

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/354542130>

# Lightweight Deep Network-Enabled Real-Time Low-Visibility Enhancement for Promoting Vessel Detection in Maritime Video Surveillance

Article in *Journal of Navigation* · September 2021

DOI: 10.1017/S0373463321000783

CITATIONS

3

READS

197

3 authors:



Yu Guo

Wuhan University of Technology

10 PUBLICATIONS 20 CITATIONS

SEE PROFILE



Yuxu Lu

Wuhan University of Technology

12 PUBLICATIONS 47 CITATIONS

SEE PROFILE



Wen Liu

Wuhan University of Technology


109 PUBLICATIONS 1,345 CITATIONS

SEE PROFILE



RESEARCH ARTICLE

# Lightweight deep network-enabled real-time low-visibility enhancement for promoting vessel detection in maritime video surveillance

Yu Guo<sup>†,1</sup>, Yuxu Lu<sup>†,1</sup> and Ryan Wen Liu<sup>1,2,\*</sup> 

<sup>1</sup> Hubei Key Laboratory of Inland Shipping Technology, School of Navigation, Wuhan University of Technology, Wuhan, China.

<sup>2</sup> National Engineering Research Center for Water Transportation Safety, Wuhan, China.

\*Corresponding author. E-mail: [wenliu@whut.edu.cn](mailto:wenliu@whut.edu.cn)

<sup>†</sup> (Yu Guo and Yuxu Lu are co-first authors)

**Received:** 2 April 2021; **Accepted:** 12 September 2021

**Keywords:** maritime video surveillance; intelligent vision; vessel detection; low-visibility enhancement; maritime safety

## Abstract

Maritime video surveillance has become an essential part of the vessel traffic services system, intended to guarantee vessel traffic safety and security in maritime applications. To make maritime surveillance more feasible and practicable, many intelligent vision-empowered technologies have been developed to automatically detect moving vessels from maritime visual sensing data (i.e., maritime surveillance videos). However, when visual data is collected in a low-visibility environment, the essential optical information is often hidden in the dark, potentially resulting in decreased accuracy of vessel detection. To guarantee reliable vessel detection under low-visibility conditions, the paper proposes a low-visibility enhancement network (termed LVENet) based on Retinex theory to enhance imaging quality in maritime video surveillance. LVENet is a lightweight deep neural network incorporating a depthwise separable convolution. The synthetically-degraded image generation and hybrid loss function are further presented to enhance the robustness and generalisation capacities of LVENet. Both full-reference and no-reference evaluation experiments demonstrate that LVENet could yield comparable or even better visual qualities than other state-of-the-art methods. In addition, it takes LVENet just 0.0045 s to restore degraded images with size 1920 × 1080 pixels on an NVIDIA 2080Ti GPU, which can adequately meet real-time requirements. Using LVENet, vessel detection performance can be greatly improved with enhanced visibility under low-light imaging conditions.

## 1. Introduction

Maritime video surveillance has become increasingly important for traffic situational awareness in a range of maritime applications (Bloisi et al., 2017). In particular, the widely-used surveillance cameras provide indispensable visual information for video surveillance to enhance transportation safety and security. As an essential function of maritime surveillance, visual data-based vessel detection has attracted considerable attention due to its practical importance. Accurate vessel detection offers excellent potential for promoting navigational safety monitoring, collision avoidance, vessel traffic services, etc. However, the quality of visual information obtained from surveillance cameras will directly affect the efficiency of vessel detection. With the rapid rise of deep learning, it is feasible to build learning-based methods that significantly enhance visual information quality in real time, potentially leading to improved vessel detection in maritime supervision. Maritime safety and security can thus be guaranteed in ports and other maritime infrastructures.

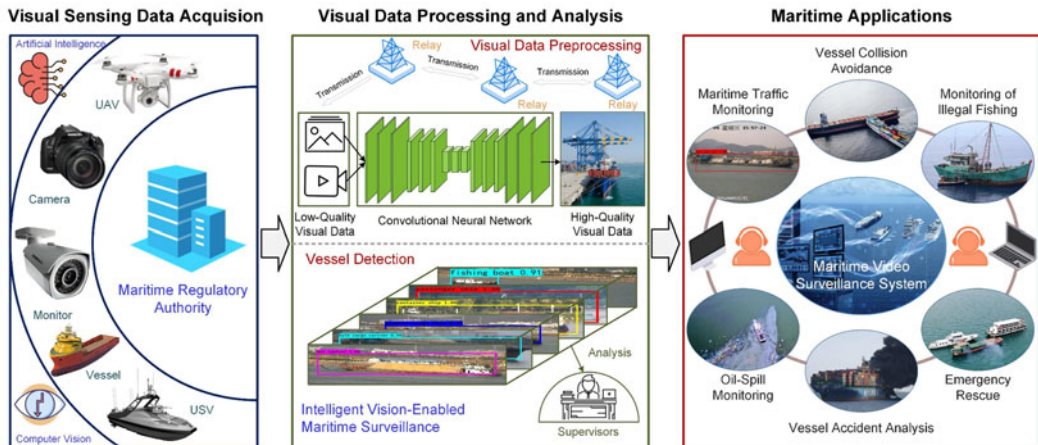
In the current literature, many vision-based technologies of vessel detection and tracking have been developed (Zhang et al., 2017; Yang et al., 2018; Liu et al., 2021a, 2021b) for effective maritime surveillance. These methods have been verified to generate satisfactory vessel detection results in normal lighting environments. Under low-light imaging conditions, however, much valuable visual information is buried in the dark, especially the moving objects of interest. To overcome the environmental effects, Nie et al. (2019) designed a vessel detection method under different weather conditions. However, the method did not enhance the essential information hidden in the dark and the study only expanded the dataset by simulating various environments to improve the robustness of the vessel detection network. Therefore, it is necessary to develop an efficient low-visibility enhancement network as the preprocessing step. The advantage of this strategy is that it can directly serve any network to detect vessels under low-light environment without retraining.

Low-visibility enhancement is a topic that has attracted sustained attention. Traditional model-based methods often design special hand-crafted priors to obtain enhanced results. For instance, Fu et al. (2013) introduced the bright channel prior (BCP), different from the popular dark channel prior originally proposed by He et al. (2010), to perform luminance estimation. Fu et al. (2016) introduced a weighted variational model, which is significantly different from previous models. It is capable of separately estimating reflection and illumination components. Subsequently, Guo et al. (2017) constructed a structure prior model to optimise the illumination component. These methods have been verified to produce satisfactory results on traditional images. However, due to the significant differences in textural structures, these methods often fail to enhance maritime images accurately under poor imaging conditions.

Furthermore, many studies on the enhancement of low-visibility maritime images have been carried out to obtain satisfactory visual performance. Yang et al. (2019a, 2019b) proposed a coarse-to-fine luminance estimation-based marine low-visibility enhancement method. Inspired by the convolutional neural network (CNN), Guo et al. (2020) constructed a low-light maritime image enhancement framework by combining traditional model estimation and deep learning. Although these methods can obtain satisfactory visual effects, it is challenging to complete real-time processing tasks due to the high computational complexity. To eliminate these potential limitations, a lightweight neural network for real-time low-visibility enhancement is proposed. Compared with previous studies, the proposed low-visibility enhancement network (LVENet) differs from other competing methods as follows.

- A CNN-enabled LVENet is designed to enhance the visual qualities of images captured under low-light imaging conditions. In particular, the network is designed based on the Retinex theory. Therefore, high-quality images can be obtained quickly with fewer calculations.
- To guarantee real-time maritime surveillance, it is proposed to replace ordinary convolutions with depthwise separable convolutions to significantly reduce network parameters and increase computational speed.
- A synthetically-degraded image generation method and a hybrid loss function are proposed to further enhance the robustness and generalisation capacities of the lightweight neural network.
- Image enhancement and vessel detection experiments under low-visibility conditions have demonstrated that the method has superior image enhancement results and can improve detection accuracy. In addition, the network obtains the best running time compared with other competing image enhancement methods.

The rest of this paper is categorised into the following sections. Section 2 presents a review of related works. In Section 3, the LVENet is introduced in detail. Network training details and numerical experiments are presented in Section 4. Finally, Section 5 summarises the main contributions and future works.



**Figure 1.** The workflow of the proposed deep learning-based maritime video surveillance system. Note that UAV and USV denote unmanned aerial vehicle and unmanned surface vehicle, respectively.

## 2. Related works

### 2.1. Maritime video surveillance system

Both the traditional remote sensing-based technology, for example, synthetic aperture radar (Yang et al., 2018; Chaturvedi, 2019; Liu et al., 2021), and the automatic identification system (AIS) (Zhang et al., 2016, 2017) have been salient achievements in maritime surveillance. However, these methods require special equipment to be installed on both vessels and observation stations. Unfortunately, some vessels are not equipped with these devices, which leads to the failure of maritime supervision. Meanwhile, some illegal vessels attempt to escape detection and surveillance by shutting down related equipment intentionally. Therefore, the maritime video surveillance system is crucial to improve maritime supervision and emergency rescue capabilities further. Though videos and images can provide more intuitive information for managers, long-term observation will cause visual fatigue and eventually vital information may be overlooked. With the emergence of deep learning technology and target detection, many high-quality vessel detection methods have been proposed, making it possible to build an autonomous maritime video surveillance system.

As shown in Figure 1, the workflow of the proposed deep learning-based maritime video surveillance system consists of three components: visual sensing data acquisition (VSDA), visual data processing and analysis (VDPA), and maritime applications. VSDA can obtain video and image information collected by various devices in the ocean, on land, and in the sky. The VDPA receives and processes visual information through special hardware, assisting supervisors to complete different maritime surveillance and management tasks comprised in maritime applications. However, the visual data collected by VSDA is easily affected by weather conditions which can directly affect the normal operation of VDPA. In previous studies, the step of visual data preprocessing is usually ignored, which will cause serious problems. For example, directly detecting the target on the video/image captured under low-light conditions will significantly reduce the accuracy of vessel detection and even result in detection failure. For the sake of better supervision, it is important to develop a low-visibility enhancement preprocessing network to promote the robustness and accuracy of vessel detection.

### 2.2. Low-visibility enhancement methods

In the current literature, low-visibility enhancement methods can be divided into three broad categories: plain methods, model-based methods, and deep learning-based methods.

### 2.2.1. Plain methods

Histogram equalisation (Pisano et al., 1998) and its improved versions (Kim, 1997; Chen and Ramli, 2003; Tan et al., 2012) belong to the classic low-light enhancement methods that can force the greyscale histogram to the full range by contrast stretching. In particular, histogram equalisation-based methods have the capacity of guaranteeing contrast enhancement. However, the contrast stretching-based methods easily cause over- and under-enhancement. To further improve image quality, gamma correction individually conducts a non-linear operation on each pixel to enhance illumination. Because the relationship between adjacent pixels is ignored, the restored images may be inconsistent with the latent sharp versions.

### 2.2.2. Model-based methods

Unlike plain methods, model-based methods focus on decomposing the image and then processing the components separately. Inspired by the popular Retinex theory (Land, 1977), the colour image  $I$  with red, green, and blue (RGB) channels can be decomposed as follows:

$$I(x) = \tilde{I}(x) * L(x), \quad (1)$$

where  $x$  is the pixel index,  $*$  denotes the element-wise multiplication operator,  $I$ ,  $\tilde{I}$ , and  $L$  are the captured low-visibility image, reflection map, and illumination map, respectively. It is well known that the reflection map contains textural details and rich colour. In contrast, the illumination map has only the luminance information with single channel. Previous attempts estimated the smoothed illumination using a Gaussian filter, for example, single-scale Retinex (Jobson et al., 1997a, 1997b), multi-scale Retinex (MSR) (Jobson et al., 1997a, 1997b), and MSR with colour restoration (Jiang et al., 2015). Though these methods can enhance the image illumination, direct reduction of illumination will cause problems of unnatural effects and over-enhancement. Wang et al. (2013) thus developed the naturalness-preserved enhancement method, which can significantly improve the image contrast and make the illumination more natural. However, under certain circumstances (excessively dark regions), the enhancement results appear an unnatural grey colour. Fu et al. (2013) presented the BCP-regularised method to generate smooth illumination. However, the image details are over-smooth due to implementing a variational approach to process the reflection. Fu et al. (2016) used a novel weighted variational model (SRIE) for estimating the reflection and illumination and found that logarithmic transformation can significantly highlight the details of dark regions. SRIE can enhance the target version by controlling the illumination and suppressing unwanted noise. To further improve the enhancement results, the low-light image enhancement framework (LIME) (Guo et al., 2017) was presented to refine the illumination by introducing a structural prior.

### 2.2.3. Deep learning-based methods

Driven by deep learning, CNN has been widely used in image processing, for example, in denoising (Lu et al., 2021), dehazing (Li et al., 2017), deraining (Yang et al., 2019a, 2019b), and super-resolution (Dong et al., 2016). Due to the strong non-linear learning capability of CNN, deep learning-based methods have been widely presented to perform low-visibility image enhancement. For instance, Lore et al. (2017) constructed an encoder-decoder structure (LLNet) to optimise the illumination without over-amplifying the lighter regions. Inspired by MSR (Jobson et al., 1997a, 1997b), Shen et al. (2017) proposed an end-to-end MSR-based network (MSRNet) to directly learn the projection between low-light and latent sharp images. However, the restored results generated by MSRNet often appear unnatural in practice. In 2018, a Retinex-enabled network (RetinexNet) (Wei et al., 2018) was proposed, which introduced a component separation method and employed a decomposition network. However, the images enhanced by RetinexNet often suffer from serious distortion problems due to inaccurate component estimation. To solve this problem, Zhang et al. (2019) proposed a strong low-light image enhancer, which can correctly deal with the issue of image distortion.

Though current studies on plain, model-, and learning-based methods have made breakthroughs, these methods fail to directly serve maritime supervision due to the enormous difference between maritime and traditional imaging scenarios. In addition, most strategies require expensive calculations and fail to perform real-time video monitoring in maritime applications.

### 2.3. Automatic vessel detection methods

Vessel detection is important in maritime video surveillance, which aims automatically to detect the moving vessels of interest in specific waters. It is able to identify abnormal vessel behaviours in a timely manner, leading to improving maritime traffic safety. In the literature, several vessel detection methods have been implemented to enable intelligent maritime surveillance. For instance, Zhu et al. (2010) proposed an intact hierarchy method to detect vessels from spaceborne optical images. The visual monitoring strategy of cage culture (Hu et al., 2011) was proposed to detect and track vessels automatically. More recently, Chen et al. (2019) adopted a multi-view learning method which is capable of extracting a highly coupled vessel appearance and shape. These methods have verified their detection performance through massive experiments, but they generally require extensive calculations. Therefore, it is still hard to detect dynamic vessels in real time.

The development of CNN has enabled many novel vessel detection methods to be designed. Wu et al. (2018) proposed a novel inshore ship detection method, which is able to estimate the location of possible ship heads and the rough ship directions through global search. Kim et al. (2018) introduced a deep learning-enabled novel probabilistic ship detection and classification system. Shao et al. (2020) designed a saliency-aware CNN framework for ship detection, which mainly includes comprehensive ship discriminative features. In addition, the hybrid kernelised correlation and anomaly cleaning were combined with tracking moving vessels from maritime visual sensing data. Meanwhile, Chen et al. (2020) proposed a coarse-to-fine cascaded CNN (CFCCNN) to distinguish vessels with similar visual appearance. These methods can satisfy the real-time process requirements on high-performance equipment and further improve vessel detection accuracy. However, in low-light imaging environments, the accuracy of vessel detection will be significantly reduced. The primary reason for worse performance is that vital information is hidden in the dark and fails to be extracted by the vessel detection network. Therefore, to improve vessel detection results in low-visibility conditions, it is necessary to improve visual quality for installed imaging cameras in a maritime video surveillance system.

## 3. LVENet

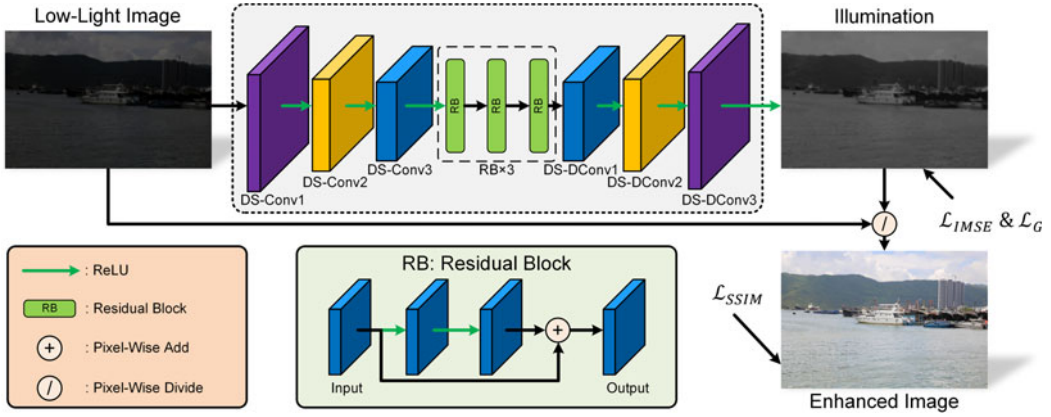
This section is dedicated to developing the proposed LVENet. The flowchart of LVENet is summarised in Figure 2. This study designs a lightweight CNN for learning the features of maritime low-visibility scenes. In particular, depthwise separable convolution is used instead of traditional convolution to reduce model parameters and improve calculation speed. Given that it is a recent advance, no research has been conducted on the adoption of depthwise separable convolution for low-light image enhancement thus far. Furthermore, a hybrid loss function is constructed to supervise the network training and enhance the network generalisation.

### 3.1. Network architecture

Inspired by the Retinex theory, the proposed LVENet also assumes that a single low-visibility image  $I$  can be decomposed into reflection  $\tilde{I}$  and illumination  $L$ . Once a satisfactory illumination component is obtained, the enhanced image (i.e., reflection component) can be easily obtained by

$$\tilde{I}(x) = \frac{I(x)}{L(x)}. \quad (2)$$





**Figure 2.** The flowchart of the proposed LVENet. DS-Conv and DS-DConv represent depthwise separable convolution and depthwise separable deconvolution, respectively.

Since the illumination component  $L$  only contains the single-channel brightness information, it is possible to obtain a stable enhancement effect with fewer calculations. Therefore, LVENet is devoted to learning the mapping between low-visibility image  $I$  and illumination component  $L$ .

The architecture of LVENet is visually shown in Figure 2. The input of the network is the three-channel low-visibility image  $I$ , and the output is the estimated illumination map  $\hat{L}$  with single channel. In particular, LVENet first adopts convolutional layers and then deconvolutional layers. The first three convolutional layers, respectively, exploit 8, 16, and 32 filters to generate feature maps. Subsequently, three residual blocks (He et al., 2016) are displayed to learn residual mapping, making it easier for the network to detect subtle differences. Mathematically, the  $k$ -th feature map of the  $m$ -th residual block is defined as follows

$$g_k^m = \mathcal{A}(\mathcal{R}_k^m + g_k^{m-1}), \quad (3)$$

where  $\mathcal{A}(\cdot)$  denotes the rectified linear units (ReLU) activation function,  $\mathcal{R}_k^m$  is the  $k$ -th output feature map generated by two convolutions of the  $m$ -th residual block, and  $g_k^{m-1}$  represents the  $k$ -th input map obtained by the  $(m-1)$ -th residual block. Finally, three deconvolutional layers are employed to guarantee that the output illumination component  $\hat{L}$  has the same spatial resolution as the input  $I$ . Note that the ReLU activation function is displayed after each convolutional layer, and all convolutional/deconvolutional layers are composed of depthwise separable convolution/deconvolution to simplify calculations and reduce model parameters. In particular, the configuration details of the LVENet have been shown in Table 1. Furthermore, the robust Adam technique (Kingma and Ba, 2014) is exploited for optimisation. The  $l$ -th updated model parameters  $\theta_l$  can be updated by

$$\theta_l = \theta_{l-1} - \partial \cdot \frac{\hat{u}_l}{\sqrt{\hat{v}_l + e}}, \quad (4)$$

where  $\partial$  and  $e$  represent the learning rate and minimal constant for avoiding outliers,  $\hat{u}_l$  and  $\hat{v}_l$ , respectively, denote the exponential decay averages of the  $(l-1)$ -th gradient  $u_{l-1}$  and the  $(l-1)$ -th square gradient  $v_{l-1}$ , and  $\theta_{l-1}$  is network parameters through the  $(l-1)$ -th updating. After obtaining the illumination component  $\hat{L}$ , the enhanced image is finally generated, i.e.,

$$\hat{I}(x) = \frac{I(x)}{\max(\hat{L}(x), \gamma)}, \quad (5)$$

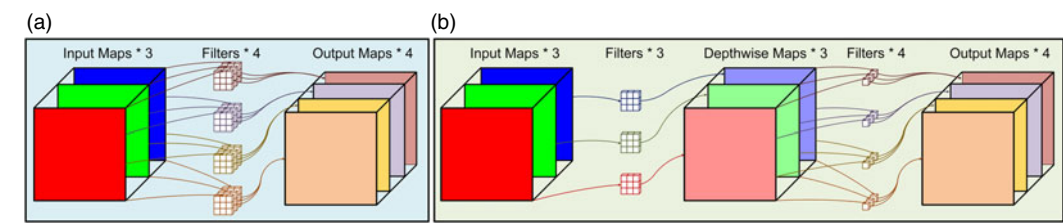
with  $\gamma$  being a particular parameter to avoid over-enhancement effect and suppress outliers. In the experiments,  $\gamma = 0.15$  was empirically selected to obtain satisfactory imaging performance.

Table 1. Configurations of proposed LVENet.

Layer Type	DS-Conv1		DS-Conv2		DS-Conv3		RB				DS-DConv1		DS-DConv2		DS-DConv3	
	C	C	C	C	C	C	C	C	C	C	TC	C	TC	C	TC	C
Size	3	1	3	1	3	1	3	1	3	1	3	1	3	1	3	1
Channel	8	8	16	16	32	32	32	32	32	32	16	16	8	8	3	3
Stride	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Pad	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0
Groups	8	1	16	1	32	1	32	1	32	1	16	1	8	1	3	1

Note: C: convolution; DC: transposed convolution.





**Figure 3.** Usage case of traditional convolution and depthwise separable convolution.

**Table 2.** Comparison of the parameter quantities between traditional convolution (Conv) and depthwise separable convolution (DS-Conv) in Figure 3.

Layer	Layer 1			Layer 2			parameter quantity
	filters	kernel	size	filters	kernel	size	
Conv	4	3	3	—	—	—	$4 \times 3 \times 3^2 = 108$
DS-Conv	3	1	3	4	3	1	$3 \times 1 \times 3^2 + 4 \times 3 \times 1^2 = 39$

**3.2. Depthwise separable convolution and deconvolution**

The powerful depthwise separable convolution was originally proposed by Howard et al. (2017). Unlike traditional convolution, depthwise separable convolution is actually composed of two convolutions. Figure 3 and Table 2 describe a classic case to distinguish the traditional convolution and depthwise separable convolution. Specifically, this case takes the three-channel feature map as input and the four-channel feature map as output. Four filters of the traditional convolutional layer simultaneously process each channel input and obtain four maps. It is worth mentioning that each filter contains three kernels of size  $3 \times 3$ . Therefore, the parameter quantity of the traditional convolution is 108.

In contrast, the depthwise separable convolution splits this process into two steps: depthwise and pointwise convolutions. Depthwise convolution is adopted to obtain the spatial characteristic information (i.e., depthwise feature) of each channel. Inspired by grouped convolution, all filters in depthwise convolution only process the feature map of the corresponding channel. It is obvious that the number of depthwise feature maps obtained by depthwise convolution is equal to the input. Like traditional convolution, pointwise convolution directly performs convolution with kernel size of  $1 \times 1$  on all depthwise feature maps. To sum up, this depthwise separable convolution only contains 39 parameters. By comparison, the depthwise separable convolution can reduce the model parameters by nearly 64% in this case.

To restore the image resolution faster, a depthwise separable deconvolution module is designed, including the depthwise deconvolution and pointwise convolution. Unlike depthwise convolution, depthwise deconvolution replaces the traditional convolution with transposed convolution (Radford et al., 2015) to construct a feature map with a larger pixel size. A  $1 \times 1$  convolution is then applied after the depthwise deconvolution for pointwise feature extraction.

**3.3. Hybrid loss function**

To promote the network performance, a hybrid loss function is introduced to constrain the estimated illumination component  $\hat{L}$  and enhanced image  $\hat{I}$ . In this work, two loss functions (i.e., the gradient loss function  $\mathcal{L}_G$  and the illumination-based mean square error loss function  $\mathcal{L}_{IMSE}$ ) are proposed to make

the illumination  $L$  and estimated illumination  $\hat{L}$  as similar as possible. The gradient loss function  $\mathcal{L}_G$  is thus defined as follows

$$\mathcal{L}_G = \sum_{x \in \Omega} (\nabla_h L(x) - \nabla_h \hat{L}(x))^2 + (\nabla_v L(x) - \nabla_v \hat{L}(x))^2, \quad (6)$$

where  $\Omega$  is the entire image domain,  $\nabla_h$  and  $\nabla_v$  represent the operators of the horizontal and vertical gradients, respectively. In this work,  $\mathcal{L}_G$  can make the estimated illumination  $\hat{L}$  obtain an edge structure that is similar to the ground truth  $L$ .

In the low-visibility enhancement task, the researchers considered that the dark region of the estimated version should receive more attention to ensure the effectiveness of enhancement. Thus, a simple method was proposed to divide the whole estimated illumination component  $\hat{L}$  into two parts: the high illumination component  $\hat{L}_h$  and the low illumination component  $\hat{L}_l$ . Since the single-channel illumination only contains the luminance information of the scene, a specific value  $\tau$  was thus chosen to guarantee robust estimation. When the value of a certain pixel in  $\hat{L}$  (i.e., illumination of the pixel) is higher than  $\tau$ , it is regarded as  $\hat{L}_h$ ; while it is regarded as  $\hat{L}_l$  otherwise. The final  $\mathcal{L}_{\text{IMSE}}$  can be thus written as follows

$$\mathcal{L}_{\text{IMSE}} = \lambda_l \sum_{x \in \Omega} (L_l(x) - \hat{L}_l(x))^2 + \lambda_h \sum_{x \in \Omega} (L_h(x) - \hat{L}_h(x))^2, \quad (7)$$

where  $L_l$  and  $L_h$ , respectively, denote the ground truth corresponding to  $\hat{L}_l$  and  $\hat{L}_h$ ,  $\lambda_l$  and  $\lambda_h$  represent the trade-off parameters. According to the comparison experiments, the reliable coefficients, i.e.,  $\tau = 0.4$ ,  $\lambda_l = 0.8$  and  $\lambda_h = 0.2$ , were chosen to generate the satisfactory low-visibility enhancement results. To preserve the essential structure, illumination, and contrast in the final enhanced image  $\hat{I}$ , the structural similarity loss function  $\mathcal{L}_{\text{SSIM}}$  is exploited, i.e.,

$$\mathcal{L}_{\text{SSIM}} = 1 - \text{SSIM}(\tilde{I}, \hat{I}), \quad (8)$$

where  $\tilde{I}$  is the original clear image,  $\text{SSIM}(\cdot)$  denotes the calculation operation of structural similarity (SSIM). The terms of the SSIM metric will be explained in Section 4. To sum up, the hybrid loss function  $\mathcal{L}$  of the network can be defined as follows

$$\mathcal{L} = \lambda_1 \mathcal{L}_G + \lambda_2 \mathcal{L}_{\text{IMSE}} + \lambda_3 \mathcal{L}_{\text{SSIM}}, \quad (9)$$

where  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  denote the penalty coefficients. These were set as:  $\lambda_1 = 0.5$ ,  $\lambda_2 = 1.3$ , and  $\lambda_3 = 2.0$ .

## 4. Experimental results and analysis

This section introduces the details of implementation of network training. All the qualitative and quantitative experiments conducted are also discussed. Note that all experiments and training were performed in Python 3.7 and Matlab 2019a environment running on a PC with Intel (R) Core (TM) i5-10600KF CPU @ 4.10 GHz and a Nvidia GeForce RTX 2080 Ti GPU.

### 4.1. Evaluation metric

To fully evaluate the enhancement performance, five full-reference evaluation metrics and two no-reference metrics were employed in the experiments. In particular, first five full-reference metrics were adopted to measure the enhanced and latent sharp images: PSNR, SSIM, FSIM, FSIMc, and VSI. Two no-reference metrics, NIQE and BTMQL, are then exploited to blindly evaluate the quality of enhanced image. Finally, the mean average precision (mAP) is introduced to evaluate the accuracy of vessel detection. The definitions of these evaluation metrics are given as follows:

- PSNR: *Peak signal-to-noise ratio* (Wang and Bovik, 2009) is a widely-used evaluation metric to measure image quality. The PSNR value between the restored image  $\hat{Y}$  and the target version  $Y$  is given by

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

where  $M$  is the maximum pixel value, and  $\text{MSE}(\hat{Y}, Y)$  represents the operation of calculating the mean square error (MSE) between  $\hat{Y}$  and  $Y$ .

- SSIM: *Structural similarity* (Wang et al., 2004) is proposed based on the assumption that the human visual system can judge image structure similarity objectively. SSIM consists of three main components: luminance comparison  $l(\hat{Y}, Y)$ , contrast comparison  $c(\hat{Y}, Y)$ , and structure comparison  $s(\hat{Y}, Y)$ . The mathematical expression of SSIM can be extracted as

$$\text{SSIM}(\hat{Y}, Y) = \frac{(2\mu_{\hat{Y}}\mu_Y + c_1)(2\sigma_{\hat{Y}Y} + c_2)}{(\mu_{\hat{Y}}^2 + \mu_Y^2 + c_1)(\sigma_{\hat{Y}}^2 + \sigma_Y^2 + c_2)}, \quad (11)$$

where  $\mu_{\hat{Y}}$  and  $\mu_Y$ , respectively, represent the mean values of  $\hat{Y}$  and  $Y$ ,  $\sigma_{\hat{Y}}$  and  $\sigma_Y$  denote the corresponding standard deviations,  $\sigma_{\hat{Y}Y}$  is the covariance value.  $c_1$  and  $c_2$  are special constants to prevent outliers.

- FSIM and FSIMc: *Feature similarity* (Zhang et al., 2011) considers all pixels in an image to have different importance. For instance, pixels at the edge of an object are more important than pixels in other background regions. Therefore, the improvement direction of this evaluation index focuses on distinguishing necessary pixels and giving them appropriate weights. FSIM is similar to SSIM, coupled with the phase congruency term  $S_{pc}$  and gradient magnitude term  $S_G$ . The definition of FSIM is described as follows

$$\text{FSIM}(\hat{Y}, Y) = \frac{\sum_{x \in \Omega} S_{pc}(x) S_G(x) \max(\text{PC}_{\hat{Y}}(x), \text{PC}_Y(x))}{\sum_{x \in \Omega} \max(\text{PC}_{\hat{Y}}(x), \text{PC}_Y(x))}, \quad (12)$$

with  $\text{PC}_{\hat{Y}}$  and  $\text{PC}_Y$  being the phase congruency of  $\hat{Y}$  and  $Y$ . On the basis of FSIM, FSIMc further considers chromaticity information  $S_C(x)$  and uses  $\lambda$  to further adjust the importance of chromatic components, i.e.,

$$\text{FSIMc}(\hat{Y}, Y) = \frac{\sum_{x \in \Omega} S_{pc}(x) S_G(x) S_C^\lambda(x) \max(\text{PC}_{\hat{Y}}(x), \text{PC}_Y(x))}{\sum_{x \in \Omega} \max(\text{PC}_{\hat{Y}}(x), \text{PC}_Y(x))}. \quad (13)$$

- VSI: *Visual saliency-induced index* (Zhang et al., 2014) considers that distortion will cause visual saliency changes, which strongly correlates with distortion. Therefore, the GBAS model-based visual saliency map (VS map) is used to evaluate the distortion. The mathematical definition of VSI between  $\hat{Y}$  and  $Y$  is given by

$$\text{VSI}(\hat{Y}, Y) = \frac{\sum_{x \in \Omega} S(x) \max(\text{VS}_{\hat{Y}}(x), \text{VS}_Y(x))}{\sum_{x \in \Omega} \max(\text{VS}_{\hat{Y}}(x), \text{VS}_Y(x))}, \quad (14)$$

where  $S$  denotes the local similarity of  $\hat{Y}$  and  $Y$ ,  $\text{VS}_{\hat{Y}}$  and  $\text{VS}_Y$  are the VS maps corresponding to  $\hat{Y}$  and  $Y$ .

- NIQE: *Natural image quality evaluator* (Mittal et al., 2013) is proposed according to the statistical law observed in natural images. NIQE is a quality perception collection of statistical features constructed based on a simple and successful airspace natural scene statistics model.
- BTMqi: *Blind tone-mapped quality index* (Gu et al., 2016) is an efficient and effective no-reference objective quality metric. In particular, BTMqi can automatically evaluate standard low dynamic

range images created by different tone mapping operators without accessing the original high dynamic range image.

- mAP: *Mean average precision* (Everingham et al., 2010) is an evaluation index widely used in target detection. mAP is obtained by calculating the mean value of average precision for all classes. In this work, average precision can be obtained by the area under the precision/recall curve of detections.

Note that more details of the implementation of PSNR, SSIM, FSIM/FSIMc, VSI, NIQE, BTMQL, and mAP can be found in the literature (Wang et al., 2004; Wang and Bovik, 2009; Everingham et al., 2010; Zhang et al., 2011, 2014; Mittal et al., 2013; Gu et al., 2016). To obtain better enhancement performance, the values of full-reference metrics should become higher, whereas the values of no-reference metrics should be lower. In addition, higher mAP indicates more accurate detection of moving vessels in maritime applications.

#### 4.2. Synthetically-degraded image generation

Though many datasets that pair low-light and clear images already exist, a network model trained by these datasets cannot perform maritime surveillance effectively due to the distinctive nature of maritime imagery. Meanwhile, it is challenging to collect paired low-light/clear maritime images. Therefore, it is crucial to generate more realistic low-visibility images based on clear maritime surveillance images. The traditional low-visibility image generation method simply multiplies all image pixels by a definite coefficient, which causes all regions of the image to be darkened to the same degree. Unlike synthetic images, natural low-light images usually have both bright and dark regions. Therefore, an assumption was proposed that applying a stronger darkening factor (i.e., lower value) to the darker regions can make the composite image more realistic. Inspired by this prior, a simple but effective illumination estimation-based method was designed to obtain the darkening weight of each pixel. Specifically, the maximum value of each pixel of the RGB image  $\tilde{I}$  in the three channels (i.e., R, G, B) was found, which can be written as

$$\bar{L}(x) = \max_{C \in \{R, G, B\}} \tilde{I}^C(x), \quad (15)$$

with  $\tilde{I}^C$  and  $\bar{L}$  being the single-channel image of  $\tilde{I}$  and coarse illumination component. Subsequently, a guided filter was adopted to smooth the details and preserve the significant edges as much as possible, which can be given by

$$\hat{L}(x) = G_*(\tilde{I}^{\text{Gray}}(x)) \otimes \bar{L}(x), \quad (16)$$

where  $\hat{L}$  and  $\tilde{I}^{\text{Gray}}$  represent the darkening weight (i.e., refined illumination component) and guided image (i.e., grayscale image corresponding to  $\tilde{I}$ ),  $\otimes$  and  $G_*$  are the convolutional operator and guided filter associated with  $\tilde{I}^{\text{Gray}}$ . In this paper, the local window radius and regularisation parameter are respectively set as 7 and  $10^{-3}$ . More details on the description of guided filter can be found in He et al. (2013). Finally, the specific darkening coefficient  $\varepsilon$ , weight  $\hat{L}$ , and clear image  $\tilde{I}$  were multiplied to obtain the final synthetically-degraded image, i.e.,

$$I(x) = \varepsilon \hat{L}(x) * \tilde{I}(x). \quad (17)$$

#### 4.3. Implementation details

To test the implementation of the proposed model for maritime surveillance, 1,000 images were selected and captured as a training dataset containing 800 SeaShips images (Shao et al., 2018), 100 outdoor images chosen by the MIT-Adobe FiveK dataset (Bychkovsky et al., 2011), and 100 maritime images captured by the authors' SLR camera. Then 8,000 patches with the size of  $256 \times 256$  were generated using cropping, rotating, and scaling. In the numerical experiments, the training epoch was set as 60. During training, the learning rate was set to  $10^{-3}$  in the first 30 epochs and  $10^{-4}$  in the last 30 epochs. In the

training, the inputs were obtained by Equation (17) mentioned in Section 4.1 with darkening coefficient  $\varepsilon$  ranging between (0.4, 0.8). It took about 3 h to train the low-visibility enhancement network LVENet with the Pytorch 1.5.0. The Python source code is available at <https://github.com/gy65896/LVENet>.

#### 4.4. Comparisons with other competing methods

In this work, the model-based methods of BCP (Fu et al., 2013), JIEP (Cai et al., 2017), and SRIE (Fu et al., 2016) and the deep learning-based methods of RetinexNet (Wei et al., 2018), MBLLEN (Lv et al., 2018), LightenNet (Li et al., 2018), and KinD (Zhang et al., 2019) are involved as the competitors on image enhancement experiments. To ensure a fair comparison, the optimal parameters of these state-of-the-art methods are directly adopted according to the authors' codes. Furthermore, YOLOv4 (Bochkovskiy et al., 2020), a well-designed target detection network, is used to analyse the improvement of detection effect by low-visibility enhancement.

- BCP: *Bright channel prior* (Fu et al., 2013). This method proposes a BCP-regularised variational framework for low-visibility enhancement. An alternating direction optimisation method is adopted to effectively handle the resulting minimisation problem. In particular, BCP can effectively achieve low-visibility enhancement through an alternate direction optimisation-based method.
- JIEP: *Joint intrinsic-extrinsic prior* (Cai et al., 2017). This method has the capacity of jointly estimating both illumination and reflection components from an observed image. It is able to capture the luminance by illumination prior, estimate the reflectance with rich details by texture prior, and preserve the structural features by shape prior.
- SRIE: *Simultaneous reflectance and illumination estimation* (Fu et al., 2016). SRIE is proposed for better prior representation based on the logarithmic transformation. This method can effectively retain more details and suppress unwanted noise.
- RetinexNet: *Deep Retinex decomposition-based network* (Wei et al., 2018). Based on the assumption that original images can be decomposed into illumination and reflection components, a deep Retinex-based network is presented to enhance imaging quality. RetinexNet uses a decomposition network and an illumination adjustment network to enhance the reflectance and illumination components, respectively.
- MBLLEN: *Multi-branch low-light enhancement network* (Lv et al., 2018). MBLLEN is devoted to simultaneously handling various factors, including artefact, contrast, brightness, and noise. It employs different modules to extract abundant features and perform enhancement via multiple subnets. Finally, the enhanced image can be generated through multi-branch fusion.
- LightenNet: *LightenNet for weakly illuminated enhancement* (Li et al., 2018). LightenNet is a trainable CNN for enhancement of weakly illuminated images. In particular, LightenNet takes the weakly illuminated image as an input and outputs its related illumination. The final enhanced image can thus be obtained according to Retinex theory.
- KinD: *Kindling the darkness* (Zhang et al., 2019). KinD first exploits a simple and effective module to decompose the original image into illumination and reflection components. Two sub-networks are then designed to enhance the illumination and reflection components separately. These two enhanced components are finally combined to generate the latent sharp image.
- YOLOv4: *You only look once v4* (Bochkovskiy et al., 2020). YOLOv4 is composed of four parts: CSPDarknet53 backbone, PANet path-aggregation neck, SPP additional module, and YOLOv3 anchor-based head. In this paper, this model is trained for 50 epochs and uses 7,000 SeaShips images (Shao et al., 2018) as the dataset.

#### 4.5. Full-reference image quality assessment

This subsection is devoted to verifying the superior performance of the proposed method. Specifically, 27 clear maritime images were randomly selected, as shown in Figure 4, for the experiments, and 81



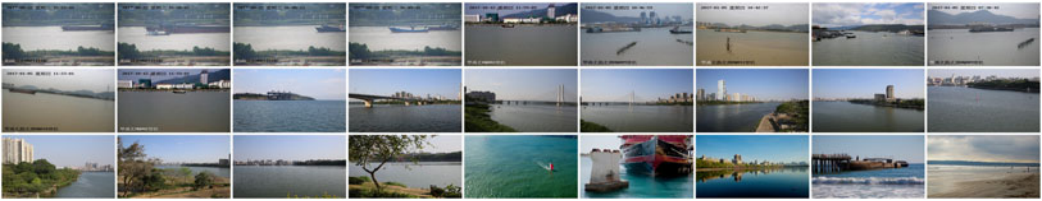


Figure 4. Twenty-seven selected sharp maritime images for synthetic experiments.

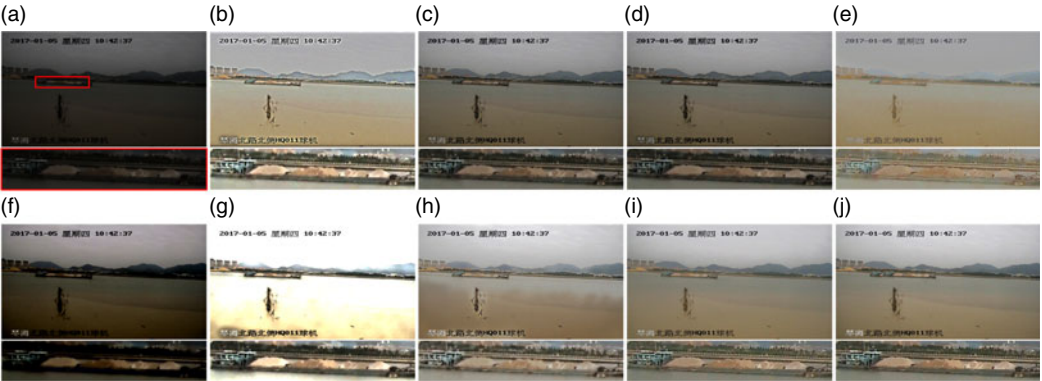
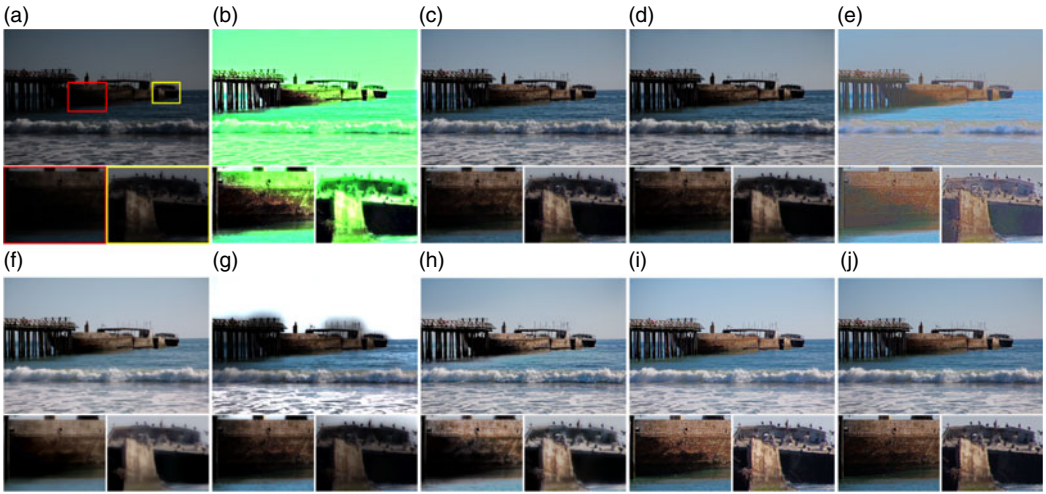


Figure 5. Comparisons of synthetic experiments on one image from Figure 4. From top-left to bottom-right: (a) synthetic low-light image with  $\varepsilon = 0.5$ , enhanced versions obtained by (b) BCP, (c) JIEP, (d) SRIE, (e) RetinexNet, (f) MBLLEN, (g) LightenNet, (h) KinD, (i) LVENet, and (j) ground truth (GT), respectively.

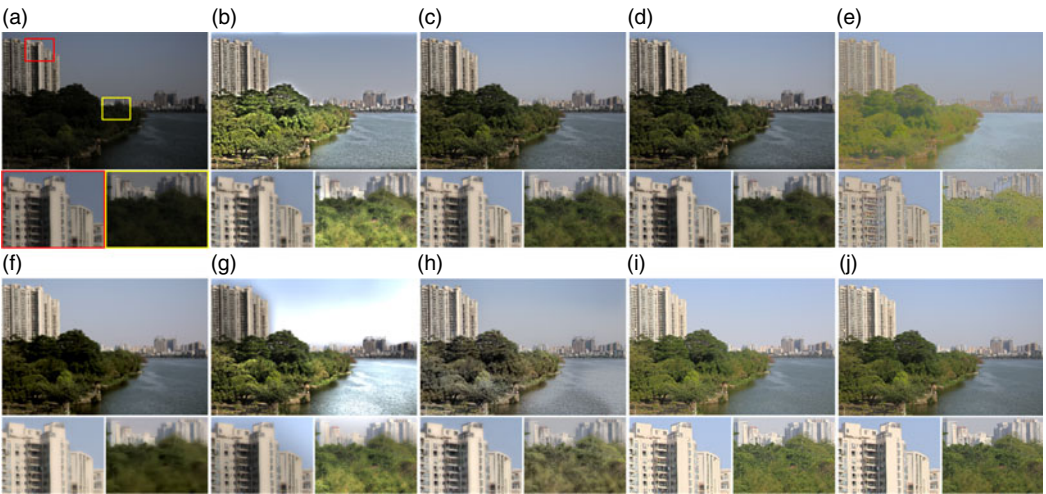
low-visibility images were synthesised using Equation (17) with  $\varepsilon = \{0.5, 0.6, 0.7\}$ . Furthermore, three model-based methods (i.e., BCP, JIEP, and SRIE) and four learning-based methods (i.e., RetinexNet, MBLLEN, LightenNet, and SRIE) were also introduced to compete with the LVENet.

To make better visual comparisons, three typical synthetic low-visibility images generated from clear images (Figure 4) and their corresponding enhanced versions obtained by various methods are shown in Figures 5–7. It can be clearly found that BCP usually causes distortion and artefact problems. Especially in Figure 6, the enhanced version obtained by BCP has serious colour abnormalities. Although JIEP, SRIE, and MBLLEN can guarantee natural visual effects, these enhanced images have the problem of insufficient enhancement. Meanwhile, it can be clearly observed from the magnified regions in Figure 7 that MBLLEN makes the enhanced image excessively smooth. Unlike JIEP, SRIE, and MBLLEN, LightenNet has the problem of over-enhancement and a black halo exists on the edges of the object of interest. RetinexNet and KinD cause different degrees of colour distortion and unnaturalness in the image. The reason behind these phenomena may be that the competing methods fail to extract the structural features in maritime images. In contrast, the proposed network is capable of learning more meaningful features, resulting in effectively restoring visual colour and details in low-light regions.

To quantitatively evaluate the enhancement effect of restored images, five full-reference evaluation metrics (PSNR, SSIM, FSIM, FSIMc, and VSI) are introduced. The calculation results of each metric are shown in Table 3. Meanwhile, the data in Table 3 are visualised as a Kiviati diagram shown in Figure 8 to compare the enhanced performance more intuitively. It can be found that the proposed method is the optimal approach in the calculation results of all five full-reference metrics. The superior performance benefits from imposing more significant penalties on dark pixels and adopting maritime images as datasets during training.



**Figure 6.** Comparisons of synthetic experiments on one image from [Figure 4](#). From top-left to bottom-right: (a) synthetic low-light image with  $\varepsilon = 0.6$ , enhanced versions obtained by (b) BCP, (c) JIEP, (d) SRIE, (e) RetinexNet, (f) MBLLEN, (g) LightenNet, (h) KinD, (i) LVENet, and (j) ground truth (GT), respectively.



**Figure 7.** Comparisons of synthetic experiments on one image from [Figure 4](#). From top-left to bottom-right: (a) synthetic low-light image with  $\varepsilon = 0.7$ , enhanced versions obtained by (b) BCP, (c) JIEP, (d) SRIE, (e) RetinexNet, (f) MBLLEN, (g) LightenNet, (h) KinD, (i) LVENet, and (j) ground truth (GT), respectively.

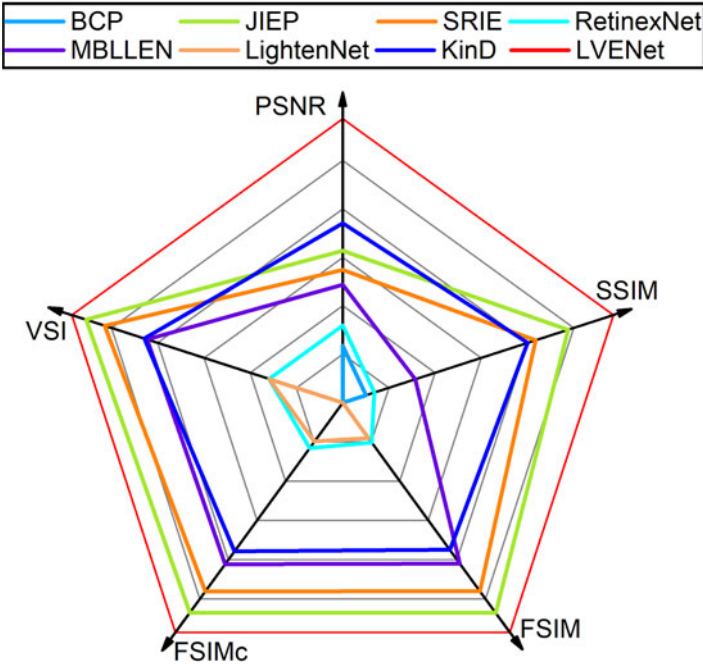
#### 4.6. No-reference image quality assessment

This subsection is dedicated to verifying the effectiveness of the proposed method for realistic low-light maritime images. In this work, the LVENet is compared with seven different image enhancement methods: BCP, JIEP, SRIE, RetinexNet, MBLLEN, LightenNet, and KinD. Specifically, the LVENet and other competitive methods are exploited to enhance three real low-visibility maritime examples. The enhancement results and their associated magnified versions are shown in [Figure 9](#). Meanwhile, two evaluators (i.e., NIQE and BTMQUI) are adopted to compare all the imaging methods.



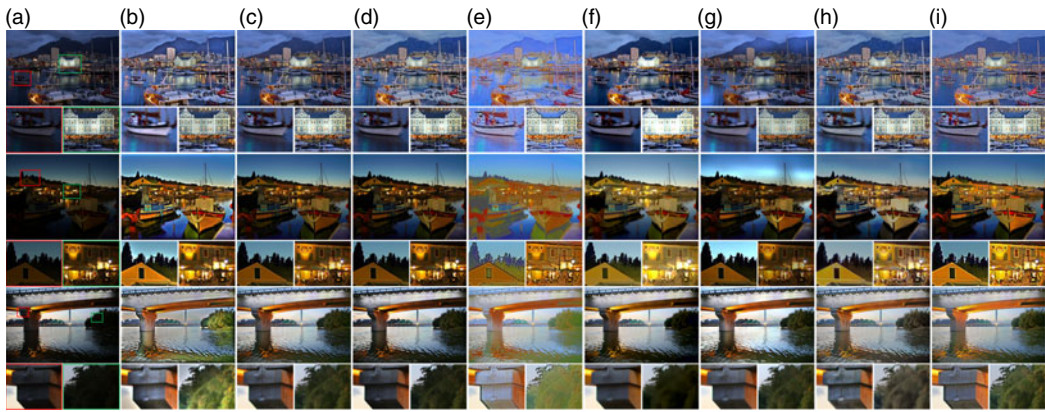
**Table 3.** Comparison of LVENet with PSNE, SSIM, FSIM, FSIMc, and VSI (mean±std) framework and state-of-the-art methods on 81 synthetic images. The best, second best, and third best results are highlighted in red, blue, and green, respectively.

Method	PSNR	SSIM	FSIM	FSIMc	VSI
BCP	15.53±3.00	0.772±0.082	0.841±0.035	0.835±0.043	0.945±0.018
JIEP	20.40±2.81	0.939±0.027	0.972±0.012	0.971±0.012	0.993±0.003
SRIE	19.42±2.57	0.912±0.037	0.958±0.017	0.957±0.017	0.989±0.005
RetinexNet	16.58±1.23	0.779±0.075	0.866±0.038	0.865±0.038	0.959±0.017
MBLLEN	18.67±4.43	0.813±0.064	0.941±0.017	0.940±0.017	0.981±0.006
LightenNet	12.63±3.55	0.753±0.087	0.863±0.040	0.860±0.040	0.959±0.016
KinD	21.79±2.27	0.905±0.039	0.933±0.015	0.931±0.016	0.982±0.005
LVENet	27.19±3.73	0.977±0.011	0.984±0.007	0.984±0.007	0.996±0.002



**Figure 8.** Kiviati diagram that visualises the calculation results from Table 3. All calculated values of each metric are normalised, i.e., the best value is 1 and the worst is 0.

Through visual comparison, it can be observed that BCP tends to produce an unnatural white halo on the target of interest and is accompanied by a certain degree of colour distortion. The restored images of JIEP and SRIE seem very natural, but their insufficient enhancement may make it hard to find essential information. Although RetinexNet has successfully enhanced the illumination, it is evident that the colour of the enhanced versions becomes extremely unnatural, which will directly lead to the failure of vessel detection. MBLLEN can obtain satisfactory visual effects, however, the improved images of MBLLEN have the problem of insufficient enhancement and the risk of over-smoothing. The combined effect of these two issues may directly erase valuable information hidden in the dark. The enhanced versions of LightenNet have the problem of overexposure and boundary artefacts in certain regions. Although KinD can achieve almost the same visual effect as the proposed method through expensive calculations, both no-reference metrics show that the proposed recovery effect has better performance.



**Figure 9.** Comparisons of realistic experiments on images 1–3 (from top to bottom: image 1, image 2, and image 3). From left to right: (a) low-light image, and enhanced versions obtained by (b) BCP, (c) JIEP, (d) SRIE, (e) RetinexNet, (f) MBLLEN, (g) LightenNet, (h) KinD, and (i) LVENet, respectively.

**Table 4.** NIQE and BTMQUI comparisons of LVENet with other methods on all test images in Figure 9. The best, second best, and third best results are highlighted in red, blue, and green, respectively.

Method	Image 1		Image 2		Image 3		mean±std	
	NIQE	BTMQUI	NIQE	BTMQUI	NIQE	BTMQUI	NIQE	BTMQUI
Low-Light	3.351	5.070	3.637	7.172	3.073	4.205	3.353±0.230	5.482±1.246
BCP	3.052	3.899	2.584	2.971	2.699	4.264	2.778±0.337	3.711±0.545
JIEP	2.854	3.461	2.871	4.394	2.621	2.486	2.782±0.114	3.447±0.779
SRIE	3.075	2.891	2.937	4.676	3.018	2.902	3.010±0.057	3.490±0.839
LightenNet	3.771	3.665	4.722	1.918	2.747	3.705	3.747±0.807	3.096±0.833
RetinexNet	3.768	2.844	3.249	2.556	4.251	3.245	3.756±0.409	2.881±0.283
MBLLEN	2.937	3.192	2.922	4.312	3.865	2.845	3.241±0.441	3.450±0.626
KinD	3.311	3.195	3.524	3.010	2.595	2.723	3.143±0.397	2.976±0.194
LVENet	2.843	2.775	2.830	2.999	2.581	2.649	2.751±0.121	2.808±0.145

As shown in Table 4, the LVENet has superior performance on both NIQE and BTMQUI metrics in most cases. Besides, Figure 10 shows more cases of low-visibility maritime images obtained by this method, illustrating that the proposed LVENet can effectively obtain buried important information.

#### 4.7. Vessel detection after low-visibility enhancement

In the maritime video surveillance system, the low-visibility environment seriously reduces the accuracy of manual or autonomous vessel detection. The LVENet is designed to solve the problems of detection failure and inaccurate identification. To verify that the LVENet can improve detection performance, YOLOv4 (Bochkovskiy et al., 2020) is employed to detect vessels in both original low-visibility images and their enhanced versions obtained by LVENet.

Figure 11 shows the detection effects of synthetic low-visibility images, enhanced images, and normal visibility images (i.e., original image). By comparison, detection failures and errors often occur in low-visibility environments. However, the enhanced images can achieve almost the same detection effect as the normal visibility image. Meanwhile, Figure 12 shows the visual results of vessel detection on the real low-visibility images and the enhanced images. It can be clearly seen that the preprocessing

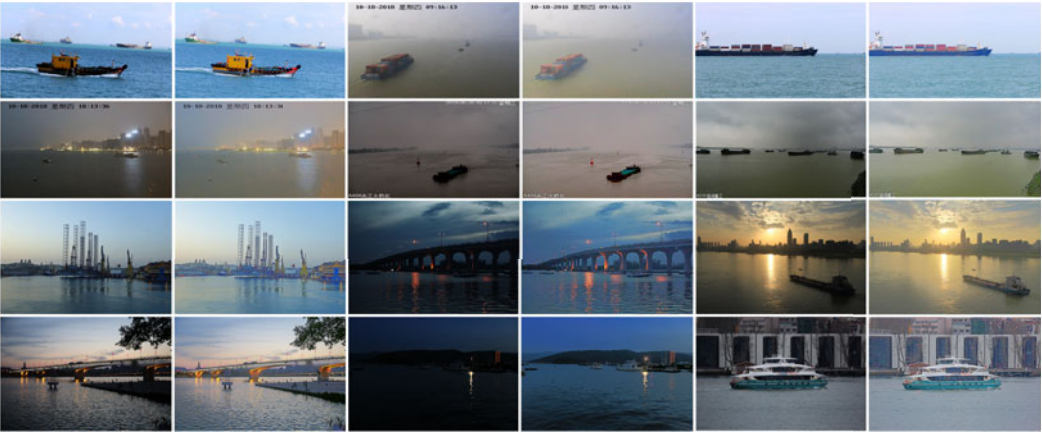


Figure 10. More low-visibility enhancement cases generated by the proposed LVENet.

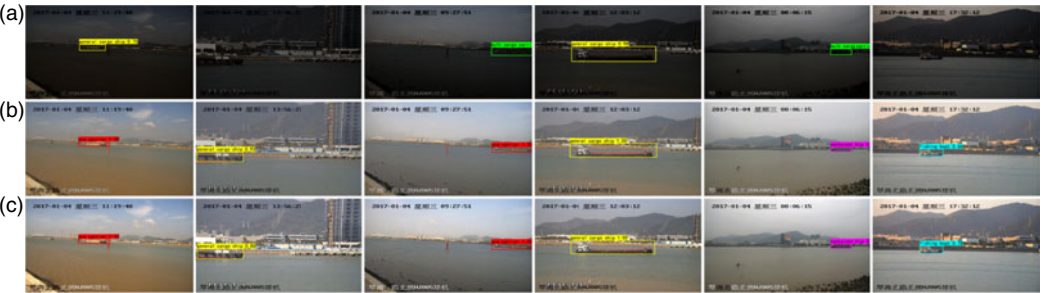


Figure 11. Vessel detection experiments on synthetic low-visibility maritime images. From top to bottom: the vessel detection results on (a) low-light images, (b) enhanced images obtained by LVENet, and (c) ground truth (GT), respectively.



Figure 12. Vessel detection experiments on real low-visibility maritime images. From top to bottom: the vessel detection results on (a) low-light images and (b) enhanced images obtained by LVENet, respectively.

of images through the LVENet can significantly improve the detection performance in natural low-visibility maritime environments. In low-light conditions, YOLOv4 tends to distinguish vessel types incorrectly and even produces the problem of detection failure. However, using the proposed method as preprocessing module can enable the vessel detection model to collect features better and to improve accuracy. Furthermore, Table 5 qualitatively displays the detection performance of all images using mAP indicators to further reveal the improvement of the LVENet for vessel detection effect in a low-visibility environment.

**Table 5.** Comparison of mAP results (unit: %).

Method	Dataset (Syn)	Dataset (Real)	Dataset (All)
Low-Light	45.83%	7.74%	37.28%
LVENet	70.83%	64.29%	57.96%
GT	71.62%	—	—

Note: Dataset (Syn), Dataset (Real), and Dataset (All) contain 18 images in Figure 11, 12 images in Figure 12, and 30 images in Figures 11 and 12, respectively.

**Table 6.** Comparison of running times (unit: second) and model parameters of several image enhancement methods.

Method	Platform	Parameters	Image size		
			480 × 640	720 × 1080	1080 × 1920
BCP	Matlab	—	0.8881	2.2966	6.3425
JIEP	Matlab	—	4.9418	13.2827	22.3051
SRIE	Matlab	—	10.4099	21.6104	26.4617
LightenNet	Matlab (Caffe)	28,929	4.0522	11.6948	32.9052
RetinexNet	Python (Tensorflow)	8,536,655	0.0978	0.2460	0.9082
MBLLEN	Python (Tensorflow)	450,171	0.1004	0.2516	0.5762
KinD	Python (Tensorflow)	35,963,861	0.0577	1.3596	2.7646
LVENet	Python (Pytorch)	6,351	0.0038	0.0041	0.0045

Note: RetinexNet, MBLLEN, KinD, and the proposed LVENet method are accelerated by 2080 Ti GPU.

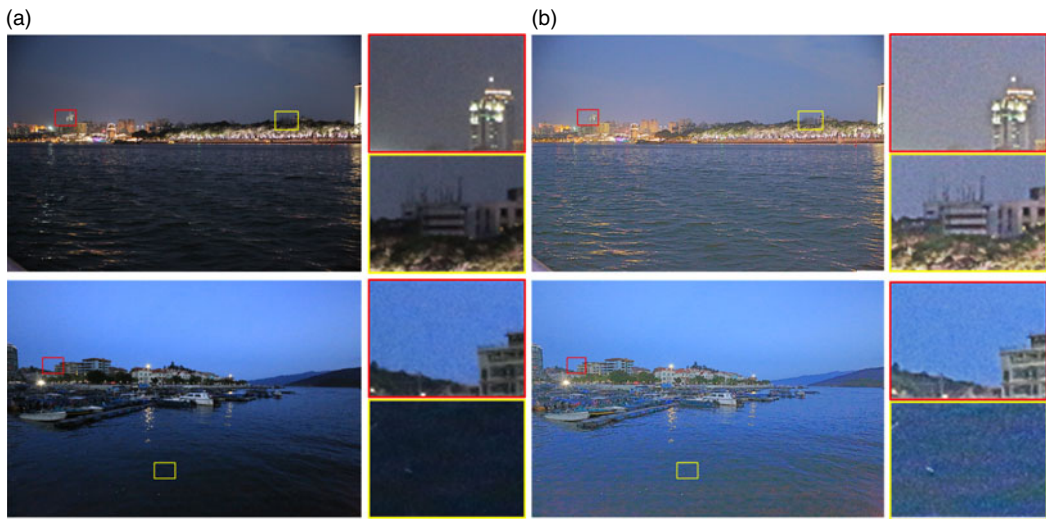
**4.8. Running time analysis**

The lightweight structure of LVENet leads to fast low-visibility enhancement. In this section, LVENet is sufficiently compared with seven image enhancement methods. It is noted that other competing imaging methods are performed using Matlab, Matlab (Caffe), and Python (Tensorflow) platforms, according to the original codes provided by the authors. To analyse the network complexity of deep learning-based approaches, the model parameters of all methods are counted. Meanwhile, three types of low-visibility datasets with sizes of 480 × 640, 720 × 1080, and 1080 × 1920 are selected for all methods to run. In particular, each dataset contains 20 images. The model parameters and average calculation times of the three scale images are shown in Table 6. Despite other slower Matlab implementations, it is fair to compare RetinexNet, MBLLEN, KinD, and the LVENet. The results illustrate that the proposed method takes first place by a large margin in terms of calculation speed. LVENet takes just 1 s to process 222 (1920 × 1080) images under the acceleration of 2080Ti GPU.

**5. Conclusion and future work**

This paper proposes a low-visibility enhancement network (termed LVENet) that directly reconstructs enhanced images via a lightweight CNN. In particular, it is inspired by the Retinex theory to estimate the illumination component using a depthwise separable convolutional network. Finally, the enhanced image is generated by dividing the input low-visibility image with the estimated illumination. In addition, a novel synthetically-degraded image strategy and a hybrid loss function were also designed to improve network performance and provide maritime images as training datasets to support better maritime surveillance. The benefit of the LVENet is that it requires relatively few calculations to obtain a





**Figure 13.** Failure enhancement cases. From left to right: (a) low-light image, and (b) enhanced versions obtained by LVENet, respectively.

satisfactory visual enhancement effect. Its superior performance has been verified on both full and no-reference assessments. Furthermore, experiments on detecting vessels after low-visibility enhancement and computing running time were also conducted to prove the practicality and real-time capability of the proposed LVENet.

Although the effectiveness of the method has been proven in many experiments, it is sometimes difficult to obtain perfect visual effects under certain circumstances. Figure 13 visually illustrates two failed cases related to low-visibility enhancement. The proposed method enhanced the brightness, but the enhanced images are still disturbed by unwanted noise carried in the input low-visibility images. However, it is still worthy of consideration since it enables better imaging results than other competitive methods. In future work, the authors will focus on developing a noise-insensitive learning method to further enhance image quality and vessel detection in maritime videos. The authors believe that there is significant potential to exploit the proposed method to guarantee the safety of vessel navigation under maritime video surveillance.

**Acknowledgements.** This work was supported by the NSFC (Grant No. 51609195) and the National Key R&D Program of China (Grant No. 2018YFC0309602).

## References

- Bloisi, D. D., Previtali, F., Pennisi, A., Nardi, D. and Fiorini, W. (2017). Enhancing automatic maritime surveillance systems with visual information. *IEEE Transactions on Intelligent Transportation Systems*, **18**(4), 824–833.
- Bochkovskiy, A., Wang, C. Y. and Liao, H. Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. arXiv: 2004.10934.
- Bychkovsky, V., Paris, S., Chan, E. and Durand, F. (2011). Learning Photographic Global Tonal Adjustment with Database of Input/Output Image Pairs. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Colorado Springs, CO, USA.
- Cai, B., Xu, X., Guo, K., Jia, K., Hu, B. and Tao, D. (2017). A Joint Intrinsic-Extrinsic Prior Model for Retinex. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy.
- Chaturvedi, S. K. (2019). Study of synthetic aperture radar and automatic identification system for ship target detection. *Journal of Ocean Engineering and Science*, **4**(2), 173–182.
- Chen, S. D. and Ramli, A. R. (2003). Minimum mean brightness error bi-histogram equalization in contrast enhancement. *IEEE Transactions on Consumer Electronics*, **49**(4), 1310–1319.
- Chen, X., Wang, S., Shi, C., Wu, H., Zhao, J. and Fu, J. (2019). Robust ship tracking via multi-view learning and sparse representation. *The Journal of Navigation*, **72**(1), 176–192.

- Chen, X., Yang, Y., Wang, S., Wu, H., Tang, J., Zhao, J. and Wang, Z. (2020). Ship type recognition via a coarse-to-fine cascaded convolution neural network. *The Journal of Navigation*, **73**(4), 813–832.
- Dong, C., Chen, C. L., He, K. and Tang, X. (2016). Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **38**(2), 295–307.
- Everingham, M., Gool, L. V., Williams, C. K. I., Winn, J. and Zisserman, A. (2010). The PASCAL visual object classes (VOC) challenge. *International Journal of Computer Vision*, **88**, 303–338.
- Fu, X., Zeng, D., Huang, Y., Ding, X. and Zhang, X. P. (2013). A Variational Framework for Single Low Light Image Enhancement Using Bright Channel Prior. *Proceedings of the IEEE Global Conference on Signal and Information Processing*, Austin, TX, USA.
- 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.
- Gu, K., Wang, S., Zhai, G., Ma, S., Yang, X., Lin, W., Zhang, W. and Gao, W. (2016). Blind quality assessment of tone-mapped images via analysis of information, naturalness, and structure. *IEEE Transactions on Multimedia*, **18**(3), 432–443.
- Guo, X., Li, Y. and Ling, H. (2017). LIME: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, **26**(2), 982–993.
- Guo, Y., Lu, Y., Liu, R. W., Yang, M. and Chui, K. T. (2020). Low-light image enhancement with regularized illumination optimization and deep noise suppression. *IEEE Access*, **8**, 145297–145315.
- He, K., Sun, J. and Tang, X. (2010). Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **33**(12), 2341–2353.
- He, K., Sun, J. and Tang, X. (2013). Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35**(6), 1397–1409.
- He, K., Zhang, X., Ren, S. and Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA.
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv: 1704.04861.
- Hu, W. C., Yang, C. Y. and Huang, D. Y. (2011). Robust real-time ship detection and tracking for visual surveillance of cage aquaculture. *Journal of Visual Communication and Image Representation*, **22**(6), 543–556.
- Jiang, B., Woodell, G. A. and Jobson, D. J. (2015). Novel multi-scale retinex with color restoration on graphics processing unit. *Journal of Real-Time Image Processing*, **10**, 239–253.
- Jobson, D. J., Rahman, Z. and Woodell, G. A. (1997a). Properties and performance of a center/surround retinex. *IEEE Transactions on Image Processing*, **6**(3), 451–462.
- Jobson, D. J., Rahman, Z. and Woodell, G. A. (1997b). 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.
- Kim, Y. T. (1997). Contrast enhancement using brightness preserving bi-histogram equalization. *IEEE Transactions on Consumer Electronics*, **43**(1), 1–8.
- Kim, K., Hong, S., Choi, B. and Kim, E. (2018). Probabilistic ship detection and classification using deep learning. *Applied Sciences*, **8**(6), 936–953.
- Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. arXiv: 1412.6980.
- Land, E. H. (1977). The Retinex theory of color vision. *Scientific American*, **237**(6), 108–129.
- Li, B., Peng, X., Wang, Z., Xu, J. and Feng, D. (2017). AOD-Net: All-in-One Dehazing Network. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy.
- Li, C., Guo, J., Poriki, F. and Pang, Y. (2018). Lightnet: A convolutional neural network for weakly illuminated image enhancement. *Pattern Recognition Letters*, **104**(1), 15–22.
- Liu, R. W., Nie, J., Garg, S., Xiong, Z., Zhang, Y. and Hossain, M. S. (2021a). Data-driven trajectory quality improvement for promoting intelligent vessel traffic services in 6G-enabled maritime IoT systems. *IEEE Internet of Things Journal*, **8**(7), 5374–5385.
- Liu, R. W., Yuan, W., Chen, X. and Lu, Y. (2021b). 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., Yang, M. and Liu, R. W. (2021). DSPNet: Deep Learning-Enabled Blind Reduction of Speckle Noise. *Proceedings of the International Conference on Pattern Recognition*, Milan, Italy.
- Lv, F., Lu, F., Wu, J. and Lim, C. (2018). MBLLEN: Low-Light Image/Video Enhancement Using CNNs. *Proceedings of the British Machine Vision Conference*, Newcastle, UK.
- Mittal, A., Soundararajan, R. and Bovik, A. C. (2013). 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 International Conference on Intelligent Transportation Systems*, Auckland, NZ.

- Pisano, E. D., Zong, S., Hemminger, B. M., DeLuca, M., Johnston, R. E., Muller, K., Braeuning, M. P. and Pizer, S. M. (1998). Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. *Journal of Digital Imaging*, **11**(4), 193–200.
- Radford A., Metz L. and Chintala S. (2015). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *Proceedings of the International Conference on Learning Representations*, San Juan, PR.
- 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.
- Shao, Z., Wang, L., Wang, Z., Du, W. and Wu, W. (2020). Saliency-aware convolution neural network for ship detection in surveillance video. *IEEE Transactions on Circuits and Systems for Video Technology*, **30**(3), 781–794.
- Shen, L., Yue, Z., Fen, F., Chen, Q., Liu, S. and Ma, J. (2017). MSR-net: Low-light image enhancement using deep convolutional network. arXiv: 1711.02488.
- Tan, L. T., Sim, K. S. and Tso, C. P. (2012). Image enhancement using background brightness preserving histogram equalization. *Electronics Letters*, **48**(3), 155–157.
- Wang, Z. and Bovik, A. C. (2009). Mean squared error: Love it or leave it? *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. and Li, B. (2013). Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, **22**(9), 3538–3578.
- Wei, C., Wang, W., Yang, W. and Liu, J. (2018). Deep Retinex Decomposition for Low-Light Enhancement. *Proceedings of the British Machine Vision Conference*, Newcastle, UK.
- Wu, F., Zhou, Z., Wang, B. and Ma, J. (2018). Inshore ship detection based on convolutional neural network in optical satellite images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **11**(11), 4005–4015.
- Yang, C. S., Park, J. H. and Rashid, A. H. A. (2018). An improved method of land masking for synthetic aperture radar-based ship detection. *The Journal of Navigation*, **71**(4), 788–804.
- Yang, M., Nie, X. and Liu, R. W. (2019a). Coarse-to-Fine Luminance Estimation for low-Light Image Enhancement in Maritime Video Surveillance. *Proceedings of the IEEE International Conference on Intelligent Transportation Systems*, Auckland, NZ.
- Yang, W., Liu, J., Yang, S. and Guo, Z. (2019b). Scale-free single image deraining via visibility enhanced recurrent wavelet learning. *IEEE Transactions on Image Processing*, **28**(6), 2948–2961.
- Zhang, L., Zhang, L., Mou, X. and Zhang, D. (2011). FSIM: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, **20**(8), 2378–2386.
- Zhang, L., Shen, Y. and Li, H. (2014). VSI: A visual saliency-induced index for perceptual image quality assessment. *IEEE Transactions on Image Processing*, **23**(10), 4270–4281.
- Zhang, W., Goerlandt, F., Kujala, P. and Wang, Y. (2016). An advanced method for detecting possible near miss ship collisions from AIS data. *Ocean Engineering*, **124**, 1–16.
- Zhang, W. B., Kopca, C., Tang, J., Ma, D. and Wang, Y. (2017). A systematic approach for collision risk analysis based on AIS data. *The Journal of Navigation*, **70**(5), 1117–1132.
- Zhang, Y., Zhang, J. and Guo, X. (2019). Kindling the Darkness: A Practical Low-Light Image Enhancer. *Proceedings of the ACM International Conference on Multimedia*, Nice, France.
- Zhu, C., Zhou, H., Wang, R. and Guo, J. (2010). A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features. *IEEE Transactions on Geoscience and Remote Sensing*, **48**(9), 3446–3456.