




RESEARCH ARTICLE

Attention-guided lightweight generative adversarial network for low-light image enhancement in maritime video surveillance

Ryan Wen Liu,^{1,2,3}  Nian Liu,¹ Yanhong Huang,^{1*} and Yu Guo^{1*}

¹Hubei Key Laboratory of Inland Shipping Technology, School of Navigation, Wuhan University of Technology, Wuhan, China

²Chongqing Research Institute, Wuhan University of Technology, Chongqing, China

³Hainan Institute, Wuhan University of Technology, Sanya, China.

*Corresponding authors. E-mails: yhuang@whut.edu.cn; yuguo@whut.edu.cn

Received: 13 March 2022; Accepted: 23 July 2022; First published online: 30 August 2022

Keywords: maritime surveillance; vessel traffic service; low-light image enhancement; vessel detection; generative adversarial network

Abstract

Benefiting from video surveillance systems that provide real-time traffic conditions, automatic vessel detection has become an indispensable part of the maritime surveillance system. However, high-level vision tasks generally rely on high-quality images. Affected by the imaging environment, maritime images taken under poor lighting conditions easily suffer from heavy noise and colour distortion. Such degraded images may interfere with the analysis of maritime video by regulatory agencies, such as vessel detection, recognition and tracking. To improve the accuracy and robustness of detection accuracy, we propose a lightweight generative adversarial network (LGAN) to enhance maritime images under low-light conditions. The LGAN uses an attention mechanism to locally enhance low-light images and prevent overexposure. Both mixed loss functions and local discriminator are then adopted to reduce loss of detail and improve image quality. Meanwhile, to satisfy the demand for real-time enhancement of low-light maritime images, model compression strategy is exploited to enhance images efficiently while reducing the network parameters. Experiments on synthetic and realistic images indicate that the proposed LGAN can effectively enhance low-light images with better preservation of detail and visual quality than other competing methods.

1. Introduction

A maritime video surveillance system accurately records the real-time situation of surrounding waters, which is an essential part of vessel traffic supervision (Huang et al., 2021; Liu et al., 2021). Through the video surveillance system, managers can effectively and efficiently carry out early warning, remote emergency handling and forensic investigations (Nie et al., 2019; Chen et al., 2021). Moreover, vessel detection has been widely used in maritime surveillance, vessel rescue and other fields with significant application value. It greatly promotes the intelligence of maritime traffic supervision, and becomes an important auxiliary means of maritime video supervision. Many vessel detection methods have been proposed, which have accurate detection results under normal lighting conditions. However, the current mainstream vessel detection methods mostly rely on high-quality inputs. Due to the poor weather and low illumination in the marine environment, it is difficult for shore-based cameras to obtain clear images. The maritime images captured by shore-based surveillance cameras tend to have low contrast and high noise, making it difficult for supervisors to automatically detect targets hidden in

the dark (Guo et al., 2022). Such degraded images make it more difficult to perform target detection robustly and accurately. In addition, influenced by the traffic environment, the images captured by surveillance cameras are characterised by a complex background and small (vessel) targets. These result in less structural information that can be used for image restoration, thereby increasing the difficulty of maritime surveillance. Image enhancement aims to improve the quality of images commonly using computer vision technologies. Low-light image enhancement can provide clear and reliable inputs for vessel detection tasks, which has positive significance for improving the intelligence of maritime supervision.

Traditional enhancement methods are mainly classified into two categories: histogram equalisation-based methods and retinex-based methods. Retinex-based methods are widely used to enhance non-uniformly illuminated images. This kind of method first estimates the light component of a low-illumination image by manual priors and parameter adjustment, then obtains the enhanced image directly or indirectly according to the retinex theory (Land, 1964). For example, Wang et al. (2013) proposed a bi-log transformation for mapping illumination to strike a balance between details and naturalness. A method was proposed by Guo et al. (2016) effectively to achieve low-illumination image enhancement (LIME) by estimating only the illumination map. A weighted variable model was proposed by Fu et al. (2016) to estimate the illumination and reflection components of low-illumination images to improve non-uniform illumination images. As traditional methods suffer from problems such as artificial prior and parameter optimisation, deep learning-based methods are widely used to enhance images, such as SICE (Cai et al., 2018), RetinexNet (Wei et al., 2018), etc. These mainstream methods have been proven to have good enhancement effects in traditional imaging scenarios. However, most deep learning-based methods are not feasible in aiding vessel detection because they do not take into account the characteristics of marine images. Maritime images are characterised by a high percentage of sky background and small vessel targets, which makes less information available for image recovery. All of these undoubtedly increase the difficulty of low-light image enhancement in maritime applications.

Therefore, the existing image enhancement methods need further improvement when applied to maritime scenes. To efficiently enhance image quality in practical applications, we develop a lightweight generative adversarial network (LGAN) to facilitate maritime supervision tasks. The LGAN uses the residual networks and pyramidal dilated convolutional (PDC) layer for powerful feature extraction. Meanwhile, the attention mechanism is introduced to enable adaptive enhancement of different illumination regions based on the light distribution maps. In addition, a local discriminator is designed to guide LGAN to obtain more realistic enhancement results. The following points are the main contributions of this paper.

- To guarantee efficient image enhancement under low-light imaging conditions, LGAN is proposed for promoting video-based maritime surveillance systems.
- An attention mechanism, which captures the robust illumination map, is further incorporated into the proposed LGAN network. It is capable of handling low light and preventing overexposure.
- Comprehensive experiments on image enhancement and vessel detection are performed to evaluate the method in terms of both effectiveness and robustness.

The remaining part of this paper is divided into the following sections. A brief review of related works is provided in Section 2. Section 3 introduces the LGAN in detail. Section 4 shows the experimental details and demonstrates the effectiveness of the LGAN through extensive experiments. Finally, the work is summarised and future work is discussed in Section 5.

2. Related works

There are numerous low-visibility image enhancement methods, which can be broadly classified into three classes: histogram equalisation-based methods, retinex-based methods and deep learning-based methods. In this section, these three methods will be introduced separately.

2.1. Histogram equalisation-based methods

Histogram equalisation (HE) improves image contrast by smoothing and non-linearly stretching the dynamic range of an image histogram (Pizer *et al.*, 1987). This kind of method can be subdivided into global and local HE, *i.e.*, GHE and LHE (Senthilkumaran and Thimmiaraja, 2014). GHE directly adjusts the overall greyscale of the degraded image according to the global information (Reza, 2004). Quadrant dynamic HE (QDHE), proposed by Ooi and Isa (2010), obtains better enhanced images without noise amplification and excessive enhancement. However, the GHE-based enhancement method cannot extract detailed information from low-light images. In contrast, LHE could effectively acquire small-scale details in low-light images to achieve local enhancement (Kim *et al.*, 2001). Valuchamy and Subramani (2019) introduced an efficient enhancement method based on improved adaptive gamma correction and HE. Although the HE-based methods can be implemented efficiently, they do not suppress noise generation and may lead to image distortion (Wang *et al.*, 2020).

2.2. Retinex-based methods

Retinex theory (Land, 1964) was first introduced by Edwin Land in the 1970s. It states that the physical properties of an object seen by human eyes are independent of the external environment and the illumination of incident light, but only directly related to the reflective properties of this observed object. The classical retinex-based imaging methods include single-scale retinex (SSR), multi-scale retinex (MSR), and MSR with colour restoration (MSRCR) (Choi *et al.*, 2008). SSR (Jobson *et al.*, 1997b) cannot compensate for both details and illumination enhancement at the same time. Although MSR (Lin and Shi, 2014) has been proposed based on SSR, it suffers from edge blurring when enhancing images. MSRCR (Petro *et al.*, 2014) can be effective in reducing random noise and restoring image colours. As the correlation between different colour components is neglected in colour restoration, it often suffers from colour inversion and distortion problems (Jobson *et al.*, 1997a; Ma *et al.*, 2017). Kim *et al.* (2019) used the maximum value obtained during diffusion as the illumination component. After adjusting the illumination component by global stretching, it is combined with the reflection component to generate the final enhancement image with local refinement.

2.3. Deep learning-based methods

Compared with traditional imaging methods, deep learning-based low-light image enhancement methods have gained significant attention from both academia and industry. These methods are mainly divided into two classes: supervised and unsupervised learning methods (Zhao and Shi, 2019). The supervised learning methods require training model with labelled paired data to optimise network parameters. For example, Lore *et al.* (2017) designed a stacked sparse denoising autoencoder to enhance low-light images, demonstrating the effectiveness under different imaging conditions. Cai *et al.* (2018) trained an enhancer for improving the contrast of underexposed or overexposed images. As these methods require paired low-visibility and normal images for supervised training, there is a higher demand on the datasets. In contrast, the unsupervised or self-supervised learning methods are free from the intrinsic limitation of labelled data. The unsupervised learning-based methods train models based on the loss functions and the light components of degraded images. Guo *et al.* (2020) proposed to consider the low-light image enhancement as an image-specific curve estimation task. Xu *et al.* (2020) developed a decomposition-and-enhancement framework in the frequency domain, which obtains the enhancement results by recovering low-frequency information and high-frequency details from degraded images. More recently, a novel multi-branch topology residual block (MTRB) strategy (Lu *et al.*, 2022) was proposed to restore low-light images. Benefiting from the development of generative adversarial networks (GAN), Jiang *et al.* (2021) proposed the EnlightenGAN to train the model using unpaired images. However, it is intractable to effectively restore the low-light images directly using existing deep learning methods in maritime scenes.

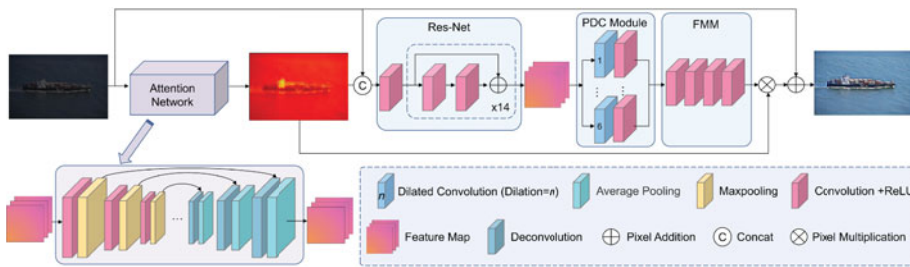


Figure 1. Attention-guided generative adversarial network. It consists of attention network and enhancement network. Further, the enhancement network includes three parts: Res-Net, PDC module and FFM.

3. LGAN: lightweight generative adversarial network

In low-illumination environments such as cloudy weather or nighttime, the images captured by shore-based video cameras may lose important image details. This phenomenon would lead to a negative impact on maritime surveillance tasks. Currently, there are few low-illumination image enhancement methods suitable for marine scenes. To assist vessel detection, we design a LGAN for maritime image enhancement. The LGAN tends to deliver a self-regularised light distribution map as a self-adjusting attention map to the network for better enhancement results. In this section, the network architecture and loss function will be introduced in detail.

3.1. Network architecture

We design the LGAN with reference to the attention-guided GAN, as shown in Figure 1. In the generator, the attention network estimates the light distribution of low-light images to distinguish the underexposed and normally exposed regions. The attentional feature maps enable the enhancement network to focus on local information of the degraded images, which can effectively solve the problem of over-enhancement. The enhancement network obtains an enhanced image by performing a series of transformations on the degraded image. It mainly consists of three components, i.e., residual network (Res-Net), PDC module, and feature fusion module (FFM). Res-Net is a residual network with 15 residual blocks for extracting deep features of low-light images. Each residual block contains two convolutions and a shortcut path. The residual blocks learn features by skip connection, which is a flexible structure (Liu et al., 2022).

Although the attention-guided GAN can effectively enhance low-illumination images, the long computational time does not satisfy the requirement of real-time vessel detection. To reduce the computational effort and to realise real-time computation, we propose the lightweight version (i.e., LGAN) based on the attention-guided GAN shown in Figure 1. When designing the LGAN, the attention network was first removed to reduce the number of network parameters. Accordingly, the self-regularised light distribution map is exploited to implement the attention mechanism, which can effectively prevent overexposure while reducing the computational effort. At the same time, the feature extraction network is changed from Res-Net-15 to Res-Net-5 to further simplify the network. Further, we replace the standard convolution in the model with the depthwise separable convolution, and replace the residual block in Res-Net with the inverse residual. To avoid information loss, only the activation function of the last point convolution is retained, and the batch normalisation is removed. The structure of the generator in LGAN is shown in Figure 2. The generator G of LGAN consists of three main parts, i.e., inverted residual network (Inverted Res-Net), PDC module and FFM.

The Inverted Res-Net is the backbone of our LGAN, which avoids information loss while deepening the network hierarchy. The addition of the inverse residual module can effectively strengthen the ability of the LGAN to extract features from low-illumination images. The PDC module is applied to multi-scale spatial feature extraction, which minimises the loss of structural information. FFM fuses the deep and multi-scale features extracted by the Inverted Res-Net and PDC module.

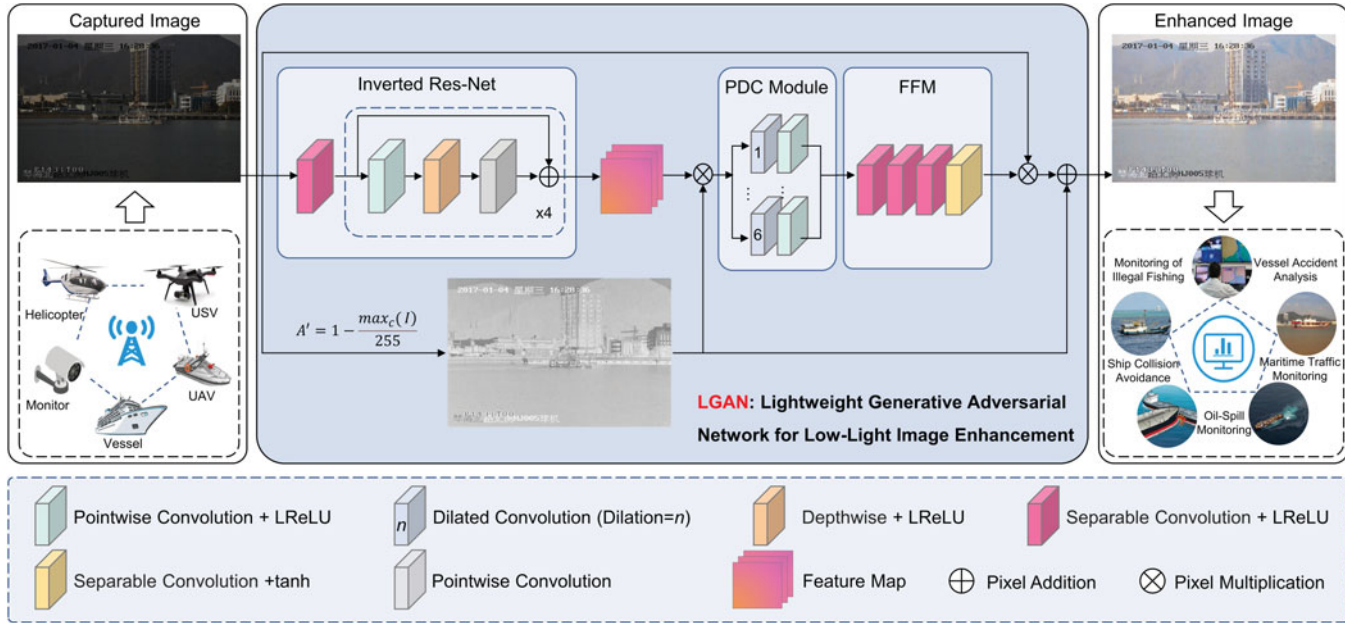


Figure 2. Flowchart of LGAN. LGAN has three main components, i.e., Inverted Res-Net, PDC module and FFM. In particular, the attention mechanism is implemented by introducing a self-regularised light distribution map in LGAN.

The original low-illuminance images and the brightness distribution maps are stitched together as the inputs of LGAN. The inputs first pass through a residual network structure with five inverted residual modules to extract features of the low-light images. Dilated convolution can increase the receptive fields and capture multi-scale context information without increasing the amount of network calculations. We thus add the PDC module after the Inverted Res-Net to extract multi-scale spatial features to avoid or minimise the loss of structural information. The PDC module includes a group of dilated convolutions which have different dilation rates. It has four parallel paths, each of which contains a dilated convolution and a convolution kernel of size 1×1 . The related dilation rates are one, two, four and six, respectively. The features extracted by Inverted Res-Net are spliced with the features extracted by each path of the PDC module. Four convolutional layers are then exploited to fuse the spliced features. They are multiplied with the illuminance distribution map and then added to the original low-illuminance image for final output.

Normally, there is an inconsistent degree of information loss in different illuminance regions of an image, thus we designed a discriminator, D, which could autonomously give different attention to different illuminance regions. The discriminator D consists of seven convolutional layers and two fully connected layers. In particular, multiple convolutional layers are used progressively to extract image features. The fully connected layers determine whether the image is real based on the extracted features.

The input images of the discriminator D are first passed through five convolutional layers for feature extraction. The feature maps extracted from the fifth convolutional layer pass through two branches. One branch is used to calculate the loss with the illuminance distribution map. The other branch works as input to the sixth convolutional layer. After two convolutional layers and fully connected layers, the extracted features are summarised to obtain the discriminant results. In order to reduce the parameter redundancy, we obtain a lighter GAN model by re-training this discriminator after pruning it.

3.2. Loss function of LGAN

Different optical devices and sensors may cause different non-linear distortions. Although the images are processed when constructing the dataset, there are still varying degrees of pixel-level offset between images. Therefore, it is obviously difficult to obtain accurate outputs using only loss functions based on pixel point differences, such as L_1 and L_2 loss functions. To improve the overall perceptual quality of the recovered image, we propose an improved loss function, which mainly consists of the generator loss L_G and the discriminator loss L_D .

3.2.1. Perceptual loss

The VGG-16 network pre-trained on the ImageNet dataset has strong feature extraction capabilities. It is typically applied to extract high-level features from both generated images and target images. The similarity among images is measured by calculating the pixel-level differences between these higher-level features. Therefore, the content perception loss defined in this paper is as follows:

$$L_{\text{con}} = \frac{1}{N} \|\phi_j(G(I)) - \phi_j(\hat{I})\|_2^2, \quad (1)$$

where ϕ_j represents the high-level feature extracted in the j -th convolutional layer of VGG-16, G represents the network of generator, N is the total number of training samples, I and \hat{I} denote the input and target images, respectively.

3.2.2. Colour loss

It is necessary to measure the differences in colour between the enhanced and normally-illuminated images. To better evaluate the colour differences, the Gaussian blur function is employed to weaken edge details in an image (Ignatov et al., 2017). In this work, the Euclidean distance between the images

after Gaussian blurring is given by

$$L_{col} = \|g(G(I)) - g(\hat{I})\|_2^2, \tag{2}$$

where g denotes the Gaussian blur function.

3.2.3. Adversarial loss

Adversarial loss urges the generator to generate a more natural enhanced image in terms of colour, texture, contrast, etc. Its definition is as follows:

$$L_{adv} = \log(1 - D(G(I))), \tag{3}$$

where D represents the network of the discriminator.

3.2.4. Gradient loss

As maritime images captured in low-visibility environments tend to be heavily affected by noise, image quality may suffer to some degree. The gradient loss is thus designed to achieve the denoising effect by calculating the square of the gradient difference between images. It makes the enhanced image spatially smooth. The gradient loss can be expressed as

$$L_{gd} = \|\nabla_x I - \nabla_x \hat{I}\|_2^2 + \|\nabla_y I - \nabla_y \hat{I}\|_2^2, \tag{4}$$

where ∇_x and ∇_y , respectively, denote the gradients in the x and y directions.

Compared with normal convolutions, deep separable convolutions can reduce the amount of parameters while causing a partial loss of network performance. The L_2 loss is then applied to ensure structural similarities between the generated and clear images. The L_2 loss function is defined as follows:

$$L_2 = \|I' - \hat{I}\|_2^2, \tag{5}$$

where I' represents the feature obtained by the Inverted Res-Net.

Therefore, the total loss of generator in LGAN is defined as follows:

$$L_G = \omega_{con}L_{con} + \omega_{col}L_{col} + \omega_{adv}L_{adv} + \sigma L_{gd} + \mu L_2, \tag{6}$$

where L_{con} , L_{col} , L_{adv} , L_{gd} denote the content perceived loss, colour loss, adversarial loss and gradient difference loss, respectively, ω_{con} , ω_{col} , and ω_{adv} represent the corresponding weight parameters, σ and μ are weighting parameters.

Based on the distribution of light in different light regions between the enhanced and sharp images, the discriminator D obtains the probability that the input is true or false. Thus, the discriminator loss L_D can be written as follows:

$$L_D = -\log(D(\hat{I})) - \log(1 - D(G(I))) + \gamma L_{map}, \tag{7}$$

where γ denotes the weight parameter, L_{map} represents the loss between the feature maps extracted from discriminator and the light distribution maps. The L_{map} can be written as follows:

$$L_{map} = \frac{1}{N} \|D(G(I))_5 - A\|_2^2 + \frac{1}{N} \|D(\hat{I})_5 - 0\|_2^2, \tag{8}$$

where $D(\cdot)_5$ represents the fifth layer convolution of discriminator, A indicates the light distribution map. Since \hat{I} indicates a real clear image, zero means it has no specific region on which to focus.



Figure 3. Some examples of paired clear/low-light images. The original sharp images are on the top row, above the corresponding synthetic low-light images.

4. Experimental results and analysis

The synthesis dataset adopted for network training is described in this section. Details of the experimental platform and experimental parameter settings are also introduced. To demonstrate the effectiveness of the LGAN, both quantitative and qualitative experiments were conducted on both synthetically-degraded and original images.

4.1. Synthetically-degraded image generation

The LGAN is proposed to enhance low-light maritime images. It learns how to map from degraded images to latent sharp images through supervised learning. It thus needs many paired images to complete the network training. However, it is hard to capture both low-visibility and sharp images jointly in a dynamic environment. The existing datasets could not be directly employed to train our network for the enhancement of low-light maritime images.

In this paper, a mixed dataset of realistic and synthetic images is used to complete the network training. The synthetic low-illumination images are obtained by globally reducing the brightness of normal images. The dataset we used is partially derived from the publicly available SeaShips dataset (Shao et al., 2018). By filtering out poor-quality images (e.g., motion blur and focus blur), 500 normally-illuminated images of size 1920×1080 were randomly obtained. Each clear image was translated from the RGB (red, green, blue) colour space to the Hue-Saturation-Value (HSV) colour space. The illumination component V of each image was then reduced by a randomly selected scale factor in the range from 0 to 0.5. After transforming them back into RGB colour space, we finally obtained 500 pairs of clear/low-light images. Figure 3 shows the synthesised results from some clear images.

To increase the diversity of training data samples, another part of the training dataset was derived from the LOL dataset (Wei et al., 2018), which has 500 pairs of low-illumination/clear images acquired from the real world with a resolution size of 600×400 . To reduce the training time, the training images were first randomly cropped into image blocks of size 128×128 before being fed into the network. In addition, the dataset was enlarged using popular augmentation methods, such as random translation and horizontal and vertical flipping. Such operations can effectively strengthen the generalisation ability of an image enhancement network. Results with different augmentation methods are shown in Figure 4.

4.2. Implementation details

During the network training, the initial learning rate, decay rate and decay coefficient are set to 0.002, 1,000 and 0.96, respectively. Moreover, the exponential decay method is used to reduce the learning rate. Both Adam optimiser and momentum optimiser are exploited in the generator and discriminator for network parameter optimisation. The LGAN is trained using a small batch training method with a batch size of eight and a number of iterations of 200. The weight parameters ω_{con} , ω_{col} , ω_{adv} , γ , σ and μ are 2, 11, 0.01, 0.05, 10 and 1, respectively.

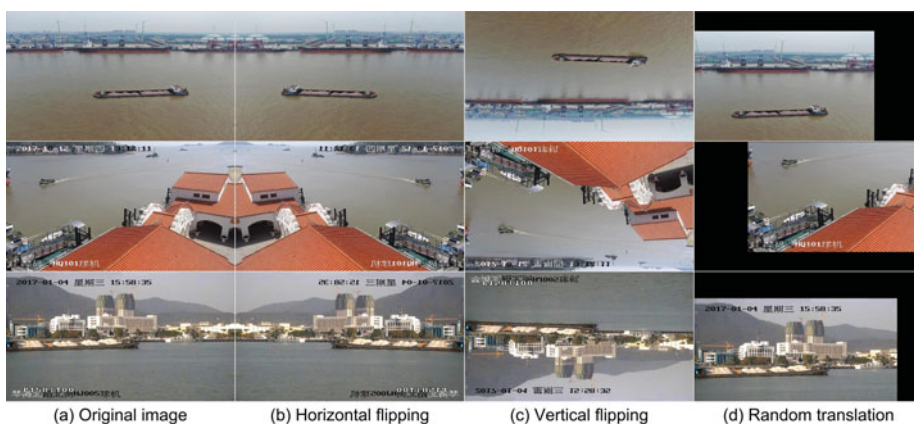


Figure 4. Examples of different augmentation methods: (a) original images, and augmented images obtained by (b) horizontal flipping, (c) vertical flipping, (d) random translation.

In this paper, each convolutional layer adopts the same filling method. Meanwhile, to increase the nonlinearity of the LGAN, only the last convolutional layer of the generator uses the sigmoid function as the activation function. Leaky rectified linear unit (Leaky ReLU) is added after all other convolution layers, and the negative slope of Leaky ReLU is set to 0.2 .

To reduce the computing time of the LGAN, the self-regularised light distribution map (Jiang et al., 2021) can be obtained through the following equation

$$A = 1 - \frac{\max_c(I)}{255}, \quad (9)$$

where $\max_c(\cdot)$ represents the maximum value of each channel of input I . The light distribution maps obtained by Equation (9) are stitched with the low-visibility image as the inputs to LGAN.

4.3. Comparisons with other competing methods

In this subsection, the following comparison methods are chosen to participate in the image enhancement experiments.

- SRIE: Simultaneous reflectance and illumination estimation (Fu et al., 2016) divides the low-visibility image into reflectance and illumination components. The recovered image could be obtained using the retinex theory. It effectively improves the whole brightness and contrast of low-visibility images.
- LIME: LIME (Guo et al., 2016) is a simple and efficient enhancement method that refines the initial light map through a prior as the final light map, and generates the result based on the light map.
- RetinexNet: RetinexNet (Wei et al., 2018) constrains the consistency of reflectance for low-light/clear images. Numerous experiments have shown that it can obtain a better low-visibility enhancement effect and represent image decomposition well.
- DPED: DSLR photo enhancement dataset (DPED) (Ignatov et al., 2017) is an end-to-end enhancement method, which can directly generate enhanced images from low-visibility images through the network.
- LightenNet: LightenNet (Li et al., 2018) directly learns the mapping of low-illumination images to light components through the convolutional neural network, then enhances the images based on retinex theory.

4.4. Evaluation metric

To evaluate the performance of different methods, two full reference evaluation metrics, i.e., peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM), are employed to evaluate the similarity between images. Moreover, two no-reference metrics, i.e., NIQE and BRISQUE, are chosen to assess the visual effect of generated images.

- PSNR: PSNR (Wang and Bovik, 2009) is an objective measure of the level of image distortion or noise. The higher its value, the less noise the image contains. The formula for calculating PSNR is as follows:

$$\text{PSNR}(\hat{Y}, Y) = 10 \log \frac{M^2}{\text{MSE}(\hat{Y}, Y)}, \quad (10)$$

where M denotes the maximum pixel value, and $\text{MSE}(\cdot)$ indicates the mean square error.

- SSIM: SSIM (Wang et al., 2004) is used to measure the brightness, contrast and structural similarities between two images. Therefore, SSIM displays favourable agreement with human observers. Its value has a range from 0 to 1. The closer the value is to 1, the more similar the two images are.
- NIQE: NIQE (Mittal et al., 2012b) does not need to adopt a reference image, but simply uses the measurable deviation of the target image to obtain the image quality evaluation value according to the natural statistical characteristics. The NIQE metric is calculated as follows:

$$(v_1, v_2, \Sigma_1, \Sigma_2) = \sqrt{(v_1 - v_2)^T \left(\frac{\Sigma_1 + \Sigma_2}{2} \right)^{-1} (v_1 - v_2)}, \quad (11)$$

where v_1 , v_2 and Σ_1 , Σ_2 denote the mean vectors and covariance matrices of the Gaussian distribution models of the natural and enhanced images, respectively.

- BRISQUE: BRISQUE (Mittal et al., 2012a) is a reference-free spatial domain image quality evaluation algorithm. The principle of BRISQUE is to extract local normalisation coefficients (MSCN) from images to fit them to generalised Gaussian distribution and asymmetric generalised Gaussian distribution. In addition, it extracts 36 features using multi-scale images and derives objective quality scores using support vector machine.

4.5. Experimental results on syntactic maritime images

Figures 5–7 visualise the enhancement results of LGAN and other comparable methods on synthetic low-illumination images. It can be seen that SRIE and DPED have limited ability to enhance images with dark regions. Their results have lower overall brightness compared with other enhanced images. The images produced by RetinexNet have obvious differences in structure and colour compared with the sharp images. LightenNet has limited effect on image enhancement, whose restored images have irregular colour spots. As shown in Figure 7, LIME has better visual effect, but suffers from colour distortion. In contrast, the proposed LGAN has a more stable enhancement effect on images with different illumination levels. This is mainly due to the incorporation of attention mechanism, which enables the LGAN to perform adaptive enhancement and have powerful generalisation ability.

PSNR and SSIM are used to further verify the enhancement ability and structure preservation ability of LGAN. From Table 1, it can be seen that LGAN performs well in PSNR and SSIM evaluation metrics compared with others. It indicates that LGAN not only significantly improve the contrast of images and retain more detailed information, but also suppress unwanted noise. Thus, the visual qualities of low-light images are significantly improved accordingly.

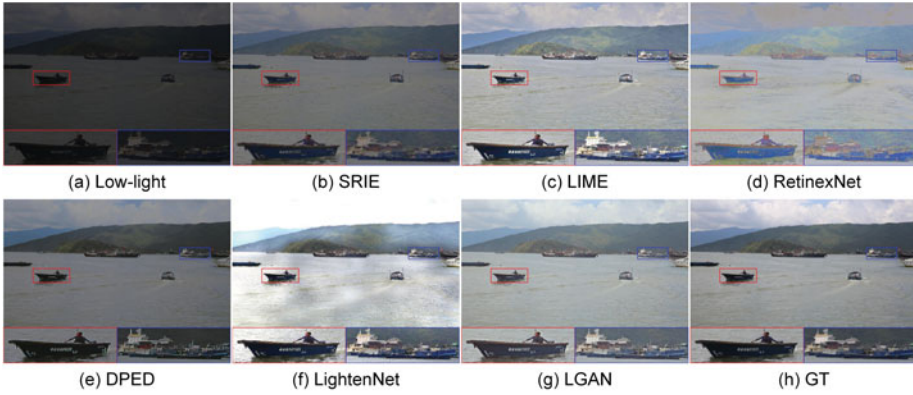


Figure 5. Visual comparisons of proposed LGAN and mainstream methods on Image 1: (a) synthetic low-light image, outputs obtained by (b) SRIE, (c) LIME, (d) RetinexNet, (e) DPED, (f) LightenNet, (g) LGAN, (h) ground truth (GT).

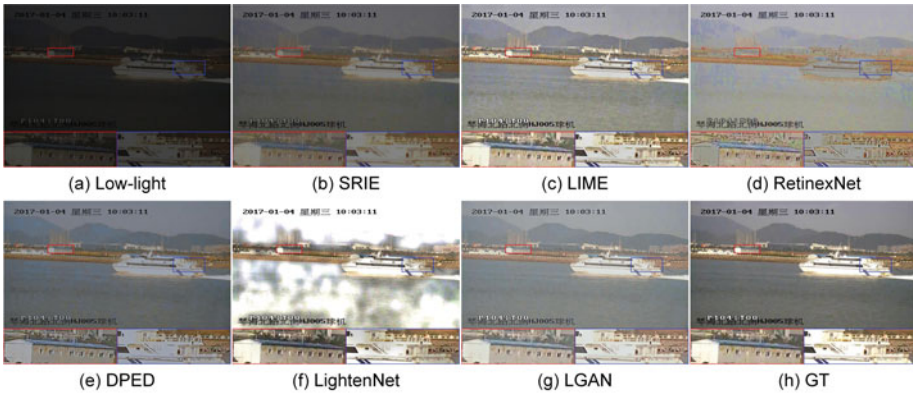


Figure 6. Visual comparisons of proposed LGAN and mainstream methods on Image 2: (a) synthetic low-light image, outputs obtained by (b) SRIE, (c) LIME, (d) RetinexNet, (e) DPED, (f) LightenNet, (g) LGAN, (h) ground truth (GT).

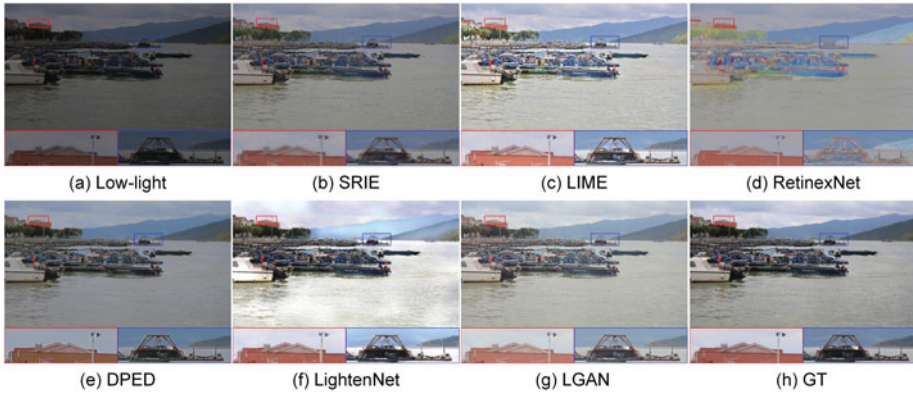


Figure 7. Visual comparisons of proposed LGAN and mainstream methods on Image 3: (a) synthetic low-light image, outputs obtained by (b) SRIE, (c) LIME, (d) RetinexNet, (e) DPED, (f) LightenNet, (g) LGAN, (h) ground truth (GT).

Table 1. Comparisons of PSNR and SSIM values for different competing methods on Images 1–3.

Image	Evaluation metric	Methods					
		SRIE	LIME	RetinexNet	DPED	LightenNet	LGAN
Image 1	PSNR	13001	6.858	15.044	12.893	4.383	21.904
	SSIM	0.658	0.416	0.771	0.642	0.298	0.955
Image 2	PSNR	13.596	6.708	14.353	10.479	6.259	20.521
	SSIM	0.582	0.342	0.866	0.480	0.327	0.953
Image 3	PSNR	13.403	7.535	15.636	13.180	6.084	20.910
	SSIM	0.701	0.540	0.794	0.727	0.459	0.944

Table 2. Comparisons of NIQE and BRISQUE values for different competing methods on Images 4–6.

Image	Evaluation metric	Methods					
		SRIE	LIME	RetinexNet	DPED	LightenNet	LGAN
Image 4	NIQE	4.934	4.522	4.652	4.625	5.546	4.241
	BRISQUE	0.491	0.493	0.487	0.502	0.534	0.482
Image 5	NIQE	5.961	6.567	5.513	5.068	7.492	5.502
	BRISQUE	0.489	0.492	0.495	0.492	0.501	0.487
Image 6	NIQE	4.824	6.082	9.867	5.351	6.781	4.627
	BRISQUE	0.501	0.492	0.483	0.497	0.472	0.395

4.6. Experimental results on realistic maritime images

To further verify the superiority of LGAN for low-light maritime image enhancement, numerous experiments were conducted on realistic low-illumination images. The comparisons of visual effects from different methods are shown in Figure 8. LIME can effectively improve the brightness of low-light images. However, the enhanced results often suffer from the problem of local overexposure. The images enhanced by LIME are not natural enough in practice. RetinexNet can cause colour distortion in the images which is not suitable for auxiliary vessel detection. The images generated by LightenNet have trouble with unnatural colour transitions. Although the images produced by DPED and SRIE are visually pleasing, they both have a certain level of detail loss. Compared with LGAN, the enhanced images of SRIE are more blurred, which indicates that SRIE is less effective in recovering details. Although DPED can achieve the purpose of improving illumination, the enhanced images show colour deviation and colour distortion. In contrast, the LGAN is more capable of enhancing the brightness and contrast, while being able to retain more image details. The images enhanced by LGAN are more natural and have better visual appearances.

In addition, NIQE and BRISQUE were selected to evaluate quantitatively the imaging quality for different image enhancement methods. As illustrated in Table 2, although the NIQE score of LGAN is lower than DPED in Image 5, the BRISQUE metrics of LGAN perform better than other competing methods. It shows that LGAN performs better enhanced effects on low-illumination maritime images.

4.7. Running time analysis

To verify the advantages of LGAN in terms of model size and running time, this subsection compares the proposed LGAN method with three other methods, i.e., RetinexNet, DPED and LightenNet. As shown in Table 3, two maritime images with different resolutions were chosen to compare the

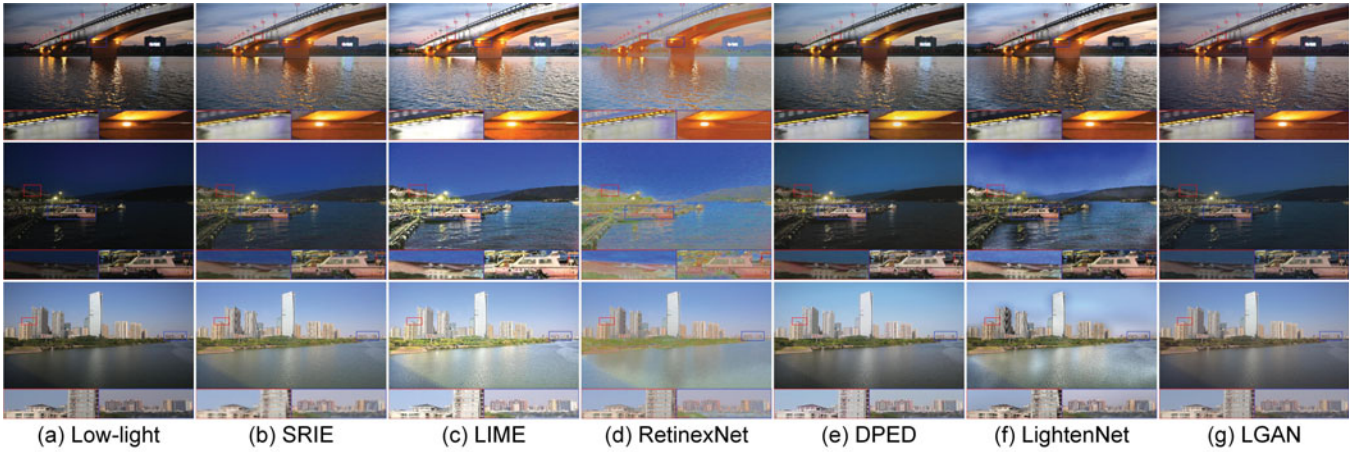


Figure 8. Visual comparisons of proposed LGAN and mainstream methods on Images 4–6. The three rows, from top to bottom, are Image 4, Image 5 and Image 6, respectively.

Table 3. Comparisons of running time (in seconds) and model size for different competing methods.

Method	Platform	Model size (in Megabytes or Kilobytes)	Image size	
			480 × 270	640 × 360
RetinexNet	Python	8 · 069M	1 · 700	1 · 793
DPED	Python	1 · 532M	0 · 051	0 · 062
LightenNet	Matlab	227K	4 · 452	5 · 127
LGAN	Python	282K	0 · 022	0 · 036

running time for different image enhancement methods. All experiments were performed on a computer with the Intel (R) Core (TM) i5-10600KF CPU @ 4.10 GHz and Nvidia GeForce GTX 2080 Ti GPU.

Although the model size of the LGAN is slightly larger than LightenNet, LGAN generates shorter running time than both DPED and LightenNet. This is because only standard convolution is used in DPED for feature extraction. LightenNet uses an additional intermediate operation such as bootstrap filtering in the image enhancement process, which consumes some extra time. In contrast, LGAN uses deeply separable convolution instead of standard convolution to significantly reduce computational effort. In addition, the model pruning and compression operations allow the model size to be greatly compressed without compromising the imaging performance. These operations allow LGAN to satisfy the demands of real-time low-light image enhancement in maritime surveillance tasks.

4.8. Experiment on vessel detection under low-light environment

This subsection presents vessel detection experiments designed to demonstrate that the proposed LGAN can improve the robustness and accuracy of maritime target detection in low-light imaging environments. In the current literature, many target detection methods have been proposed to promote traffic situation awareness. In this paper, the popular YOLOv4 (Bochkovskiy et al., 2020) is used to perform the vessel detection experiments. The YOLOv4 model employed in our experiments is trained using the SeaShips dataset. We adopt two main vessel detection schemes. One solution directly uses YOLOv4 to detect vessels on low-light images. The other solution first uses the LGAN to recover degraded images, and then performs the YOLOv4-based vessel detection.

The vessel detection results on synthetic and realistic scenarios are shown in Figures 9 and 10, respectively. It is obvious that the implementation of YOLOv4 alone in a low-illumination environment is prone to missed and undetectable detection. If the proposed LGAN is first exploited to perform low-light image enhancement, the intrinsic features of targets of interest could be highlighted from low-light environments. Therefore, the robustness and accuracy of vessel detection are significantly improved accordingly.

5. Conclusion and future work

This paper presents a lightweight image enhancement network, called LGAN, designed for enhancing maritime images under low-light imaging conditions. The LGAN uses an attention mechanism to enhance low-light images locally and prevent overexposure. Both mixed loss functions and local discriminators are introduced to reduce the loss of detail as well as to improve visual image quality. Extensive experiments demonstrated that the proposed LGAN can enhance low-illumination maritime images effectively compared with other competing methods.

Although this paper presents an effective method to enhance low-light images in maritime surveillance videos, there are still some shortcomings. (1) Images taken on sunny days are prone to local high



Figure 9. Vessel detection experiments on synthetic low-visibility maritime images (a). Vessel targets are not easily detected in the synthetic low-illumination images. After enhancement by LGAN (b), the target detection results are very close to ground truth (GT) (c).



Figure 10. Vessel detection experiments on real maritime images (a). Some vessels are not detected on real low-visibility images and there is a problem of repeated detection. In contrast, using LGAN to pre-process low-visibility images can significantly improve the robustness and accuracy of vessel detection (b).

illumination such as water reflections. The proposed method only enhances maritime images under low-light imaging environments and does not consider such problem. It is thus necessary to further extend the LGAN to handle water reflections. (2) Maritime images often suffer from a variety of adverse weather conditions, e.g., low light, haze, rain and snow, etc. However, our LGAN is only capable of enhancing low-light images. To meet the requirement of high-quality image enhancement under complex imaging conditions, there is thus a great potential to develop a unified network to restore degraded images under different adverse weather conditions. In addition, due to its superior performance in robustness and efficiency, the lightweight LGAN can be naturally deployed on devices such as unmanned surface vehicles (USVs) and unmanned aerial vehicles (UAVs). Therefore, with the LGAN-based low-light image enhancement, the accuracy and robustness of target detection, recognition and tracking would be accordingly improved for USVs and UAVs in practical applications.

Acknowledgements. This work was supported by the Research Project of Wuhan University of Technology Chongqing Research Institute (Grant No.: YF2021-13), and the Hainan Provincial Joint Project of Sanya Yazhou Bay Science and Technology City (Grant No.: 520LH057).

References

- Bochkovskiy, A., Wang, C. Y. and Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv :2004.10934.
- Cai, J., Gu, S. and Zhang, L. (2018). Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing*, **27**(4), 2049–2062.
- Chen, X., Ling, J., Wang, S., Yang, Y., Luo, L. and Yan, Y. (2021). Ship detection from coastal surveillance videos via an ensemble canny-Gaussian-morphology framework. *The Journal of Navigation*, **74**(6), 1252–1266.
- Choi, D. H., Jang, I. H., Kim, M. H. and Kim, N. C. (2008). Color Image Enhancement Using Single-Scale Retinex Based on an Improved Image Formation Model. *Proceedings of European Signal Processing Conference*, Lausanne, Switzerland.
- Fu, X., Zeng, D., Huang, Y., Zhang, X. P. and Ding, X. (2016) A Weighted Variational Model for Simultaneous Reflectance and Illumination Estimation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA.
- Guo, X., Li, Y. and Ling, H. (2016). LIME: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, **26**(2), 982–993.
- Guo, C., Li, C., Guo, J., Loy, C. C., Hou, J., Kwong, S. and Cong, R. (2020). Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA.
- Guo, Y., Lu, Y. and Liu, R. W. (2022). Lightweight deep network-enabled real-time low-visibility enhancement for promoting vessel detection in maritime video surveillance. *The Journal of Navigation*, **75**(1), 230–250.
- Huang, Z., Hu, Q., Mei, Q., Yang, C. and Wu, Z. (2021). Identity recognition on waterways: A novel ship information tracking method based on multimodal data. *The Journal of Navigation*, **74**(6), 1336–1352.
- Ignatov, A., Kobyshev, N., Timofte, R., Vanhoey, K. and Van Gool, L. (2017). DSLR-Quality Photos on Mobile Devices with Deep Convolutional Networks. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy.
- Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P. and Wang, Z. (2021). EnlightenGAN: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing*, **30**, 2340–2349.
- Jobson, D. J., Rahman, Z. U. and Woodell, G. A. (1997a). A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image Processing*, **6**(7), 965–976.
- Jobson, D. J., Rahman, Z. U. and Woodell, G. A. (1997b). Properties and performance of a center/surround retinex. *IEEE Transactions on Image Processing*, **6**(3), 451–462.
- Kim, J. Y., Kim, L. S. and Hwang, S. H. (2001). An advanced contrast enhancement using partially overlapped sub-block histogram equalization. *IEEE Transactions on Circuits and Systems for Video Technology*, **11**(4), 475–484.
- Kim, W., Lee, R., Park, M. and Lee, S. H. (2019). Low-light image enhancement based on maximal diffusion values. *IEEE Access*, **7**, 129150–129163.
- Land, E. H. (1964). The retinex. *American Scientist*, **52**(2), 247–264.
- Li, C., Guo, J., Porikli, F. and Pang, Y. (2018). Lightnet: A convolutional neural network for weakly illuminated image enhancement. *Pattern Recognition Letters*, **104**, 15–22.
- Lin, H. and Shi, Z. (2014). Multi-scale retinex improvement for nighttime image enhancement. *Optik*, **125**(24), 7143–7148.
- Liu, R. W., Guo, Y., Lu, Y., Chui, K. T. and Gupta, B. B. (2022). Deep network-enabled haze visibility enhancement for visual IoT-driven intelligent transportation systems. *IEEE Transactions on Industrial Informatics*. [online]: doi:10.1109/TII.2022.3170594.
- Liu, R. W., Yuan, W., Chen, X. and Lu, Y. (2021). An enhanced CNN-enabled learning method for promoting ship detection in maritime surveillance system. *Ocean Engineering*, **235**, 109435.

- Lore, K. G., Akintayo, A. and Sarkar, S.** (2017). LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, **61**, 650–662.
- Lu, Y., Guo, Y., Liu, R. W. and Ren, W.** (2022). MTRBNet: Multi-branch topology residual block-based network for low-light enhancement. *IEEE Signal Processing Letters*, **29**, 1127–1131.
- Ma, J., Fan, X., Ni, J., Zhu, X. and Xiong, C.** (2017). Multi-scale retinex with color restoration image enhancement based on Gaussian filtering and guided filtering. *International Journal of Modern Physics B*, **31**(16–19), 1744077.
- Mittal, A., Moorthy, A. K. and Bovik, A. C.** (2012a). No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, **21**(12), 4695–4708.
- Mittal, A., Soundararajan, R. and Bovik, A. C.** (2012b). Making a “completely blind” image quality analyzer. *IEEE Signal Processing Letters*, **20**(3), 209–212.
- Nie, X., Yang, M. and Liu, R. W.** (2019). Deep Neural Network-Based Robust Ship Detection Under Different Weather Conditions. *Proceedings of the IEEE Intelligent Transportation Systems Conference*, Auckland, New Zealand.
- Ooi, C. H. and Isa, N. A. M.** (2010). Quadrants dynamic histogram equalization for contrast enhancement. *IEEE Transactions on Consumer Electronics*, **56**(4), 2552–2559.
- Petro, A. B., Sbert, C. and Morel, J. M.** (2014). Multiscale retinex. *Image Processing On Line*, **4**, 71–88.
- Pizer, S. M., Amburn, E. P., Austin, J. D., Cromartie, R., Geselowitz, A., Greer, T., ter Haar Romeny, B., Zimmerman, J. B. and Zuiderveld, K.** (1987). Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*, **39**(3), 355–368.
- Reza, A. M.** (2004). Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. *Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology*, **38**(1), 35–44.
- Senthilkumaran, N. and Thimmiraja, J.** (2014). Histogram Equalization for Image Enhancement Using MRI Brain Images. *World Congress on Computing and Communication Technologies*, Trichirappalli, India.
- Shao, Z., Wu, W., Wang, Z., Du, W. and Li, C.** (2018). Seaships: A large-scale precisely annotated dataset for ship detection. *IEEE Transactions on Multimedia*, **20**(10), 2593–2604.
- Veluchamy, M. and Subramani, B.** (2019). Image contrast and color enhancement using adaptive gamma correction and histogram equalization. *Optik*, **183**, 329–337.
- Wang, Z. and Bovik, A. C.** (2009). Mean squared error: Love it or leave it? A new look at signal fidelity measures. *IEEE Signal Processing Magazine*, **26**(1), 98–117.
- Wang, Z., Bovik, A. C., Sheikh, H. R. and Simoncelli, E. P.** (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, **13**(4), 600–612.
- Wang, S., Zheng, J., Hu, H. M. and Li, B.** (2013). Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, **22**(9), 3538–3548.
- Wang, W., Wu, X., Yuan, X. and Gao, Z.** (2020). An experiment-based review of low-light image enhancement methods. *IEEE Access*, **8**, 87884–87917.
- Wei, C., Wang, W., Yang, W. and Liu, J.** (2018). Deep retinex decomposition for low-light enhancement. arXiv:1808.04560.
- Xu, K., Yang, X., Yin, B. and Lau, R. W.** (2020). Learning to Restore Low-Light Images via Decomposition-and-Enhancement. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA.
- Zhao, L. and Shi, G.** (2019). Maritime anomaly detection using density-based clustering and recurrent neural network. *The Journal of Navigation*, **72**(4), 894–916.