N-LoLiGan: Unsupervised Low-Light Enhancement GAN with an N-Net for Tunnel Low-Light Images

Jing Wang[1], Ganghua Xiao[1], Ziming Xiong[2], Hongliang Zhu[1], Jinqiao Cui[1], Yuzhuang Wan[1], Zhengfang Wang*[1], Lei Jia[1], and Qingmei Sui[1]

***Abstract*—Owing to the low luminance in tunnels, target features in images are not salient, which makes tunnel surface defect detection challenging. Most deep-learning-based low-light enhancement methods suffer from contrast distortion and insufficient brightness in extremely dark scenes, as well as artifacts and overexposure in scenes with large light and dark differences. To alleviate these problems, in this letter, we propose a novel unsupervised generative adversarial network (GAN), called the N-shape low-light enhancement GAN (N-LoLiGan), which can be trained using unpaired low-/normal-light images and is proven to generalize very well on tunnel images with complex lighting. In the generator, we utilize an N-Net network to balance the proportion of luminance attention information and input image channel information during the coding process through a multi-scale input layer, which allows the network to adjust both brightness and contrast levels, thereby making enhanced images appear more natural. In the discriminator, we utilize an adaptive global-local illumination mask dual-discriminator to match the illumination of the enhanced image to that of a normal image. Experimental results demonstrate that the proposed N-LoLiGan outperforms several state-of-the-art low-light image-enhancement methods in terms of visual quality and two no-reference image quality metrics. Additionally, when the proposed method is used as a preprocessing step for the tunnel surface defect detection task, it yields a higher detection rate (88.9 %) than other methods.**

***Index Terms*—low-light enhancement, unsupervised learning, tunnel defect detection.**

I. INTRODUCTION

THE safe construction and operation of tunnels are crucial. Based on geological, climatic, and other environmental factors, tunnels are subject to the occurrence of cracks and other tunnel surface defects [1]. The random growth and internal extension of cracks can lead to accidents, such as tunnel lining peeling, water seepage, and even tunnel collapse [2]. Therefore, routine tunnel inspection and health assessment are essential for ensuring the safe construction and operation of tunnels. Currently, tunnel crack detection depends primarily on manual identification and recording, and detection results are significantly affected by human subjective factors; therefore, existing methods cannot meet the standards of modern crack detection [3]. With the continuous development of computer vision technology, the nondestructive testing of traffic equipment based on image acquisition equipment and digital image processing technology has become mainstream. However, owing to the complex illumination conditions in tunnels, captured images may suffer from uneven illumination and severe local exposure, which negatively affects the recognition of lining cracks and increases the difficulty in detecting damage in later stages [4, 5]. To address this issue, one effective solution is to process low-light images using appropriate low-light enhancement algorithms, which have been successfully applied in tunnel safety monitoring, night road detection, and other areas.

Traditional methods for real-world low-light image enhancement primarily focus on adjusting brightness or contrast through fixed-tone mapping [6], thereby resulting in limited performance in challenging cases. Recently, learning-based methods have been applied to learn content-aware illumination or contrast enhancements from data using deep neural networks [7, 8]. Despite achieving satisfactory performance, such methods rely on pairs of low-light and corresponding normal-light images, which are expensive or impossible to obtain in real-world scenarios.

To eliminate reliance on paired data, several unsupervised deep-learning-based methods have been developed for image enhancement. Generative adversarial network (GAN)-based models have been designed to address illumination enhancement [9–12]. These unsupervised learning approaches can generate images with better illumination and contrast in certain cases. However, they have two common limitations in real-world low-light image enhancement tasks: 1) the contrast and illumination of enhanced images may be unsatisfactory and typically suffer from color distortion and inconsistency, and 2) artifacts and overexposure may occur in enhanced images in extremely dark scenes or scenes with light and dark differences.

This work was supported in part by the National Natural Science Foundation of China under Grant 41877230 and 61873333, in part by the Shandong Provincial Key Research and Development Program (SPKR&DP) under Grant 2019GGX101027, and in part by the Natural Science Foundation of Guangdong Province under Grant 2021A1515011782. (Corresponding author: Zhengfang Wang.)

The Authors 1 are with the School of Control Science and Engineering, Shandong University (e-mail: wangjingkz@sdu.edu.cn; xgh@mail.sdu.edu.cn; hlzhu@mail.sdu.edu.cn; 1412263612@qq.com; 1103385964@qq.com; wangzhengfangsdu@hotmail.com; jialei@sdu.edu.cn; qmsui@sdu.edu.cn).

The Authors 2 is with National Defense Engineering Academy, PLA Army Engineering University (e-mail: xzm992311@163.com).

Preprint submitted to IEEE Robotics and Automation Letters as Submission for RA-L only
Received August 9, 2022 06:16:21 Pacific Time

To address these issues, we proposed a novel unsupervised N-shape low-light enhancement GAN (N-LoLiGan). In this method, we adopted an N-Net network as the generator for a GAN. Through a multi-scale input layer, an attention diagram was fused with the sampling results of each layer. Therefore, the ability to leverage image channel information was improved, which not only solves the problems of overexposure in complex scenes, but also allows enhanced images to retain rich color information. Additionally, we designed a dual-discriminator with a global mask discriminator and adaptive local illumination discriminator to balance global and local low-light enhancement. The proposed N-LoLiGan was tested on tunnel images and evaluated on a tunnel surface defect detection task. The results indicated that our method provides a promising image-enhancement approach for real-world tunnel scenarios.

The main contributions of this study can summarized as follows.

1) We proposed an unsupervised end-to-end deep learning framework for tunnel low-light image enhancement. An N-Net network was designed to improve our model's ability to leverage image channel information. A global mask discriminator was designed to determine the number of local discriminator blocks, which enabled the network to realize adaptive dark enhancement and brightness suppression.

2) N-LoLiGan was compared to several state-of-the-art methods through comprehensive experiments. Results were derived through visual quality assessment and unreferenced image quality assessment. All results consistently confirmed the superiority of N-LoLiGan. Additionally, we tested N-LoLiGan without domain adaptation as a preprocessing step for a tunnel image dataset. The results demonstrated that the proposed algorithm can effectively improve the crack detection rate, even in a complex light-dark scenario.

II. RELATED WORK

Low-light image enhancement has been an important subject of study. This section briefly reviews the different methods used in this field.

A. Traditional Approaches

Traditional approaches can be divided into the following two categories. The first branch consists of histogram equalization [13] and its variants, such as AHE [14] and BPDHE [15], which allow constraining the histograms of output images to achieve enhanced results. The second branch is grounded in retinal theory [16], which states that the brightness of an image is determined by its reflectance and illumination. In the past, numerous methods, such as MSR [17], SSR [18], and AMSR [19], have leveraged this principle to retrieve and utilize illumination maps for low-light image enhancement. Recently, numerous methods have focused on the simultaneous adjustment of reflection and illumination maps to obtain improved enhancement results. Typical methods include NPE [20], SRIE [21], LIME [22], and BIMEF [23], which adjust enhancement results using different constraint models to achieve satisfactory performance. However, prior-based strategies do not adjust well to diverse scenarios, despite the fact that priors are typically obtained through numerous statistical analyses [24].

B. Supervised Learning Approaches

Learning-based models aim to learn the statistical distributions of reflection and illumination maps from a dataset and use these learned distributions to improve the generalization capabilities of retinex-based algorithms. LLNet [25] was the first seminal deep-learning-based low-light image enhancement method that applied stacked sparse denoising autoencoders to enhance low-light images. RetinexNet [26] utilizes a learning-based deep network that contains a decom-net for decomposing illumination maps and an enhanced net for adjusting illumination maps. Zhang et al. [27] proposed a maximum-entropy-based retinex model that uses a simple network to separate illumination and reflection for achieving self-supervised learning. Huang et al. [28] designed an attentional mechanism to predict illumination graphs and used the retinex model to estimate initial enhanced images.

However, these methods require a considerable amount of paired data for training, which are difficult to collect in numerous cases.

C. Unsupervised Learning Approaches

Over the past decade, GAN-based methods have gained significant popularity for image enhancement applications [29]. Chen et al. [30] considered the image enhancement problem as an image-to-image conversion problem and proposed a bidirectional GAN to enhance low-light images using unpaired images. EnlightenGAN [12] utilizes a U-Net with self-feature retention loss, self-regular attention mechanism, and global-local discriminator to realize unsupervised low-light image enhancement. The bio-inspired two-path network model [31] adopts a bionic dual-path network model with a fusion method for high- and low-frequency channels to realize image brightening and denoising. RetinexGAN [16] applies the retinex theory to a GAN network to implement a decomposition network consisting of a two-layer convolutional neural network to extract illumination and reflection images and achieve low-light image enhancement. LVENet [32] utilizes the retinex theory to estimate illumination components using a lightweight depth-wise separable convolutional layer. LiCENet [33] decomposes brightness components from the HSL color space and applies autoencoders to train weak light intensifiers to improve the brightness components.

These unsupervised methods can achieve good results for a variety of datasets. Regardless, illumination in tunnels is particularly weak and uneven, which imposes greater challenges on the enhancement of low-light images, typically leading to overexposure or artifacts. However, a low-light image enhancement method designed specifically for tunnels has not yet been studied.

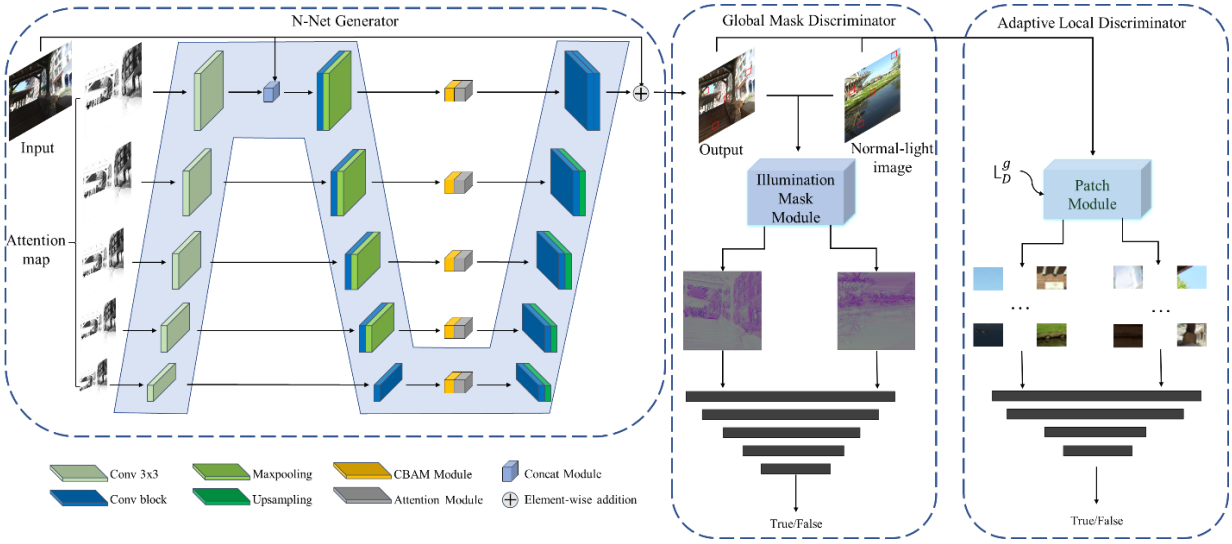

Fig. 1. Overall architecture of N-LoLiGan. The N-Net generator consists of a multi-scale input layer, encoder path, and decoder path. Each attention module has a convolutional block attention module (CBAM) feature map multiplied by a (resized) attention map. The discriminator consists of a global illumination mask discriminator and adaptive local discriminator. The illumination mask module is designed such that the global discriminator focuses on the illumination of the enhanced image rather than all information, similar to a normal-light image. Global discriminator loss is used to guide the number of local discriminator block.

III. Methods

As discussed above, owing to the dark and uneven illumination in tunnels, enhancing the dark regions while suppressing the bright regions is difficult. Therefore, enhanced images often appear with artifacts or whitening problems, thereby resulting in blurred features in enhanced images.

To solve this problem, we developed N-LoLiGan. As shown in Fig. 1, N-LoLiGan adopts an N-Net as a generator and uses an adaptive global-local illumination mask dual-discriminator to match enhanced images to normal images. Additionally, we used feature loss to guide the training process and preserve texture and structure. The details of N-LoLiGan are discussed below.

A. N-Net Generator

The proposed N-Net is an end-to-end deep network consisting of three main components. The first is a multi-scale input layer, which is used to construct an image pyramid input and achieve multi-level receptive field fusion. The second is the encoder path, which encodes the size features of input images into a channel to analyze image information and improve the efficiency of the subsequent reconstruction step. The last component is the decoder path. A bilinear upsampling layer is used to upsample low-level features, and each layer of encoder features and resized attention guidance maps is used to guide the convolution calculations of the pixels of low-level features, thereby achieving the reconstruction of images with suitable illumination.

The implementation details of each component are summarized below.

1) Multi-Scale Input Layer

A multi-scale input or image pyramid has been demonstrated to improve the ability of networks to utilize global information effectively [34]. To enhance dark regions more than bright regions, we normalized the illumination channel I of an input RGB image in a range of [0, 1] and subsequently used (1 – I) as an attention map. The N-Net employs a max-pooling layer to downsample the attention map and construct a multi-scale input using a 3 × 3 convolution layer in the encoder path. The ratio of the number of channels outputted by each multi-scale input layer to the number of channels in the corresponding encoder path was 1:3. Such a multi-scale input layer ensured the full use of the global information in the illumination map and input image, which is critical for avoiding the whitening and contrast distortion of low-light enhanced images, as will be demonstrated by our experimental results.

2) Encoder Path

The encoder path comprised the repeated application of two 3 × 3 convolution layers, each of which was followed by a LeakyReLu activation function, batch normalization layer, and 2 × 2 max pooling operation with a stride of two for downsampling. At each downsampling step, the number of feature channels was doubled.

3) Decoder Path

Every step in the decoder path comprised a bilinear upsampling layer that halved the number of feature channels, followed by two 3 × 3 convolution layers, a LeakyReLu activation function, and batch normalization layer. Skip connections utilized the CBAM [35] module and an attention module, which is key to preserving rich texture information and synthesizing normal-light images using multi-scale attention map information. The CBAM was used to boost the perceived power of the channel information from the encoder path. The attention module multiplied the CBAM output by the resized attention map and concatenated it with the upsampled decoder feature maps.

B. Adaptive Global-Local Discriminators

In this study, our network model was trained using unpaired images. To this end, we adopted adversarial learning to ensure that the distribution of enhanced images was close to that of normal-light images.

To ensure that the discriminator focuses on image brightness, a global illumination mask discriminator was designed to generate output and normal-light image illumination masks; subsequently, the PatchGan method was applied to identify true and false illumination masks. As shown in Fig. 2, the illumination mask module structure consists of multi-layer sparse convolution and skip connections, which integrate deep and shallow image features through skip connections to extract deep brightness information while retaining certain shallow features. In addition to the global illumination mask discriminator, an adaptive local discriminator was designed. Moreover, we determined the proportion of randomly cropped local patches from both the output and real normal-light images according to the results of the global illumination mask discriminator and utilized PatchGan for true and false discrimination.

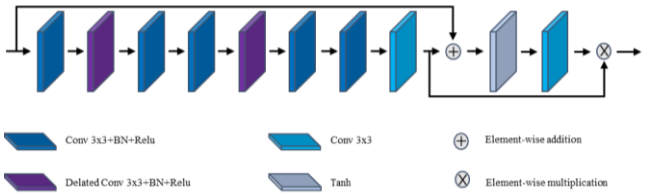

Fig. 2. Structure of the illumination mask module consisting of multi-layer sparse convolution and skip connections, which integrate deep and shallow image features through skip connections to extract deep brightness information while retaining certain shallow features.

This adaptive global-local structure ensured that all local patches in an enhanced image maintained the appearance of real normal light, even in environments with large differences between light and dark, which is critical for avoiding local over-exposure and artifacts.

C. Loss Function

Loss function is an important aspect of improving network performance to enhance images under specific illuminance conditions. The loss function of the network developed in this study is composed of adversarial loss, feature loss, and quality-perceived loss.

1) Adversarial Loss

Adversarial loss is used to ensure that $x_f \sim x_r$, the distribution of a low-illumination map, is approximately equal to that of a normal image.

We adopted the LSGAN [36] version of the relativistic average GAN loss for training our global discriminator. When updating the discriminator, we have

$$\begin{aligned} \mathrm{L}_D^g &= E_{x_r \in R}[(D_g(x_r) - \mathrm{E}_{x_f \in F} D_g(x_f) - 1)^2] \\ &+ E_{x_f \in \mathrm{F}}[(D_g(x_f) - \mathrm{E}_{x_r \in R} D_g(x_r))^2] \end{aligned} \tag{1}$$

When updating the generator, we have

$$\begin{aligned} \mathrm{L}_G^g &= E_{x_r \in R}[(D_g(x_r) - \mathrm{E}_{x_f \in F} D_g(x_f))^2] \\ &+ E_{x_f \in \mathrm{F}}[(D_g(x_f) - \mathrm{E}_{x_r \in R} D_g(x_r) - 1)^2] \end{aligned} \tag{2}$$

where R and F are the real and generated image distributions, respectively; and x_r and x_f are samples from the distributions R and F, respectively.

We adopted the LSGAN loss function for training the local discriminator.

When updating the discriminator, we have

$$\mathrm{L}_D^l = E_{x_r \in R}[(D_l(x_r) - 1)^2] + E_{x_f \in \mathrm{F}}[(D_l(x_f))^2] \tag{3}$$

When updating the generator, we have

$$\mathrm{L}_\mathrm{G}^l = E_{x_f \in F}[(D_l(x_f) - 1)^2] \tag{4}$$

2) Feature Loss

Similar to [18, 37, 38], we used the feature loss (computed on VGG features) between enhanced images and input images to preserve the content details from input images. To prevent the influence of image brightness on feature loss, we forced the network to focus only on content by performing instance normalization on the VGG feature maps before calculating the feature loss. Similar to adversarial loss, we calculated the feature loss on both the global and local images. We formulated the global feature loss L_F^g and local feature loss L_F^l as

$$\mathrm{L}_F^g = \frac{1}{WHC} \| \psi(R) - \psi(F) \|_2^2 \tag{5}$$

$$\mathrm{L}_F^l = \frac{1}{WHC} \| \psi(R_l) - \psi(F_l) \|_2^2 \tag{6}$$

where ψ denotes the VGG feature extractor; H, W, and C are the dimensions of the extracted feature maps; R and F are the real and generated images, respectively; and R_l and F_l are the random patches extracted from R and F, respectively.

3) Quality-Perceived Loss

Because obtaining paired images with low light and normal light in real-world applications is difficult, the loss functions used in existing supervised methods, such as peak signal-to-noise ratio, structural similarity, and mean-squared error, are not suitable for this application.

The natural image quality evaluator (NIQE) [39] assesses image quality using a distance metric between a model's statistics and those of an enhanced image without knowledge regarding anticipated distortions or human biases. In this study, we utilized the NIQE to design a novel loss function. If an image-enhancement method is effective, then the NIQE values of an enhanced image will be lower than those of an original image. Therefore, quality-perceived loss can be expressed by the ratio of the NIQE of an enhanced image to that of an original images. We formulate the quality-perceived loss L_Q as

$$\mathrm{L}_Q = \frac{NIQE(\mathrm{F})}{NIQE(X)} \tag{7}$$

where X and F denote the original image and enhanced image, respectively.

Therefore, the overall loss function for training N-LoLiGan is defined as

$$L = L_D^g + L_G^g + L_D^l + L_G^l + L_F^g + L_F^l + L_Q \quad (8)$$

IV. EXPERIMENTS

In this section, we present quantitative and qualitative experiments that were conducted to evaluate the performance of N-LoLiGan on our collected unpaired tunnel low-light dataset and to apply it to low-light tunnel surface defect detection. Additionally, we present an ablation study on the multi-scale input layer, CBAM attention module, illumination mask module, and adaptive local discriminator to validate the effectiveness of our model comprehensively.

A. Datasets and implementation

Unpaired training dataset: We used unpaired low-/normal-light images for training and the training dataset was provided by EnlightenGAN [12]. It contained 914 low-light images and 1016 normal-light images from public datasets. These images were converted into PNG format and adjusted to a resolution of 600 × 400 pixels. The training set was further augmented by randomly flipping images horizontally and vertically, as well as by rotating by 90°.

Unpaired tunnel low-light dataset: There are a limited number of publicly available low-light tunnel datasets. Therefore, we used an unpaired low-light tunnel dataset as a testing dataset. First, we captured low-exposure compensation photos using an iPhone-13p from various areas in different tunnels. We removed images that were excessively dark; these could not be enhanced as too many details were absent. The final dataset was composed of 173 low-light tunnel images. Examples of low-light tunnel images are presented in Fig. 3. This testing dataset includes various scenes, such as highway tunnels (Xihuan Tunnel, Fig. 3(a); Qingshiguan Tunnel, (Fig. 3(b)), railway tunnels (Shuohuang Tunnel, (Fig. 3(c)), and tunnels under construction (Xihuan Tunnel, Fig. 3(d)). Note that there are no corresponding ground-truth images any testing images. Each low-light image was resized to 600 × 400 pixels.

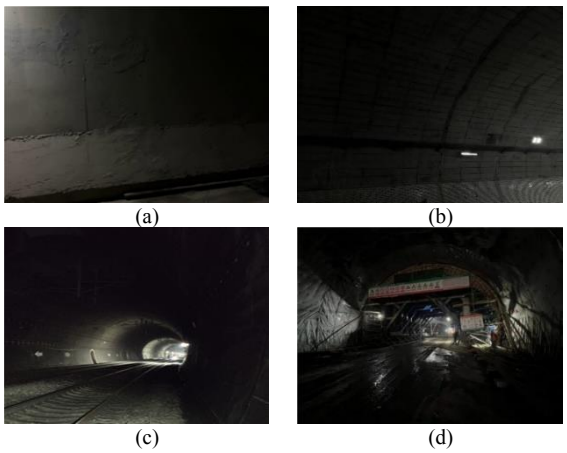

Fig. 3. Example tunnel low-light images in the testing dataset. Images (a) and (b) were captured in the Xihuan Tunnel and Qingshiguan Tunnel, respectively. Image (c) was captured in the Shuohuang Tunnel and image (d) was captured in the Xihuan Tunnel.

All experiments were conducted on a Linux workstation with three NVIDIA GTX 1080 Ti GPUs. N-LoLiGan was implemented using Pytorch. The network was optimized using the Adam optimizer with a learning rate of 10^{-4} for the first 100 epochs and a weight decay of 0.001 over the next 100 epochs. We set β_1=0.5 and β_2=0.999 , and the batch size was set to 16 for training. The total number of training epochs was set to 200 and the training time was approximately 4 h.

B. Performance Comparisons

We performed performance comparisons considering several representative state-of-the-art low-light enhancement methods. SRIE [21] is a conventional method based on retinex models. RetinexNet [26] is similar to convolutional-neural-network-based methods, but uses the retinex theory to obtain enhanced outputs. EnlightenGAN [12] was the first network to introduce successful unpaired training for low-light image enhancement.

To ensure fairness in our comparisons, we used the source code and pre-trained models provided by the original authors of these methods and used the parameter settings recommended by the authors. All the methods were trained without domain adaptation and tested on the same tunnel low-light dataset.

1) Quantitative Comparisons

We compared the quantitative results (lightness order error (LOE) and NIQE) of the proposed N-LoLiGan and the other three benchmark methods. LOE reflects the natural light retention ability of an image. The better the brightness order of an image, the more natural it appears. NIQE evaluates image quality by calculating differences in the multivariate distributions between the features of an image to be measured and those of a natural image. The smaller the value, the better the quality of an image. As listed in Table 1, LOE is computed by comparing an enhanced image to an original image. Therefore, there are no LOE values for the original images. For the testing set, the LOE and NIQE values of N-LoLiGan were 225.57 and 11.57, respectively, which were smaller than those of the other three methods, thereby demonstrating that it outperforms the other three methods in terms of quantitative measures.

TABLE I
AVERAGE LOE AND NIQE VALUES OF THE FOUR COMPARED METHODS OVER 173 TUNNEL LOW-LIGHT IMAGES IN THE TESTING DATASET (A SMALLER LOE AND NIQE INDICATE MORE PERCEPTUALLY FAVORED QUALITY)

METHOD	LOE	NIQE
ORIGINAL IMAGES		13.91
SRIE	295.55	12.93
RETINEXNET	991.44	13.26
ENLIGHTENGAN	344.01	11.83
N-LoLiGan	**225.57**	**11.57**

2) Visual Comparisons

To validate the superiority of our method further, we compared the results of N-LoLiGan to those of the three other benchmarks by visualizing output images. In Fig. 4, the first column presents original low-light images and the second to fourth columns present images enhanced by SRIE

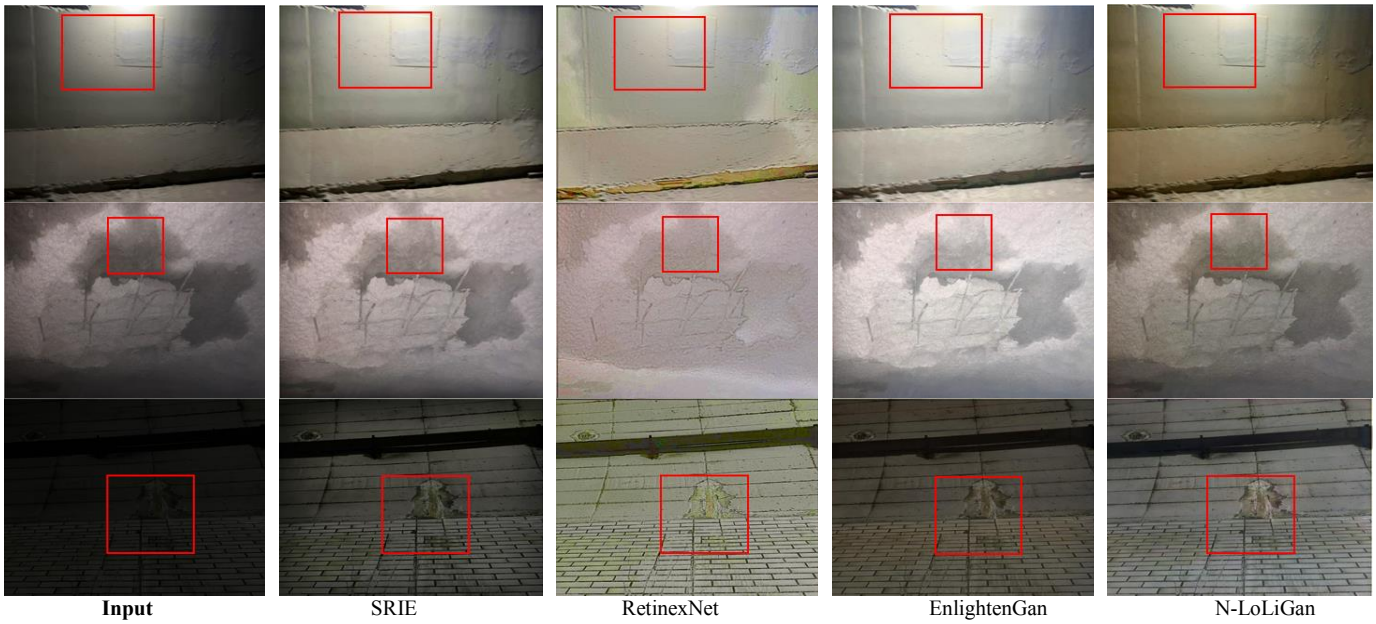

Fig. 4. Comparisons to other state-of-the-art methods. Three examples are provided from the top to the bottom rows. First example: SRIE and EnlightenGAN introduce overexposure and blur the texture features of wall details. RetinexNet generates severe color distortion. However, N-LoLiGan not only restores the wall details, but also avoids overexposure artifacts, distinctly outperforming the other methods. Second example: SRIE and EnlightenGAN generate washed-out images. RetinexNet blurs the feature information of leakage. The N-LoLiGan results retain rich leakage information by mining deeper color information from input images. Last example: N-LoLiGan produces the brightest results while retaining rich feature information. Other models either do not enhance dark details sufficiently or generate artifacts. We encourage readers to zoom in to see additional details.

Fig. 5. Visual comparisons from the ablation study of N-LoLiGan. Column 1 to 5 present low-light image inputs, results from N-LoLiGan without the multi-scale input layer, results from N-LoLiGan without the CBAM attention module, results from N-LoLiGan with only the global-local dual-discriminator, and results from the final version of N-LoLiGan, respectively. The images in columns 2–4 suffer from severe overexposure or color distortion, which are highlighted by bounding boxes. The full version of N-LoLiGan mitigates these issues and gains the most visually pleasing results. We encourage readers to zoom in to see additional details.

[21], RetinexNet [26], and EnlightenGAN [12] in order. The final column presents the results produced by N-LoLiGan.

The visualization results revealed that the results of SRIE were generally dark compared to those of the other models. RetinexNet generated unsatisfactory visual results in terms of brightness, contrast, and naturalness. EnlightenGAN produced washed-out images in environments with large differences between light and dark areas. In comparison, N-LoLiGan not only successfully learned to enhance dark areas and suppress light areas adaptively, but also retained more color information and texture details; therefore, the generated images appeared more natural.

C. Ablation Study

To demonstrate the effectiveness of the components detailed in Section 3, several ablation experiments were conducted. In particular, we designed three different experiments by removing the multi-scale input layer, CBAM attention module, and adaptive global-local illumination

mask dual-discriminator. The first column in Fig. 5 presents example input images used in our ablation experiments. The second column presents the images produced by N-LoLiGan without the multi-scale input layer. The third column presents the results produced by N-LoLiGan without the CBAM attention module. The fourth column presents the results produced by N-LoLiGan without the global-local dual-discriminator for distinguishing between low-light and normal-light images. The final column presents the images produced by the full version of N-LoLiGan.

The enhanced results for the first example in the second through fourth columns tend to contain local regions of overexposure. The enhanced results for the second and third examples from the second through fourth columns lean toward color distortion or underexposure. In contrast, the results of the full version of N-LoLiGan successfully enhance dark areas and suppress light areas, which validates the effectiveness of each component proposed in Section 3.

D. Application to Tunnel Multi-Target Detection

In general, low-light image enhancement is a preprocessing step for high-level computer vision tasks. According to the analyses above, no-reference image quality metrics may not be robust for evaluating low-light image enhancement methods. Therefore, evaluating image enhancement methods based on the performance of tunnel surface defect detection tasks may be a better approach.

In this study, we applied N-LoLiGan and the three other benchmarks to perform preprocessing to enhance the tunnel low-light images in the testing dataset. First, we applied the tunnel surface defect datasets from [40] to train the YOLACT model [41] for 2000 epochs. The other hyperparameters remained unchanged. Neither domain adaptation nor joint training were performed. Subsequently, the trained model was utilized to detect tunnel surface defects in the enhanced images generated by the tested models. Additionally, to verify that our method provides the same level of enhancement for multi-target detection, a pre-trained YOLOv5s model was used to detect people and vehicles in the testing dataset.

TABLE 2
TUNNEL MULTI-TARGET DETECTION RATES OF YOLACT [41] AND YOLOV5S USING N-LOLIGAN AND OTHER IMAGE ENHANCEMENT METHODS FOR PREPROCESSING

Method	Detection rate(%)of yolact	Detection rate(%)of yolov5
Input	62.5	89.69
SRIE	83.3	94.52
RetinexNet	72.2	94.78
EnlightenGan	86.1	96.21
N-LoLiGan	**88.9**	**97.23**

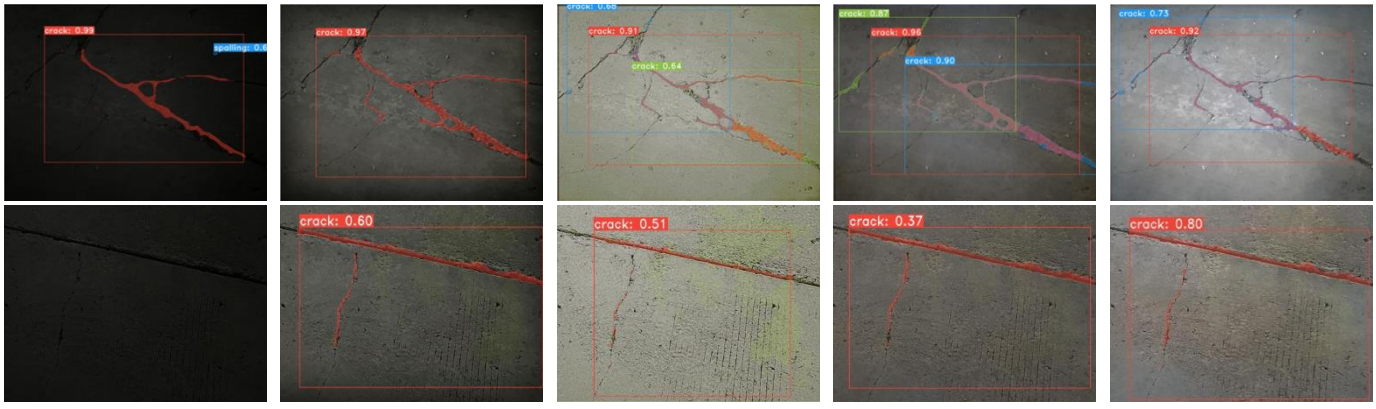

Fig. 6. Visual comparisons of YOLACT [41] tunnel surface defect detection results for tunnel low-light image inputs and various image enhancement methods.

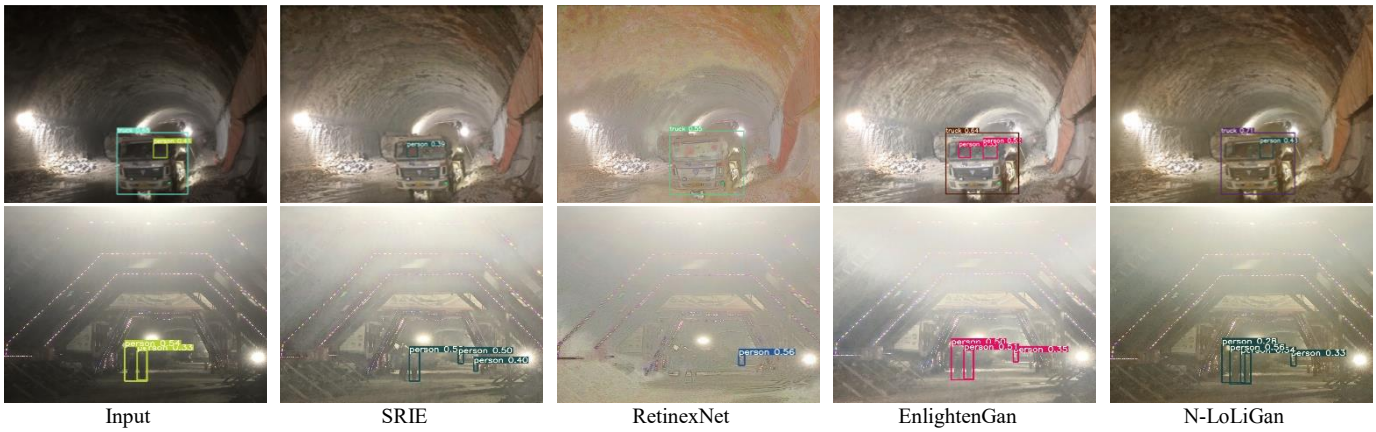

Fig. 7. Visual comparisons for YOLOv5s tunnel people and vehicle detection results for tunnel low-light image inputs and various image enhancement methods.

Some surface defect, people, and vehicle detection results from using the different methods as preprocessing steps are presented in Figs. 6 and 7. The tunnel surface defect detection rates using different methods as preprocessing steps are listed in Table 2. Compared to the other methods, N-LoLiGan identified defects more completely and with a

higher confidence and detection rate (88.9 %). These results indicate that N-LoLiGan preserves semantic details and maintains superior global consistency in addition to producing visually pleasing results.

V. Conclusion

This study proposed a novel unsupervised end-to-end network for the enhancement of low-light images in tunnels. The proposed N-LoLiGan can not only perceive the distribution of light and dark images well and realize the enhancement of dark areas and suppression of bright areas, but can also balance the contrast of images, thus making generated images appear more natural. Experimental results for real tunnel images revealed that the proposed N-LoLiGan outperforms several state-of-the-art low-light image enhancement methods in terms of quantitative and qualitative image quality evaluations, improves the tunnel surface defect detection rate from 62.5 % to 88.9 %, and improves the people and vehicle detection rate in tunnels from 89.69 % to 97.23 %. These results prove that the proposed method can operate and generalize well in other target detection fields.

References

[1] X. Liu, Y. Wang, Z. Yu, L. Zhu, H. Shi, and H. Huang, "Research on Cracks Image Detection System for Subway Tunnel," in ITOEC, 2018, pp. 188-192.
[2] Y. Meng, H. Wu and B. Niu, "A Method to Improve the Lining Images Quality in Complex Tunnel Scenes," in ICIVC, 2020, pp. 199-203.
[3] B. Niu, H. Wu and Y. Meng, "Application of CEM Algorithm in the Field of Tunnel Crack Identification," in ICIVC, 2020, pp. 232-236.
[4] L. Xu, M. Ma, "Study of the characteristics of train-induced dynamic SIFs of tunnel lining cracks based on the modal superposition approach". Engineering Fracture Mechanics, vol.233, Jun. 2020.
[5] Y. Ren, J. Huang, and Z. Hong, et al. "Image-based concrete crack detection in tunnels using deep fully convolutional networks". Construction and Building Materials, vol. 234, Mar. 2020.
[6] K. Mantiuk, S. Daly, L. Kerofsky, "Display adaptive tone mapping, " in ACM SIGGRAPH Conference, Aug. 2008, doi: 10.1145/1399504.1360667.
[7] M. Gharbi, J. Chen, J.T. Barron, S.W. Hasino, F. Durand, "Deep bilateral learning for real-time image enhancement, "ACM Trans. on Graphics (TOG), vol.36, no.4, pp. 1-12, 2017.
[8] R. Wang, Q. Zhang, C.W. Fu, X. Shen, W.S. Zheng, J. Jia, "Underexposed photo enhancement using deep illumination estimation, " in CVPR., 2019, pp. 6849-6857.
[9] W. Xiong, D. Liu, X. Shen, C. Fang, and J. Luo, “Unsupervised Low-light Image Enhancement with Decoupled Networks.” 2020.
[10] T. Ma et al., "RetinexGAN: Unsupervised Low-Light Enhancement With Two-Layer Convolutional Decomposition Networks," IEEE Access, vol. 9, pp. 56539-56550, 2021.
[11] Y. Qu, Y. Ou and R. Xiong, " Low Light Enhancement by Unsupervised Network," in RCAR, 2020, pp. 404-409.
[12] Y. Jiang et al., "EnlightenGAN: Deep Light Enhancement Without Paired Supervision," IEEE Trans. on Image Processing, vol. 30, pp. 2340-2349, 2021.
[13] M. Abdullah-Al-Wadud, M. H. Kabir, M. A. A. Dewan, and O. Chae, "A dynamic histogram equalization for image contrast enhancement, " IEEE Trans. on Consumer Electronics, vol. 53, no. 2, pp. 593–600, May 2007.
[14] 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 & Image Processing, vol. 39, no. 3, pp. 355-368, Sep. 1987.
[15] H. Ibrahim, N.S.P. Kong, "Brightness preserving dynamic histogram equalization for image contrast enhancement, " IEEE Trans. on Consumer Electronics, vol. 53, no. 4, pp. 1752–1758, Nov. 2007.
[16] E.H. Land, "The retinex theory of color vision, " Scientific American, vol. 237, no. 6, pp.108-129, Jan. 1977.
[17] 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 Trans. on Image Processing, vol. 6, no. 7, pp. 965-976, July 1997.
[18] D. J. Jobson, Z. Rahman and G. A. Woodell, "Properties and performance of a center/surround retinex," in IEEE Trans. on Image Processing, vol. 6, no. 3, pp. 451-462, March 1997.
[19] C.H. Lee, J.L. Shih, C.C. Lien, and C.C. Han, "Adaptive multiscale retinex for image contrast enhancement," in SITIS, Dec. 2013, pp. 43-50.
[20] S. Wang, J. Zheng, H.-M. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," IEEE Trans. on Image Processing, vol. 22, no. 9, pp. 3538-3548, Sep. 2013.
[21] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in CVPR, Jun. 2016, pp. 2782-2790.
[22] X. Guo, Y. Li, and H. Ling, "LIME: Low-light image enhancement via illumination map estimation," IEEE Trans. Image Process., vol. 26, no. 2, pp. 982-993, Feb. 2017.
[23] Z. Ying, G. Li, and W. Gao, "A bio-inspired multi-exposure fusion framework for low-light image enhancement," 2017, arXiv:1711.00591..
[24] Y. Liu, H. Yin, J. Wan, Z. Liu, and A. Chong, "Edge aware network for image dehazing," IEEE Signal Process. Lett., vol. 29, pp.174-178, 2022.
[25] K.G. Lore, A. Akintayo, S. Sarkar, "LLnet: A deep autoencoder approach to natural low-light image enhancement, " Pattern Recognition, no. 61, pp. 650-662, 2017.
[26] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement, " 2018, arXiv.1808.04560.
[27] Y. Zhang, X. Di, B. Zhang, and C. Wang, "Self-supervised image enhancement network: Training with low light images only," 2020, arXiv:2002.11300.
[28] W. Huang, Y. Zhu, and R. Huang, "Low light image enhancement network with attention mechanism and retinex model," IEEE Access, vol. 8, pp. 74306-74314, 2020.
[29] Z. Ying, G. Li, and W. Gao, "A bio-inspired multi-exposure fusion framework for low-light image enhancement," 2017, arXiv:1711.00591.
[30] Y.S. Chen, Y.C. Wang, M.H. Kao, and Y.Y. Chuang, "Deep photo enhancer: Unpaired learning for image enhancement from photographs with GANs," in CVPR, Jun. 2018, pp. 6306-6314.
[31] P. Tao, H. Kuang, Y. Duan, L. Zhong and W. Qiu, "BITPNet: Unsupervised Bio-Inspired Two-Path Network for Nighttime Traffic Image Enhancement," IEEE Access, vol. 8, pp. 164737-164746, 2020.
[32] Y. Guo, Y. Lu, and R. W. Liu, "Lightweight deep network-enabled real-time low-visibility enhancement for promoting vessel detection in maritime video surveillance," Journal of Navigation, vol. 75, no.1, pp. 230-250, Jan. 2022.
[33] A. Garg, X. -W. Pan and L. -R. Dung, "LiCENt: Low-Light Image Enhancement Using the Light Channel of HSL," IEEE Access, vol. 10, pp. 33547-33560, 2022.
[34] H. Fu, J. Cheng, Y. Xu, D. W. K. Wong, J. Liu and X. Cao, "Joint Optic Disc and Cup Segmentation Based on Multi-Label Deep Network and Polar Transformation," IEEE Transactions on Medical Imaging, vol. 37, no. 7, pp. 1597-1605, July 2018.
[35] S. Woo, J. Park, J.Y. Lee, I.S. Kweon, "Cbam: Convolutional block attention module,"in ECCV, Sep. 2018, pp.3-19.
[36] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in ICCV, Oct. 2017, pp. 2794_2802.
[37] Y. Dong et al., "A Deep-Learning-Based Multiple Defect Detection Method for Tunnel Lining Damages," in IEEE Access, vol. 7, pp. 182643-182657, 2019.
[38] W. Xiong, D. Liu, X. H. Shen, C. Fang, and J. B. Luo, "Unsupervised real-world low-light image enhancement with decoupled networks, "May. 2020, arXiv: 2005.02818.
[39] 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, March 2013.
[40] Y. Dong et al., "A Deep-Learning-Based Multiple Defect Detection Method for Tunnel Lining Damages," in IEEE Access, vol. 7, pp. 182643-182657, 2019.
[41] D. Bolya, C. Zhou, F. Xiao and Y. J. Lee, "YOLACT: Real-Time Instance Segmentation," in ICCV, 2019, pp. 9156-9165.