IEEE Computer Society, 2017, pp. 2813–2821.

Computer Vision and Pattern Recognition, 2020, pp. 8165–8174.

[15] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 8, pp. 2011–2023, 2020.

[16] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "Eca-net: Efficient channel attention for deep convolutional neural networks," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR. IEEE, 2020, pp. 11 531–11 539.

[17] A. Anoosheh, T. Sattler, R. Timofte, M. Pollefeys, and L. Van Gool, "Night-to-day image translation for retrieval-based localization," in 2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019, pp. 5958–5964.

[18] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, "Enlightengan: Deep light enhancement without paired supervision," CoRR, vol. abs/1906.06972, 2019.

[20] N. K. Kalantari and R. Ramamoorthi, "Deep high dynamic range imaging of dynamic scenes," ACM Trans. Graph., vol. 36, no. 4, pp. 144:1–144:12, 2017.

[21] S. Park, T. Schöps, and M. Pollefeys, "Illumination change robustness in direct visual slam," in ICRA, 2017.

[22] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, vol. 15, 2011, pp. 315–323.

[23] H. Yong, J. Huang, X. Hua, and L. Zhang, "Gradient centralization: A new optimization technique for deep neural networks," CoRR, vol. abs/2004.01461, 2020.

[24] J. Kim, M. Kim, H. Kang, and K. Lee, "U-gat-it: unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation," arXiv preprint arXiv:1907.10830, 2019.

[25] F. Lv, F. Lu, J. Wu, and C. Lim, "MBLLEN: low-light image/video enhancement using cnns," in British Machine Vision Conference 2018, 2018, p. 220.

[26] J.-Y. Zhu, T. Park, P. Isola, and A. 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.

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

[28] A. Mittal, Fellow, IEEE, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," IEEE Signal Processing Letters, vol. 20, no. 3, pp. 209–212, 2013.

[29] J. R. A, "Focus optimization criteria for computer image processing [j]," in Microscope, 1976, pp. 163–180.

[30] R. Mur Artal and J. D. Tardós, "ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras," IEEE Transactions on Robotics, vol. 33, no. 5, pp. 1255–1262, 2017.