IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 30, 2021, accepted January 23, 2022, date of publication February 7, 2022, date of current version February 17, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3147493

Maritime Infrared Image Super-Resolution Using Cascaded Residual Network and Novel Evaluation Metric

ZONGJIANG GAO [1] AND JINHAI CHEN [2]

[1]Navigation College, Dalian Maritime University, Dalian 116026, China
[2]National-Local Joint Engineering Research Center for Marine Navigation Aids Services, Navigation College, Jimei University, Xiamen 361021, China

Corresponding author: Zongjiang Gao (gaojiang@dlmu.edu.cn)

This work was supported in part by the Fundamental Research Funds of the Central Universities under Grant 3132020127; in part by the Navigation College of Jimei University, National-Local Joint Engineering Research Center for Marine Navigation Aids Services under Grant 2020; and in part by the Natural Science Foundation of Liaoning under Grant 2019-ZD-0162.

ABSTRACT Infrared (IR) cameras have been important surveillance sensors for autonomous surface vessels; however, their detection ranges are limited by low resolution. In this study, we collect maritime IR images, analyze the characteristics of those images, and develop datasets for training and testing. Then, a new maritime IR image super-resolution network, maritime infrared super-resolution using cascaded residual network, is developed to reconstruct IR images using a scale of 4. Moreover, different loss functions have different effects on output images; a loss function is set to be a combination of three loss functions, including mean absolute error, mean squared error, and perceptual loss. Peak signal-to-noise ratio and structural similarity index measure cannot effectively describe super-resolution performance. As the novel evaluation metric, Canny edge detection method is used because edges are important for human and target detection algorithms. Finally, experiments are conducted and the results demonstrate that the developed residual network can achieve high-quality reconstructed maritime IR images.

INDEX TERMS Maritime infrared image, residual network, super-resolution.

I. INTRODUCTION

Infrared (IR) cameras are extensively used at sea in applications such as maritime search and rescue, waterway management, and sea farm management. Moreover, IR cameras have been important surveillance sensors for autonomous surface vessels on situational awareness and environmental perception [1]. IR cameras can detect temperature differences at night and compensate for radar and automatic identification system (AIS) limitations in detecting targets at sea, e.g., pirate/illegal fishing/smuggling boats, survival crafts, and people falling into the water.

However, the resolution of shipborne thermal cameras is considerably less than that of shipborne visible light cameras because of the size of photosensitive detectors, manufacturing process, and cost [2], [3]. Image resolution determines the fine degree of image details. The higher the resolution of the same image, the smaller the size represented by a pixel and the clearer the details in the image. Typically, the pixels of shipborne IR cameras are $<$200,000, with the highest value being $\sim$300,000 pixels. Small targets on seas only comprise a dozen or dozens of pixels in images because of the low resolution of IR sensors, making them difficult to be detected by ship officers and detection algorithms. The detection range of marine IR cameras for small ships is $\sim$2–5 nautical miles, whereas that for men overboard is only $\sim$1–2 nautical miles. The aforementioned detection distances are insufficient for collision avoidance or human rescue. Therefore, research on maritime IR image super-resolution is important; it has considerable significance for situational awareness and environmental perception.

Hence, in this study, we propose a model known as maritime IR super-resolution based on a cascaded residual network (MISR-CRN). In this study, four operations were performed: (i) the characteristics of maritime IR images were analyzed; (ii) The loss function and performance evaluation metric were improved as per the characteristics of maritime IR images; (iii) the network was designed to meet the super-resolution (SR) requirement; and (iv) a confirmation experiment was conducted.

The associate editor coordinating the review of this manuscript and approving it for publication was Felix Albu.

The remainder of this study is organized as follows: certain related work on maritime IR intelligent surveillance and SR are briefly introduced in Section II; the proposed method is presented in Section III; results and analysis are presented in Section IV; finally, the work is summarized in Section V.

II. RELATED WORK

A. INFRARED RESEARCH ON UNMANNED SHIPS

For unmanned ships, using IR cameras to compensate for the limitations of radar and AIS in small target detection is a current research hotspot. The European Union's unmanned ship research project, known as ''MUNIN,'' fused data from IR cameras, visible cameras, radars, and AIS [4], [5]. To improve and optimize the perception of the navigation environment, Rolls-Royce's advanced unmanned ship application development plan used IR and visible cameras, radar, AIS, LIDAR, and other technologies [6]. Yara Birkland, a zero emission unmanned ship developed by KONGSBERG, was equipped with IR and visible cameras. The Smart Ship Specifications released by the China Classification Society proposed using advanced sensing technology and sensor information fusion technology to obtain and perceive status information required for navigation [7].

B. IMAGE SUPER-RESOLUTION

Image SR reconstruction is known as image magnification; it uses one or more frames of low-resolution images to develop a high-resolution image, which is extensively used in satellite and aerial image processing, medical image enhancement [8], text image, and fingerprint image processing. SR reconstruction technology increases the number of image pixels and detailed information that low-resolution images do not have. SR reconstruction is a pathological task; the task can be positively definite by adding constraints to determine an optimal solution [9].

SR can then be divided into three categories: interpolation, reconstruction-based, and learning-based SRs. The interpolation method is simple and includes nearest neighbor, bilinear, cubic spline, and local adaptive zoom interpolations. The interpolation method cannot reproduce image details effectively and it generates blurry images. The reconstruction-based method uses prior knowledge for image SR reconstruction based on image degradation models such as the convex set projection, maximum posterior probability, and iterative back projection methods.

In recent years, convolutional neural networks (CNNs) have been used in image SR research [10]–[12]. Images have been reconstructed well by training considerable amount of data multiple times. In 2016, Dong *et al.* proposed the first image SR (SRCNN) algorithm based on a CNN. SRCNN directly trained high-resolution and low-resolution image pairs, achieved end-to-end SR reconstruction of a single image, and eliminated feature extraction and high-resolution image block aggregation [13]. Kim *et al.* proposed a 20-layer CNN to perform the SR reconstruction of a single image and improved the calculation speed of the network via learning residuals and a large learning rate [14]. At the IEEE International Conference on Computer Vision and Pattern Recognition in the same year, Kim *et al.* proposed a deep CNN for image SR reconstruction using loop supervision and jump links. Shi *et al.* proposed a SR (ESPCN) method for obtaining high-resolution images by rearranging the feature maps obtained using subpixel convolutional layers [15]. Huang *et al.* used a bidirectional CNN to redevelop the resolution of multi-frame images. Shared weights were used to replace the complete connection in the recurrent neural network, and they connected the previous input layer to the current hidden layer by conditional convolution to enhance time dependence [16].

Most SR research studies focus on visible images with few studies focusing on IR images. First, visible images are easier to obtain compared to IR images. There are multiple visible images on the Internet. However, IR images are rare, particularly IR images on the sea for detecting and recognizing ships [17]. A model trained on a small dataset can easily overfit. Second, because IR images have low contrast, low signal-to-noise ratio, and blurred edges, the SR reconstruction of IR images is extremely difficult compared to visible images.

Because of the lack of IR images, Choi *et al.* trained a CNN on 91 visible images to enhance IR images and tested the network on IR images [18]. Li trained a CNN on BSD100, a visible image dataset, to develop high-resolution IR images [9]. Two CNN-based models were trained on visible and IR images; the results demonstrated that differences between these two models was not extremely large [19]. He *et al.* designed a cascaded deep network with multiple receptive fields, which was abbreviated CDN_MRF. CDN_MRF was trained on 120 IR images and had two different receptive field deep neural networks (DNNs) to redevelop structural and fine edges [20], [21]. As a part of the Perception Beyond the Visible Spectrum 2020 workshop, the first challenge on thermal image SR was organized in 2020 and six teams' works were introduced [22], [23].

C. LOSS FUNCTIONS

Loss functions have been used to measure the difference between reconstructed images and original high-resolution images. Currently, various loss functions have been extensively used in the SR field, including pixel loss, content loss, texture loss, and adversarial loss. Different loss functions have different impact on the reconstructed images [24], e.g., in early times, the pixel wise L2 loss was extensively used; however, L2 loss can make the overall images to be more even. In practice, researchers often combine multiple loss functions using a weighted average. In this study, we use a combination of three different loss functions: mean absolute error (MAE), mean squared error (MSE), and perceptual loss (P_loss). The combined loss function can avoid the limitations caused by a single loss function. The loss function used in this study can calculate the low level error between the

ground truth images and generated images as well as the high-level perceptual and semantic differences.

D. EVALUATION

Evaluation, including subjective and objective methods, on the reconstructed images is known as image quality assessment. Mean opinion score is one of the commonly used subjective methods that people rate in the reconstructed images [25]. In objective methods, peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) is extensively used in the SR field. Whether it is a subjective or objective method, an image with a high score reports that the reconstruction of the image is good. Usually, the reconstructed images are used as input for other tasks; therefore, in recent years task-based evaluation is gaining popularity [26]. In this study, the edges of maritime IR images are important to ship officer vision and target detection algorithm; therefore, we propose a novel task-based evaluation metric that measures the length of edges of reconstructed images.

III. PROPOSED METHOD

A. MARITIME INFRARED IMAGE

We installed an IR camera (FLIR 617CS) on a ship and collected IR images on the sea and near ports. Figure 1 shows samples of maritime infrared images. The timestamps and some icons can be removed by changing the camera settings but others cannot be removed. These icons have clear edges, whereas the sea targets have blurred edges. CNNs are apt to identify clear edges; therefore, networks trained on those images are suited better to recover icons rather than the IR target.

(a) (b)

FIGURE 1. Maritime infrared images.

Furthermore, the sky and sea occupy most IR images with the target accounting to only a small portion of the image. The grayscale of the sky and sea is almost identical; there are no discernible edges. If these images are used to train a CNN, the network will not be able to learn anything useful.

B. DATASET

The lack of sufficient data to train CNNs has always been a major problem for the infrared image SR. Because of the slow change of scene at sea, images are similar for a long time; therefore, the quantity of images collected is not sufficient to train a CNN and are used only for testing.

In this study, we used T91, BSD100 [27], and BSD200 [28] as training and validation datasets, which are classical image datasets used in SR. These datasets are composed of multiple images ranging from nature images to object specific images such as plants, people, and food. First, we converted the images to HSV and extracted the V channel. Then, we cropped the images to patches with the size and step of 44 to increase the number of training images. Finally, we obtained 15707 images in the dataset. Note that 90% of the patches were used for training, while the remaining 10% were used for validation. IR images captured on the ship were used for testing. We used bicubic interpolation to down-sample the original images by a scale of 4 to obtain low-resolution images. The network inputs low-resolution images and outputs constructed images as follows:

$$I_{SR} = SR(D \cdot I_{ori}), \tag{1}$$

where I_{ori} is the original image, D is the down-sampling operation, and SR is the SR construction.

C. NETWORK STRUCTURE

Typically, a deep CNN is effective at extracting features but suffers from the gradient vanishing problem during back-propagation. In this study, a cascaded residual network was developed to reconstruct low-resolution images; therefore, we named the network MISR-CRN. In Figure 2, the network's operations are divided into three parts: feature extraction, reconstruction, and fine-tuning. To reduce computation complexity, we used two transpose convolutions to process an input image at the original size and increased the resolution step by step. First, a transpose convolution increased an input image by two; four convolutional blocks were used to extract features; and a residual connect was adopted to learn high-frequency parts between two images. The skip connection can propagate an error to the front layers in a shortcut. Then, to reconstruct the image, a transpose convolution and four convolution blocks were used in the second part. Finally, a convolution block was adopted for fine-tuning the output image.

D. LOSS FUNCTION

A weighted combination of three different loss functions is used: MAE, MSE, and P_loss. MAE denotes the mean absolute error between a ground truth image and a generated image and is defined as follows:

$$MAE = \frac{1}{N}\sum_{p=1}^{N} \left| I_{ori}^{p} - I_{SR}^{p} \right|, \tag{2}$$

where N is the total number of pixels in the images and I^p is the value of pixel.

The MSE is used for maintaining consistency between the input and output images [29] and can be mathematically

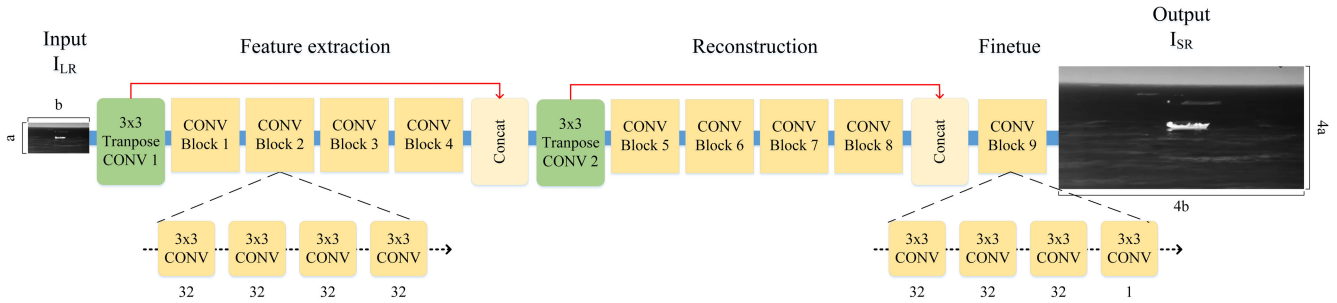

FIGURE 2. Structure of the proposed network (MISR-CRN).

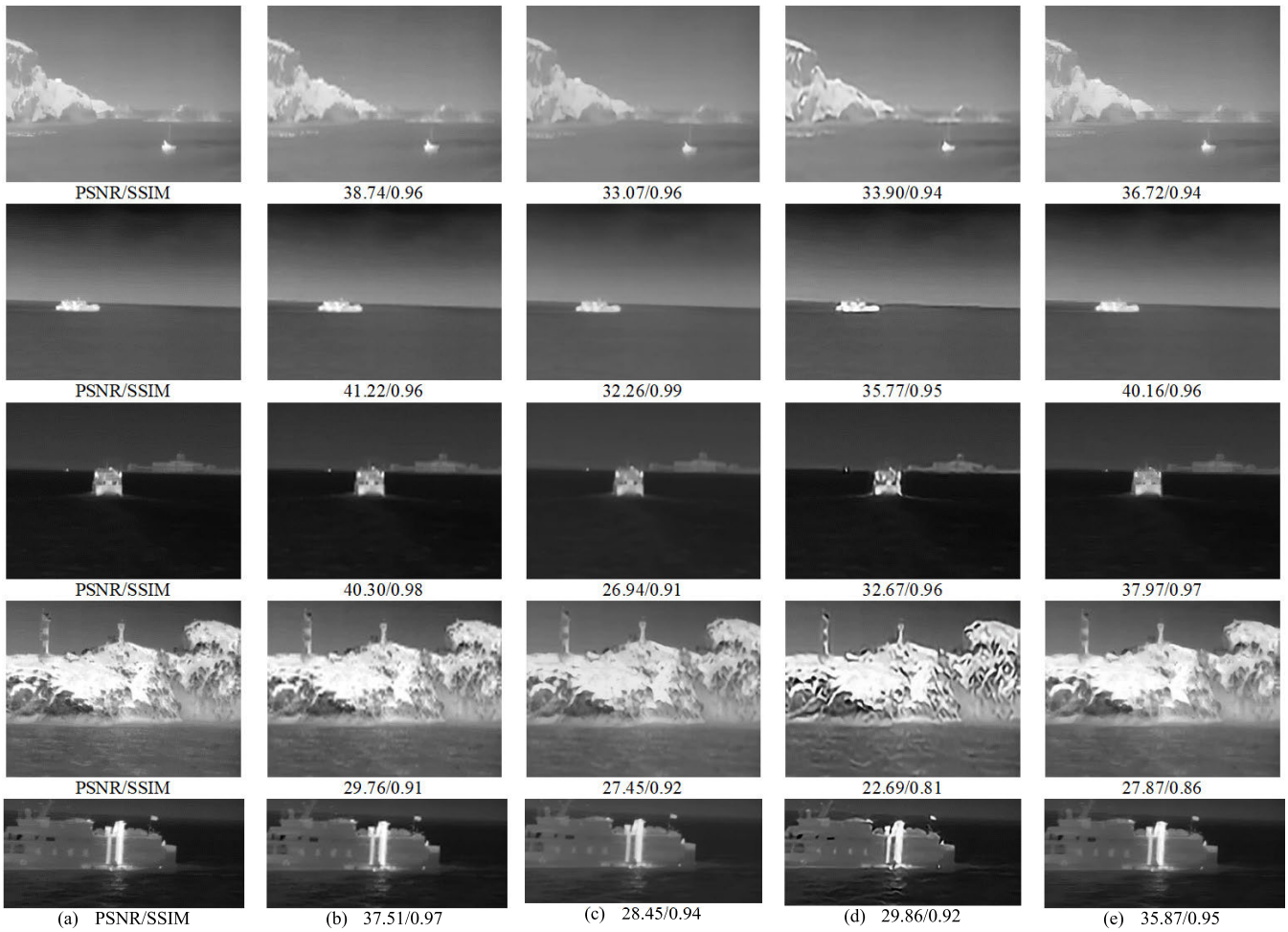

FIGURE 3. Super-resolution results on five typical maritime infrared images with factor 4: (a) original images, (b) bicubic, (c) super-resolution convolutional neural network, (d) CDN_MRF, and (e) our method.

represented as follows::

$$MSE = \frac{1}{N}\sum_{p=1}^{N}\left(I_{ori}^{p} - I_{SR}^{p}\right)^2, \tag{3}$$

P_loss measures the high-level perceptual and semantic differences between I_{ori} and I_{SR} [30]. In our experiments, a 19-layer VGG network retrained on visual and IR datasets is used as the loss network φ [31]. Then, P_loss is defined as follows:

$$P\_loss = \frac{1}{C_j H_j W_j}\left\|\varphi_j(I_{ori}) - \varphi_j(I_{SR})\right\|_2^2, \tag{4}$$

where $\varphi_j(I)$ is the output of image I at the jth layer and C, H, and W are the channel, height, and width of the output of the jth layer, respectively.

Total loss is a weighted sum of MAE, MSE, and P_loss; they have different effects on the output image, e.g., the

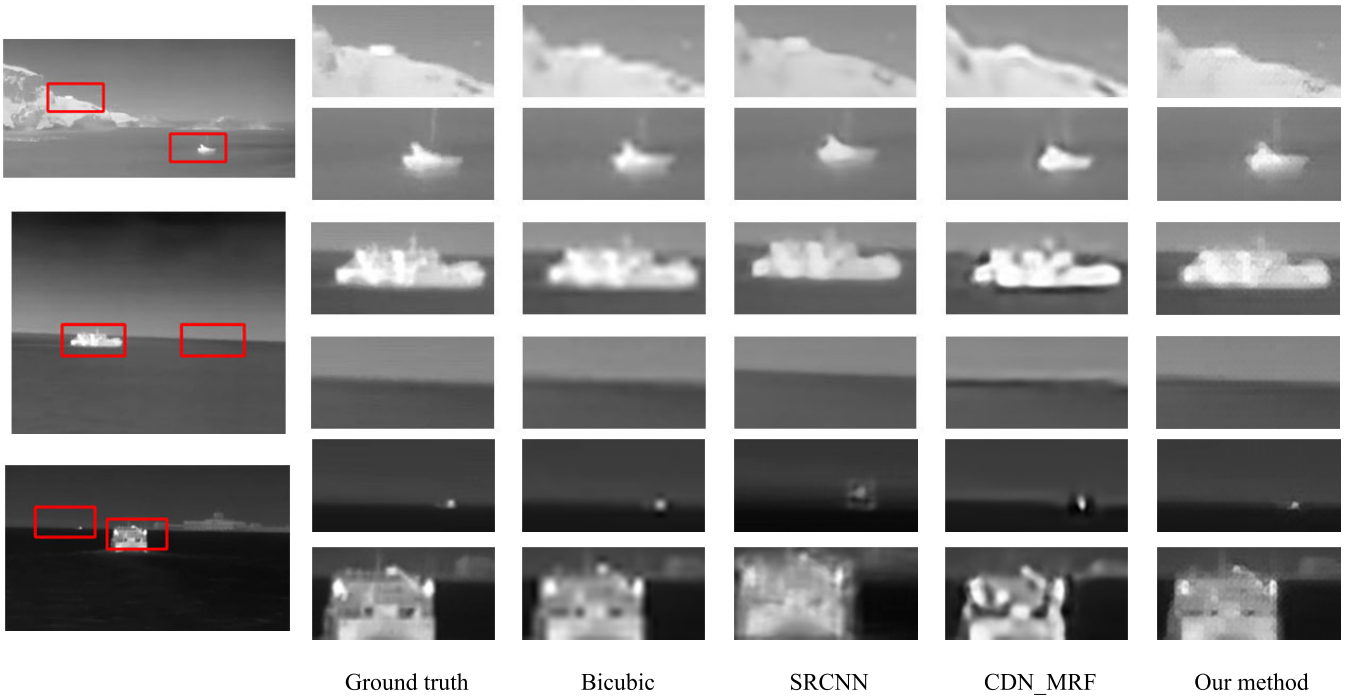

FIGURE 4. Qualitative comparison of algorithms for the ×4 scale SR.

reconstructed images obtained using MSE as the loss function are blurry. Moreover, the images obtained using P_loss will have effects such as the mosaic effect.

The total loss is defined as

$$Total\ loss = \alpha \cdot MAE + \beta \cdot MSE + \gamma \cdot P\_loss, \quad (5)$$

where α, β, and γ are weights associated with every loss, which have been empirically set. The network is supervised using the proposed loss function in Eq. (5). Adam optimizer is selected as the optimizer, the learning rate is set to 0.001, and the batch size is set to 1,000. ReLu is used as the activation function in the network. To prevent overfitting, network training is stopped when the validation loss does not show any improvement over 10 epochs. The network is trained on five NVIDIA V100 GPUs with CUDA and cuDNN in parallel.

E. PERFORMANCE EVALUATION

There is no consensus on which metrics can best describe SR performance. PSNR and SSIM are extensively used matrices for determining the difference between ground truth and outputs. However, they have been reported to have a poor correlation with human perception of visual quality [30]. For example, PSNR is defined as follows:

$$PSNR(I_{ori}, I_{SR}) = 10 \log_{10} \left(\frac{255^2}{MSE(I_{ori}, I_{SR})} \right), \quad (6)$$

where $MSE(I_{ori}, I_{SR})$ is the mean square error between I_{ori} and I_{SR}. PSNR has a reciprocal relationship with MSE. When MSE is used as the loss function, PSNR is possibly high. Therefore, these metrics cannot be used to evaluate maritime IR images because targets, such as ships at sea, only occupy a small portion of images. The majority of IR image patches are background patches with small grayscale variations; the bicubic method produces good results in an entire image; however, the edge of sea targets and sea horizon is blurred. Figure 3 shows that the bicubic method performs best in terms of PSNR and SSIM; however, the entire images are blurry. Therefore, high PSNR and SSIM do not indicate that this method is effective at recovering the details of maritime IR images.

The edges are important in human vision, target detection, tracking, and classification. If an image is reconstructed well, it should have clear edges; therefore, we extract the edges from images and use the length of edges as an evaluation metric. Canny edge detection is a popular edge detection algorithm. It can reduce the noise in IR images and suppress non-maximum; therefore, it is used to detect edges. In hysteresis thresholding, we set the minimum and maximum thresholds to 50 and 150, respectively. Edges with an intensity gradient less than the minimum value are not edges and should be discarded. Those that fall between these two thresholds are classified as edges if they are connected to "sure edge" pixels; otherwise, they are discarded.

IV. RESULTS AND DISCUSSION

To confirm the efficiency of our proposed method, we used three state-of-the-art methods for comparison, including a classic method (Bicubic interpolation), deep-learning-based methods (SRCNN [13] and CDN_MRF [20], [21]). The source codes of SRCNN and CDN_MRF are provided by

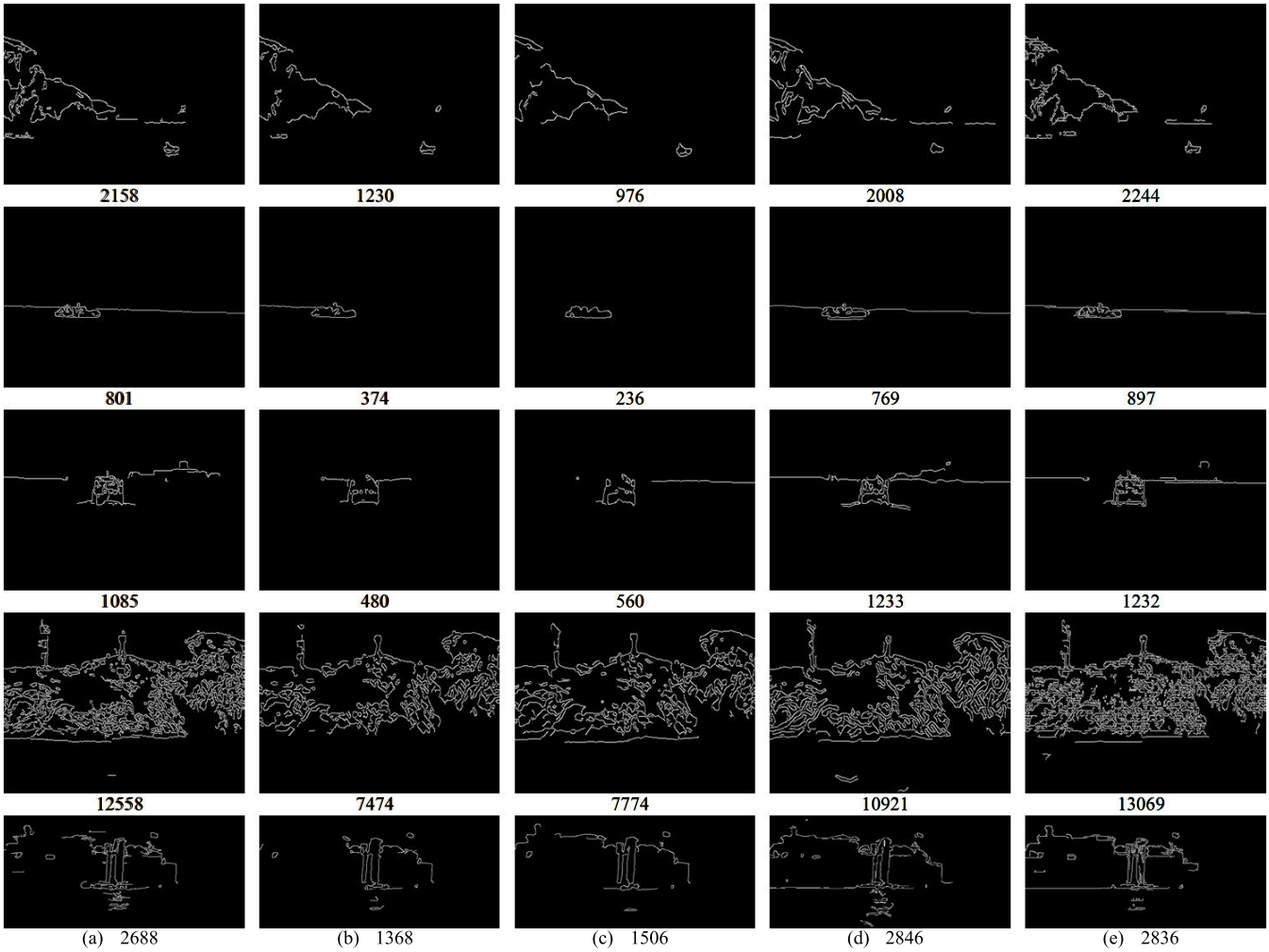

FIGURE 5. Canny detections on maritime infrared images: (a) Canny on original images, (b) Canny on Bicubic, (c) Canny on SRCNN, (d) Canny on CDN_MRF, and (e) Canny on our method. The numbers are the lengths of the edges.

their authors and the weights are unchanged. We use those codes to reconstruct the test images and evaluate them in our code.

We performed quantitative experiments to evaluate all models using five representative images, which were captured on the ship. Figure 3 shows the results for $\times 4$ scale SR. Moreover, qualitative comparisons are included, and the results are shown in Figure 4. The images reconstructed through bicubic interpolation have the highest PSNR and SSIM; however, the edges are more blurred compared to others. As described in Section III, PSNR and SSIM are evaluations performed on the overall images. The results confirm that PSNR and SSIM are unsuitable for evaluating maritime IR image reconstruction. SRCNN is ineffective on these images, indicating that the reconstructed IR images have blurred edges. The image generated by CDN_MRF has good contrast; however, the reconstruction of the edges is poor, resulting in multiple curved shapes. Although the images reconstructed by our method have lower PSNR and SSIM than the bicubic interpolation method, they have sharper edges that are important to humans and additional study. The small targets in the reconstructed images exhibit considerable brightness, which is important for the target detection algorithm.

We use Canny to detect the edges of reconstructed images and edges' lengths to evaluate SR. Figure 5 shows the quantitative comparisons. The results show that the images reconstructed using our method have the longest edges in the first two and fourth images. In the second and fifth images, although the length of the reconstructed images using our method is slightly lower than that of CDN_MRF, the horizon and outer boundaries of the ship are more complete in our reconstructed images than that of CDN_MRF. Because we use clear visual images to train our network, the quality of the reconstructed images is better than the original images.

In Canny, there are two thresholds, a minimum threshold and a maximum threshold. The different thresholds for edge detectors can produce different results. To confirm that our method can work for different thresholds, we consider

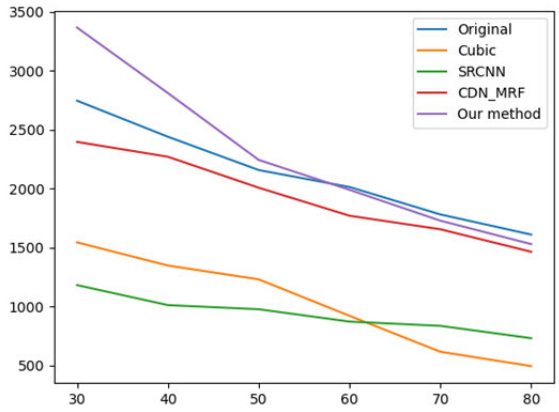

FIGURE 6. Lengths of the Canny edges on different thresholds. The vertical axis is the length of the edges and the horizontal axis is the minimum threshold.

the first image in Figure 3 as an example to confirm the edge length at different minimum thresholds. The maximum threshold is set to be 100 greater than the corresponding minimum threshold. The comparisons are depicted in Figure 6; the results demonstrate that when the minimum threshold is between 30 and 50, the edge length of the reconstructed image achieved by our method is greater than that of the original image. Our method reconstructs the images and improves the quality of original images. When the minimum threshold is between 60 and 80, the edge length of image reconstructed by this method is slightly smaller than the original image and significantly larger than other methods.

V. CONCLUSION

This study demonstrated a novel CNN-based approach for the maritime image SR. The characteristics of maritime IR images were analyzed. The network comprised three parts: feature extraction, reconstruction, and fine-tuning. It was trained on three extensively used visual image datasets. A combination of MAE, MSE, and P_loss was used to develop images with clear and natural appearances. For these experiments, Canny was a novel metric for evaluating image reconstruction compared to PSNR and SSIM. The experimental results demonstrated that the reconstructed images had improved quality. Moreover, future work will focus on identifying other evaluation metrics for evaluating maritime IR images.

REFERENCES

[1] C. A. Thieme, I. B. Utne, and S. Haugen, "Assessing ship risk model applicability to marine autonomous surface ships," *Ocean Eng.*, vol. 165, pp. 140–154, Oct. 2018, doi: 10.1016/j.oceaneng.2018.07.040.

[2] Y. Li, K. Zhao, F. Ren, B. Wang, and J. Zhao, "Research on super-resolution image reconstruction based on low-resolution infrared sensor," *IEEE Access*, vol. 8, pp. 69186–69199, 2020, doi: 10.1109/ACCESS.2020.2984945.

[3] Y. Yang, Q. Li, C. Yang, Y. Fu, H. Feng, Z. Xu, and Y. Chen, "Deep networks with detail enhancement for infrared image super-resolution," *IEEE Access*, vol. 8, pp. 158690–158701, 2020, doi: 10.1109/ACCESS.2020.3017819.

[4] Ø. J. Rødseth and Å. Tjora, "A system architecture for an unmanned ship," in *Proc. 13th Int. Conf. Comput. IT Appl. Maritime Industries (COMPIT)*, 2014, pp. 1–13.

[5] C. Fan, K. Wróbel, J. Montewka, M. Gil, C. Wan, and D. Zhang, "A framework to identify factors influencing navigational risk for maritime autonomous surface ships," *Ocean Eng.*, vol. 202, pp. 107–188, Apr. 2020, doi: 10.1016/j.oceaneng.2020.107188.

[6] E. Jokioinen *et al.*, "Remote and autonomous ships the next steps," Rolls-Royce, London, U.K., Tech. Rep. AAWA, 2016.

[7] *Smart Ship Specification 2020*, China Classification Soc., Beijing, China, Dec. 2019. [Online]. Available: https://www.ccs.org.cn/ccswz/articleDetail?id=201900001000009739

[8] G. Suryanarayana, K. Chandran, O. I. Khalaf, Y. Alotaibi, A. Alsufyani, and S. A. Alghamdi, "Accurate magnetic resonance image super-resolution using deep networks and Gaussian filtering in the stationary wavelet domain," *IEEE Access*, vol. 9, pp. 71406–71417, 2021, doi: 10.1109/ACCESS.2021.3077611.

[9] F. Li, "Research on super-resolution reconstruction of infrared imaging system," Ph.D. dissertation, Changchun Inst. Opt., Fine Mech. Phys., Chinese Acad. Sci., Changchun, China, 2018.

[10] G. Suryanarayana, E. Tu, and J. Yang, "Infrared super-resolution imaging using multi-scale saliency and deep wavelet residuals," *Infr. Phys. Technol.*, vol. 97, pp. 177–186, Mar. 2019, doi: 10.1016/j.infrared.2018.12.028.

[11] T. Yao, Y. Luo, J. Hu, H. Xie, and Q. Hu, "Infrared image super-resolution via discriminative dictionary and deep residual network," *Infr. Phys. Technol.*, vol. 107, Jun. 2020, Art. no. 103314, doi: 10.1016/j.infrared.2020.103314.

[12] Y. Zhao, X. Sui, Q. Chen, and S. Wu, "Learning-based compressed sensing for infrared image super resolution," *Infr. Phys. Technol.*, vol. 76, pp. 139–147, May 2016.

[13] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016, doi: 10.1109/TPAMI.2015.2439281.

[14] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654, doi: 10.1109/CVPR.2016.182.

[15] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1874–1883.

[16] Y. Huang, W. Wang, and L. Wang, "Bidirectional recurrent convolutional networks for multi-frame super-resolution," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 235–243.

[17] M. M. Zhang, J. Choi, K. Daniilidis, M. T. Wolf, and C. Kanan, "VAIS: A dataset for recognizing maritime imagery in the visible and infrared spectrums," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 10–16, doi: 10.1109/CVPRW.2015.7301291.

[18] Y. Choi, N. Kim, S. Hwang, and I. S. Kweon, "Thermal image enhancement using convolutional neural network," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2016, pp. 223–230, doi: 10.1109/IROS.2016.7759059.

[19] R. E. Rivadeneira, P. L. Suarez, A. D. Sappa, and B. X. Vintimilla, "Thermal image SuperResolution through deep convolutional neural network," in *Proc. 16th Int. Conf. Image Anal. Recognit. (ICIAR)*, Waterloo, ON, Canada, vol. 11663, 2019, pp. 417–426, doi: 10.1007/978-3-030-27272-2_37.

[20] Z. He, "Research on the strip noise elimination and super-resolution technologies for uncooled long-wave infrared images," Ph.D. dissertation, Zhejiang Univ., Hangzhou, China, 2019.

[21] Z. He, S. Tang, J. Yang, Y. Cao, M. Y. Yang, and Y. Cao, "Cascaded deep networks with multiple receptive fields for infrared image super-resolution," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 8, pp. 2310–2322, Aug. 2019, doi: 10.1109/tcsvt.2018.2864777.

[22] R. E. Rivadeneira *et al.*, "Thermal image super-resolution challenge—PBVS 2020," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 432–439, doi: 10.1109/CVPRW50498.2020.00056.

[23] R. E. Rivadeneira, A. D. Sappa, and B. X. Vintimilla, "Thermal image super-resolution: A novel architecture and dataset," in *Proc. VISIGRAPP*, 2020, pp. 111–119.

[24] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Trans. Comput. Imag.*, vol. 3, no. 1, pp. 47–57, Mar. 2017, doi: 10.1109/TCI.2016.2644865.
[25] S. Anwar, S. Khan, and N. Barnes, "A deep journey into super-resolution: A survey," 2019, *arXiv:1904.07523*.
[26] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for image super-resolution: A survey," 2019, *arXiv:1902.06068*.
[27] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 2, Jul. 2001, pp. 416–423, doi: 10.1109/ICCV.2001.937655.
[28] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, Aug. 2011, doi: 10.1109/TPAMI.2010.161.
[29] P. Kansal and S. Nathan, "A multi-level supervision model: A novel approach for thermal image super resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 426–431, doi: 10.1109/CVPRW50498.2020.00055.
[30] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Cham, Switzerland: Springer, 2016, pp. 694–711.
[31] Z. Gao, Y. Zhang, and Y. Li, "Extracting features from infrared images using convolutional neural networks and transfer learning," *Infr. Phys. Technol.*, vol. 105, Mar. 2020, Art. no. 103237, doi: 10.1016/j.infrared.2020.103237.

ZONGJIANG GAO received the master's degree from the Navigation College, Dalian Maritime University, in 2010. He is currently a Teacher with the Navigation College, Dalian Maritime University. His research interests include computer vision, image processing, and deep learning.

JINHAI CHEN received the Ph.D. degree from the Institute of Geographic Sciences and Natural Resources Research, CAS, in 2015. He is currently an Associate Professor with Jimei University. His research interests include marine traffic geographic information, ship trajectory data analysis, and deep learning.