



Infrared image super-resolution method based on dual-branch deep neural network

Huang Zhijian¹ · Hui Bingwei² · Sun Shujin³ · Ma Yanxin⁴

Accepted: 14 April 2023 / Published online: 28 May 2023

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

Infrared image has lower resolution, lower contrast, and less detail than visible image, which causes its super-resolution (SR) more difficult than visible image. This paper presents an approach based on a deep neural network that comprises an image SR branch and a gradient SR branch to reconstruct high-quality SR image from single-frame infrared image. The image SR branch reconstructs the SR image from the initial low-resolution infrared image using a basic structure similar to the enhanced SR generative adversarial network (ESRGAN). The gradient SR branch removes haze, extracts the gradient map, and reconstructs the SR gradient map. To obtain more natural SR image, a fusion block based on attention mechanism is adopted between these branches. To preserve the geometric structure, gradient L1 loss and gradient GAN loss are defined and added. Experimental results on a public infrared image dataset demonstrate that, compared with the current SR methods, the proposed method is more natural and realistic, and can better preserve the structures.

Keywords Infrared image · Super-resolution · ESRGAN · Gradient guidance · Haze removal

1 Introduction

Infrared imaging systems, which are widely utilized in the military, medical, and public security areas, can record environmental information under challenging weather such as darkness, rain, and fog. Compared with the million pixel resolution of visible light sensors, the resolution of infrared imaging systems is usually far lower than the resolution required for practical applications. However, increasing the resolution of infrared imaging systems by hardware, such as reducing pixel size or expanding the detector matrix, increases the production cost significantly. More importantly,

in some scenarios, such as the military, where volume and weight are typically limiting factors, increasing the resolution through hardware is not practical. Therefore, using software to improve the resolution of infrared images is the most promising technical approach.

Currently, visible image super-resolution (SR) method progressed markedly due to the rapid development of deep learning technology [1]. The single-image SR methods based on deep learning may be grouped into four categories according to the input image characteristics, the network structure, the method of feature extraction, and the way of processing information. (1) The first methods are based on interpolating, in which initial image is first scaled to the output image size by interpolation and then refines the details using a deep network. SRCNN [2] used a deep neural network first time for SR reconstruction. It only employed a three-layer network, but the effect is far better than the traditional methods. VDSR [3] adopted a residual net to build a 20-layer model with enlarged receptive field, resulting in multiple SR images. Based on this idea, some excellent models emerged, such as IRCNN [4], Memnet [5], DRCN [6], DRRN [7], and SDSR [8]. (2) The second methods employ a low-resolution (LR) image directly, avoiding the loss of detail caused by interpolation and drastically reducing the calculation time. The FSRCNN [9] constructed a fast SR network with a

✉ Huang Zhijian
274652724@qq.com

Hui Bingwei
huibingwei07@163.com

¹ School of Computer Science and Engineering, Changsha University, Changsha 410003, China

² ATR Key Laboratory, School of Electronic Science, National University of Defense Technology, Changsha, China

³ Hunan Province Key Laboratory of Industrial Internet Technology and Security, Changsha, China

⁴ College of Meteorology and Oceanology, National University of Defense Technology, Changsha, China

deconvolution layer, small convolution kernels and shared deep layers. RED [10] employed a symmetrical convolutional–deconvolution layer. ESPCN [11] extracted features in LR space and enlarged the image to the target size by a sub-pixel convolutional layer. SRResNet [12] and EDSR [13] extracted features in the LR space with residual learning and enlarged LR features by a sub-pixel convolutional layer. Luying Li [14] proposed an unsupervised face super-resolution via gradient enhancement and semantic guidance. (3) The third methods adopt dense net technique which overcomes the sparseness of effective features layer-by-layer. SRDenseNet [15] used a dense net to obtain SR images with a better visual effect. The paper [16] proposed a joint restoration convolutional neural network for low-quality image super-resolution. The paper [17] proposed a single-image SR method based on local biquadratic spline with edge constraints and adaptive optimization in transform domain. Zhang et al. [18] extracted local features by a residual dense network. DenseNet [19] used high-frequency information to augment the dense network (SRDN), which pays more attention to high-frequency regions such as edges and textures. (4) The fourth methods employ the GAN net. SRGAN [20] raised the quality of SR to a new level based on a GAN model. ESRGAN [21] increased the speed of model training and improved the quality of SR.

However, there are several problems when using the above methods directly for infrared image SR. That because these visible image SR methods do not take into account the unique characteristics of infrared image, such as: Infrared image often has low resolution, weak contrast, and few details [22, 23]; the fine geometric structure in the infrared image is easy to be destroyed in the process of super-division resulting in distortion [24, 25]; and water vapor absorption and atmospheric scattering causes blur, showing the characteristics of haze in infrared image.

Considering the above problems, this paper presents a new method for infrared image SR. The main contributions of this paper are as follows:

- (1) To obtain high-quality SR image from single-frame infrared image, we propose a dual-branch deep neural network. The image SR branch reconstructs the SR image from the initial infrared image using a basic structure similar to the ESRGAN. The gradient SR branch removes haze, extracts the gradient map, and reconstructs the high-resolution gradient map. To reduce the complexity and calculation, the gradient SR branch directly uses the intermediate-level features extracted in the image SR branch.
- (2) Since infrared image has lower contrast and less detail than visible image, enhancing the detail information in the original image is important for infrared image SR. To emphasize contrast of initial IR image, a haze removal

method based on a dark-channel prior model [26] is adopted before gradient extraction block.

- (3) To fuse the gradient SR map into the image SR map more naturally, this paper adopts a fusion block based on attention mechanism.
- (4) To preserving fine geometric structure, we designed a gradient L1 loss and gradient GAN loss, which supervises the generator training as a second-order constraint.

2 Methods

2.1 Overall process

The method includes two branches, the image SR branch and gradient SR branch, as shown in Fig. 1. The image SR branch reconstructs a SR image from an initial infrared image using a basic structure similar to ESRGAN; the gradient SR branch removes the haze first, then extracts the gradient map, and reconstructs the SR gradient map. To obtain more naturally SR image, a fusion block based on attention mechanism is adopted. To reduce calculations during the gradient SR process, several intermediate-level features from the image SR branch are used directly.

The image SR branch uses a network similar to ESRGAN [21] constructed with the basic residual-in-residual dense block (RRDB) module to reconstruct an SR image from the initial infrared image. The gradient extraction block extracts the gradient map using the Sobel or Laplacian operator. The following sections will introduce the haze removal module, the gradient SR branch, and the fusion block in detail. To better preserve the structure, we add gradient L1 loss and gradient GAN loss.

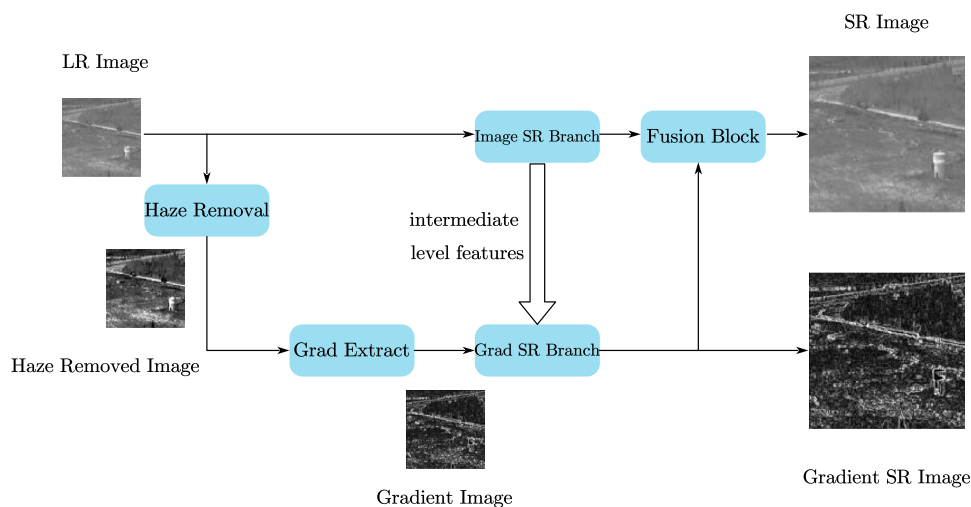
2.2 Haze removal of infrared image

Since infrared images have lower contrast and less detail than visible images [27], enhancing the detail information in the original image is important for infrared image SR. However, infrared images are usually blurred and show haze characteristics in visually because of water vapor absorption and atmospheric scattering [28]. So, a haze removal based on dark-channel prior model is adopted to emphasize contrast of initial IR image before gradient extraction block.

The dark-channel prior model [26] is a haze removal method of visible image which has RGB three channels. Its basic hypothesis is that the haze patch has very low intensity at least one color channels in most non-sky patches. In other words, the minimum intensity in such a patch should be very low. It expressed by a mathematical model as follows:

$$J^d(x) = \min_{c \in (r, g, b)} \left(\min_{y \in \Omega(x)} J^c(y) \right) \rightarrow 0 \quad (1)$$

Fig. 1 Overall framework of our method



where J^c is the color channel of the image J , $\Omega(x)$ is a local patch centered at (x,y) , and $c \in (r,g,b)$ is a pixel in the patch $\Omega(x)$. The dark-channel prior model says that if J is a haze-free outdoor image, then, except for the sky region, the intensity of J^d is low and tends to zero. Since the infrared image only has one channel, Eq. (1) can be simplified as follows:

$$J^d(x) = \min_{y \in \Omega(x)} J(y) \rightarrow 0 \tag{2}$$

According to the above dark-channel prior model (1), we can model the transmission and simplify its estimation as follows:

$$t(x) = (1 - \omega) \min_{y \in \Omega(x)} \frac{I(y)}{A} \tag{3}$$

where $I(y)$ is the original image with haze, which is known, A is the global atmospheric light value, which is unknown, and ω is the rate of haze removal in the interval $[0,1]$, whose default value is 0.95.

In practice, a simple method can be used to estimate the atmospheric light A with the following steps:

- (1) Picking the top 0.1% brightest pixels from the dark channel. These pixels are most haze opaque.
- (2) Among these pixels, the pixels with highest intensity in the input image are selected as the atmospheric light A .

After A and $t(x)$ are estimated, the haze-free image can be recovered from the haze removal model using the following formula:

$$J(x) = \frac{I(x) - A}{\max(t(x), t_0)} \tag{4}$$

where $I(x)$ is the input image, A is the estimated global atmospheric light, and $t(x)$ is the estimated transmission within the window using Eq. (3). t_0 is the lowest transmission, which means that a small amount of haze is preserved in very dense haze regions. A typical value of t_0 is 0.1.

2.3 Gradient SR branch

The gradient SR branch recovers SR gradient map from the LR gradient map with several intermediate-level features from the image SR branch. The recovered SR gradient map will be sent into the fusion block to the final SR image. The network structure of the gradient SR branch is shown in Fig. 2.

Consistent with the image SR branch has 23 RRDBs, the gradient SR branch consists of 22 Grad-Conv block, three independent 3×3 Conv blocks, and one 4 times upsampling block. Each Grad-Conv block integrates the output of the previous Grad-Conv block and the output of the current RRDB block to produce next level gradient feature. The motivation of such a scheme is that the well-designed ESRGAN can carry rich structural information, which is important for the recovery of gradient maps.

The Grad-Conv block locates between two RRDBs to extract high-level features from the gradient map. The Grad block can be a residual block or a bottleneck block. The two structures have no obvious advantages or disadvantages, and both can be utilized as practical examples. The two network structures of the Grad block are shown in Fig. 3.

2.4 Fusion block

To fuse the gradient SR map into the image SR map more naturally, this paper adopts a fusion block based on attention mechanism, which is shown in Fig. 4. First, the gradient SR map and the image SR map are put into attention block,

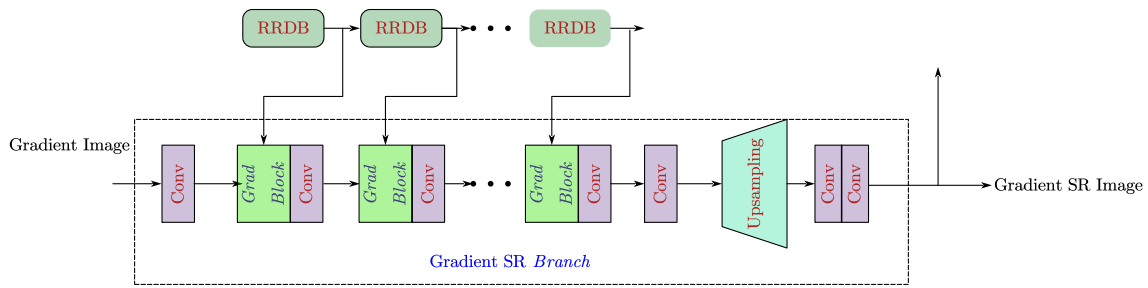


Fig. 2 The structure of gradient SR branch

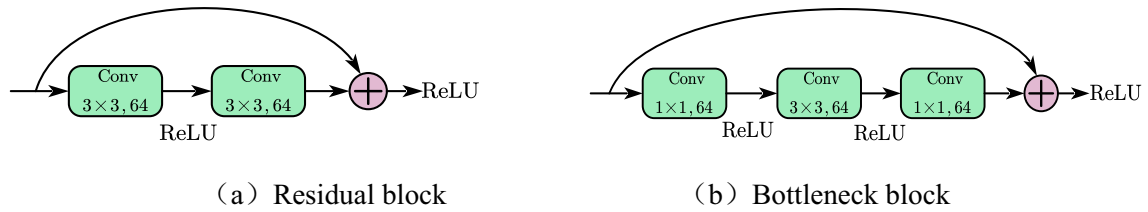


Fig. 3 Two structures of Grad block

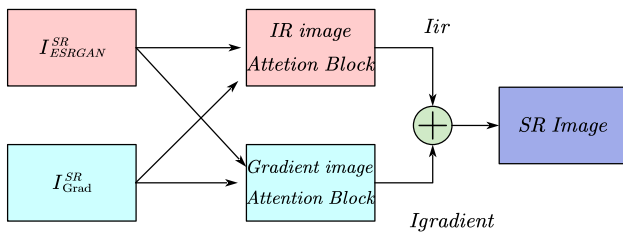


Fig. 4 Fusion block

respectively, to obtain the corresponding weights; then, the map enhancement is performed with their weights, respectively, to obtain the fused maps based on each attention; finally, the maps are fused with averaging rule to obtain the final fused image.

The calculation equation is as follows:

$$I_{ir} = w_{ir} \times I_{ESRGAN}^{SR} + w_{gr} \times I_{Grad}^{SR} \tag{5}$$

$$I_{gradient} = \varphi_{ir} \times I_{ESRGAN}^{SR} + \varphi_{gr} \times I_{Grad}^{SR} \tag{6}$$

$$I_{final}^{SR} = (I_{ir} + I_{gradient})/2 \tag{7}$$

where w_{ir} and w_{gr} denote the weights of I_{ESRGAN}^{SR} and I_{Grad}^{SR} , respectively, in the IR image attention block, and similarly, φ_{ir} and φ_{gr} denote the weights of I_{ESRGAN}^{SR} and I_{Grad}^{SR} , respectively, in the gradient image attention block. I_{ir} and $I_{gradient}$ denote, respectively, the maps enhancement in the IR image attention block and the gradient image attention block. I_{final}^{SR} is the final fused SR image.

The detail of the attention block is shown in Fig. 5. It is noting that the larger size convolution kernels are decomposed

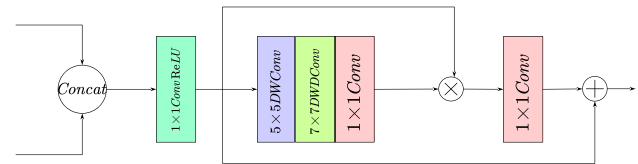


Fig. 5 Attention block

into depth-wise convolution (DWC), depth-wise dilation convolution (DWDC), and 1*1 convolution, which reduces the number of parameters while maintaining a larger receptive field and improves efficiency. To calculate weights, soft-max and average pooling are adopted in the attention block.

2.5 Loss function

To preserve geometric structure, we add gradient loss including gradient L1 loss and gradient GAN loss. So, the loss function includes image branch loss and gradient branch loss. The composition structure is as follows (Fig. 6):

- (1) L1 loss

The L_I^{pix} loss represents the absolute error between the SR image and real high-resolution image.

$$L_I^{pix} = E_I ||G(I^{LR}) - I^{HR}||_1 \tag{8}$$

where I^{LR} is the initial LR image, I^{HR} is the high-resolution (HR) image, $G(\cdot)$ is the generator whose output is the SR

image, $\|\cdot\|_1$ is the L1 norm, and $E_I(\cdot)$ is the sum of pixels in the image I .

(2) Perceptual loss

The perceptual loss L_I^{per} characterizes the error between the i -th layer feature ϕ_i of the SR image and the corresponding layer feature of the real high-resolution image:

$$L_I^{per} = E_I \|\phi_i(G(I^{LR})) - \phi_i(I^{HR})\|_1 \tag{9}$$

where $\phi_i(\cdot)$ denotes the i -th layer output of the image SR model.

(3) Image GAN loss

The image SR branch is an adversarial network (GAN). The image SR branch discriminator D and generator G are optimized using a two-player game as follows:

$$L_I^{Dis} = -E_I[\log(1 - D(I^{SR}))] - E_I[\log(D(I^{HR}))] \tag{10}$$

$$L_I^{Adv} = -E_I[\log D(G(I^{LR}))] \tag{11}$$

where I^{SR} is the SR infrared image.

(4) Gradient L1 loss

Since gradient map can reflect structural information, we use it as a second-order constraint to supervise the generator training. With the supervision in both image and gradient domains, the generator can not only learn fine appearance, but also avoid fine geometric constructure distortion. The gradient L1 loss L_{GM}^{pix} characterizes the absolute error between the generated SR gradient feature map and the HR gradient feature map.

$$L_{GM}^{pix} = E_{GM} \|M(G(I^{LR})) - M(I^{HR})\|_1 \tag{12}$$

where $M(\cdot)$ is the gradient extraction.

(5) Gradient GAN loss

To discriminate whether a gradient branch is from the HR gradient map, the gradient discriminator is defined as follows:

$$L_{GM}^{Dis} = -E_{GM}[\log(1 - D(M(I^{SR})))] - E_{GM}[\log(D(M(I^{HR})))] \tag{13}$$

To supervise the generation of SR results by adversarial learning, the adversarial loss of the gradient branch is defined

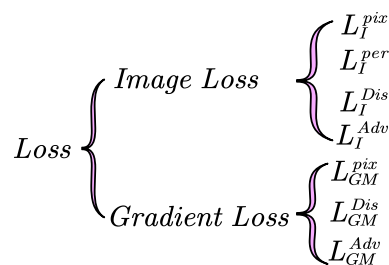


Fig. 6 Composition structure of the loss function

as follows:

$$L_{GM}^{Adv} = -E_{GM}[\log D(G(M(I^{LR})))] \tag{14}$$

(6) Overall loss

Combining the above various types of losses, the overall loss function is obtained as follows:

$$L = aL_I^{pix} + bL_I^{per} + cL_I^{Dis} + dL_I^{Adv} + eL_{GM}^{pix} + fL_{GM}^{Dis} + gL_{GM}^{Adv} \tag{15}$$

where a, b, c, d, e, f, g are the weighting parameters which meet the condition:

$$a + b + c + d + e + f + g = 1 \tag{16}$$

3 Dataset and experiment analysis

3.1 Dataset and experimental settings

In our experiments, we use the public infrared image dataset titled ‘‘A dataset for infrared detection and tracking of dim-small aircraft targets underground/air background,’’ downloaded from <http://www.csddata.org/p/387/>. The dataset contains 22 sequence images from 22 independent videos, totaling 16,177 frames. Each frame is in the 3–5 μ m mid-infrared band, having 256×256 resolution, 8-bit depth, 193 KB size, and bmp format.

Before training, we extracted 3235 frames at five frame intervals from the 16,177 frames. The samples are down-sampled by $\frac{1}{4}$ to a resolution of 64×64 , constituting the LR images, while the original images with 256×256 resolution are the corresponding HR images. So far, we have built a dataset containing 3235 LR-HR image pairs. In the experiment, we randomly selected 70% of the image pairs in the dataset are used for model training, while the remaining 30% are used for model testing.

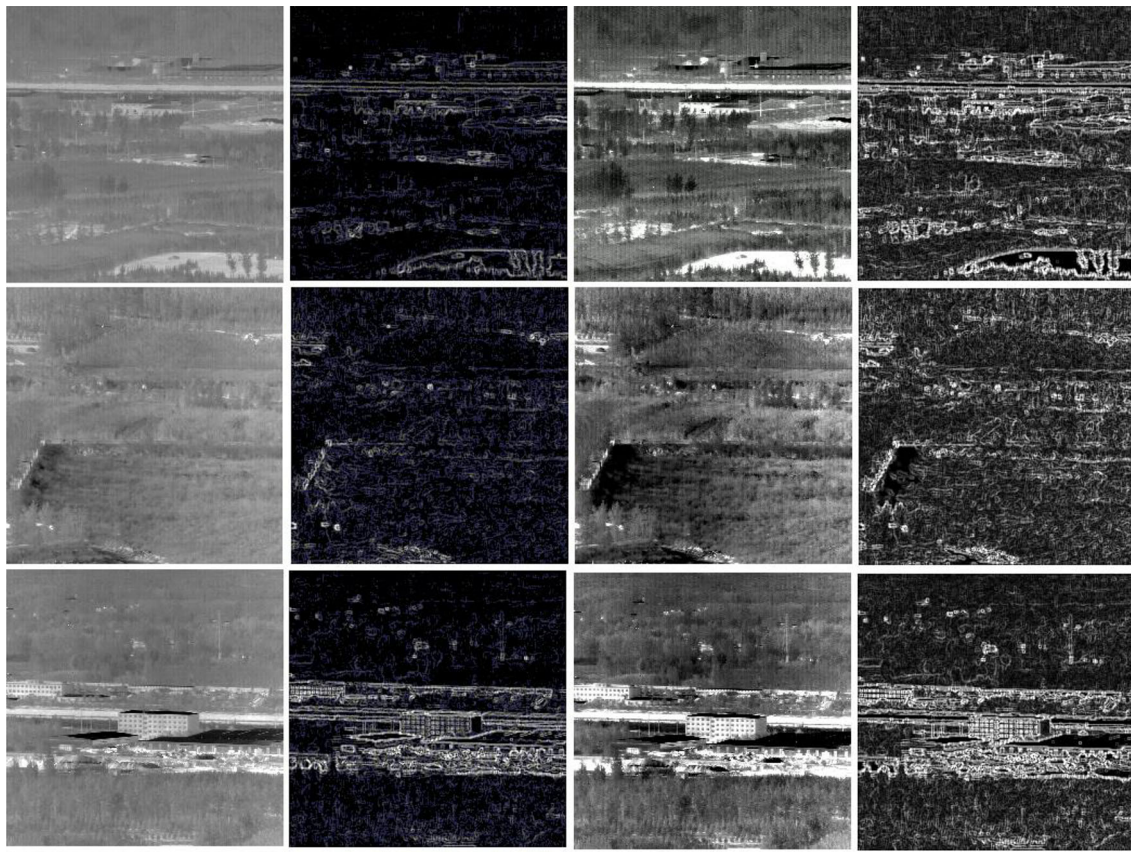


Fig. 7 Experiment results of haze removal

3.2 Infrared image haze removal experiment

The window radius Ω for haze removal is 5, the rate of haze removal ω is 0.95, and the lowest transmission t_0 is 0.1. The first and third columns of Fig. 7 show the images before and after the haze removal, respectively. Seem from the results, the contrast of image after haze removal is improved obviously compared with the original. More crucially, some blurred details become clearer.

Gradient maps are also extracted using the Sobel operator. The second and fourth columns of Fig. 7 show gradient maps before and after the haze removal. The fourth column evidently has more gradient information. This suggests that more details are recovered by removing haze.

3.3 Qualitative comparison

We compare our method to other single-frame SR methods, such as bicubic, SRResNet [12], SRGAN [20], and ESRGAN [21]. Figure 8 shows the 484th frame of data15, which contains a tower. The tower in the result of the bicubic SR is still blurry. Although the tower in the results of SRRes-

Net, SRGAN, and ESRGAN gradually becomes clear, they have different degrees of geometric distortion. The geometric structure of the tower in our method preserving better compared with above methods. In terms of visual effects, our results are more natural and realistic than other methods.

Figure 9 shows the 35th of data20 containing a house. The SR result of the bicubic is still blurry. Though the house edges in the results of SRGAN and ESRGAN become sharp, there are still noticeable geometric distortions, as indicated by red ellipses. The house edge in the SRResNet and our method is clear and with no distortion. But on finer objects, such as distant buildings marked with red boxes, our method is cleaner and more natural.

More comparisons are shown in Figs. 10, 11, and 12. Our method's better performance is primarily attributable to the following three aspects: (1) More details are recovered by haze removal; (2) gradient map is extracted and used to guide image SR; and (3) the gradient losses are added, which imposes secondary constraints on the image SR for preserving structural information. As a result, our outcome is more natural and realistic, and preserves geometric structure better.

Fig. 8 A steel tower in 484th frame of data15 [4 × super-resolution]

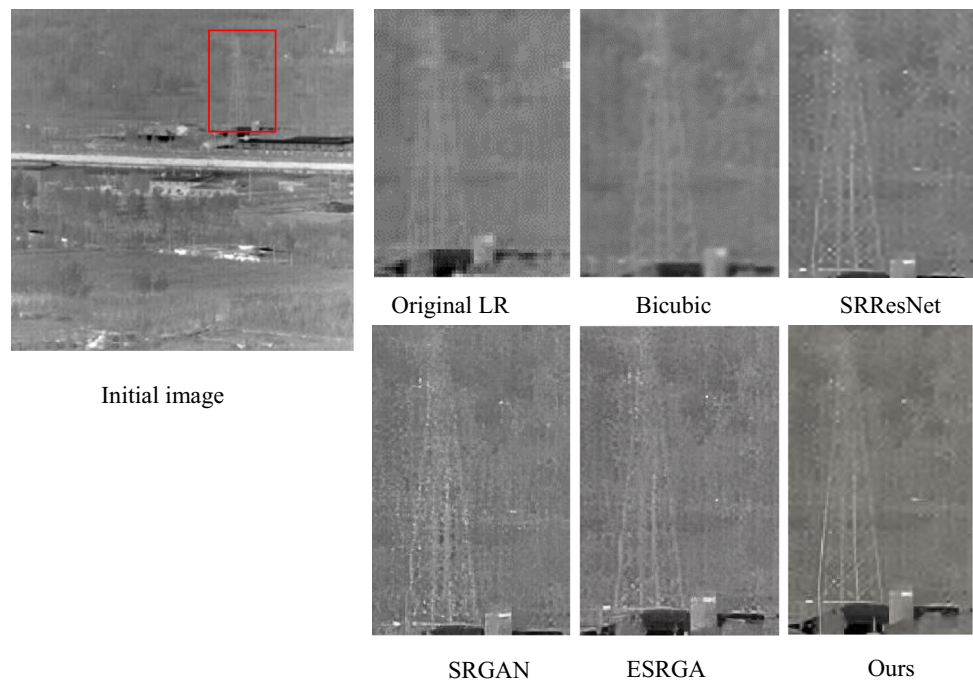
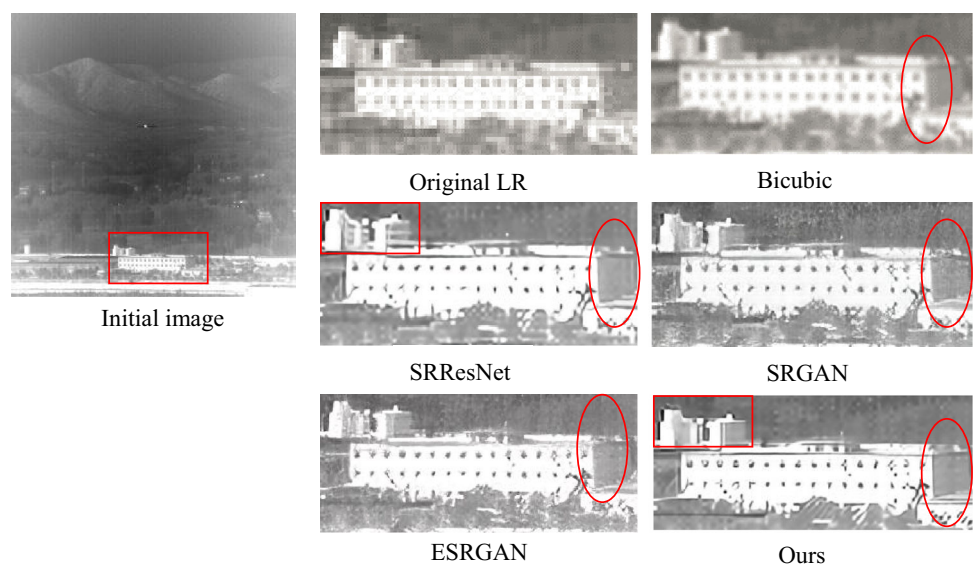


Fig. 9 A house in 35th frame of data20 [4 × super-resolution]



3.4 Quantitative comparison

We use PSNR (peak signal-to-noise ratio), SSIM (structure similarity [29]), and PI (perceptual index [30]) to quantitatively evaluate the SR performance. The value range of SSIM is [− 1 1]. The higher the PSNR and SSIM indicators are, the better, while the lower the PI is, the better.

A traditional method cubic interpolation, a filter-based method SelfExSR [31], and three deep learning methods (SRResNet [12], SRGAN [20], and ESRGAN [21]) are chosen for comparison. The results are summarized in Table 1. The best results for each indicator are shown in bolditalic,

and the next best are in *italic bold*. Our technique is best for PSNR and SSIM, as indicated in the table, whereas ESRGAN is best for PI. In case of PI, our method is somewhat inferior to ESRGAN; however, the difference is not discernible. Our method outperforms other methods on the SSIM, about 0.03 higher than the second SRResNet.

3.5 Ablation experiments

To verify the effectiveness of each part of the model, ablation experiments are also conducted on the dataset introduced in Sect. 3.1. We also use three indicators (PSNR, SSIM, and

Fig. 10 An aircraft in 348th frame of data16 [4 × super-resolution]

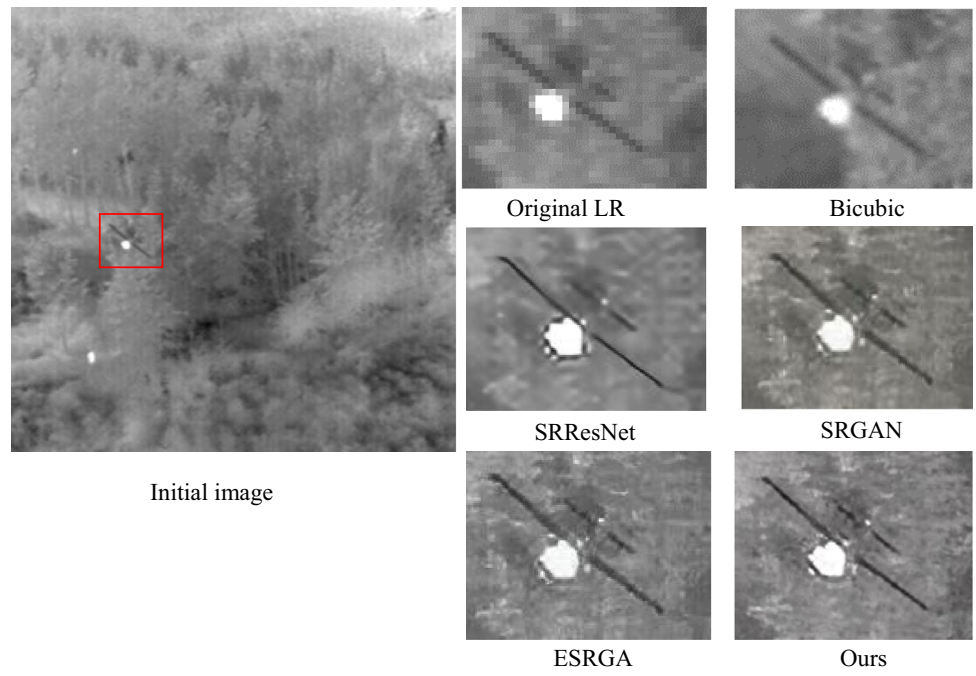


Fig. 11 A road in 26th frame of data6 [4 × super-resolution]

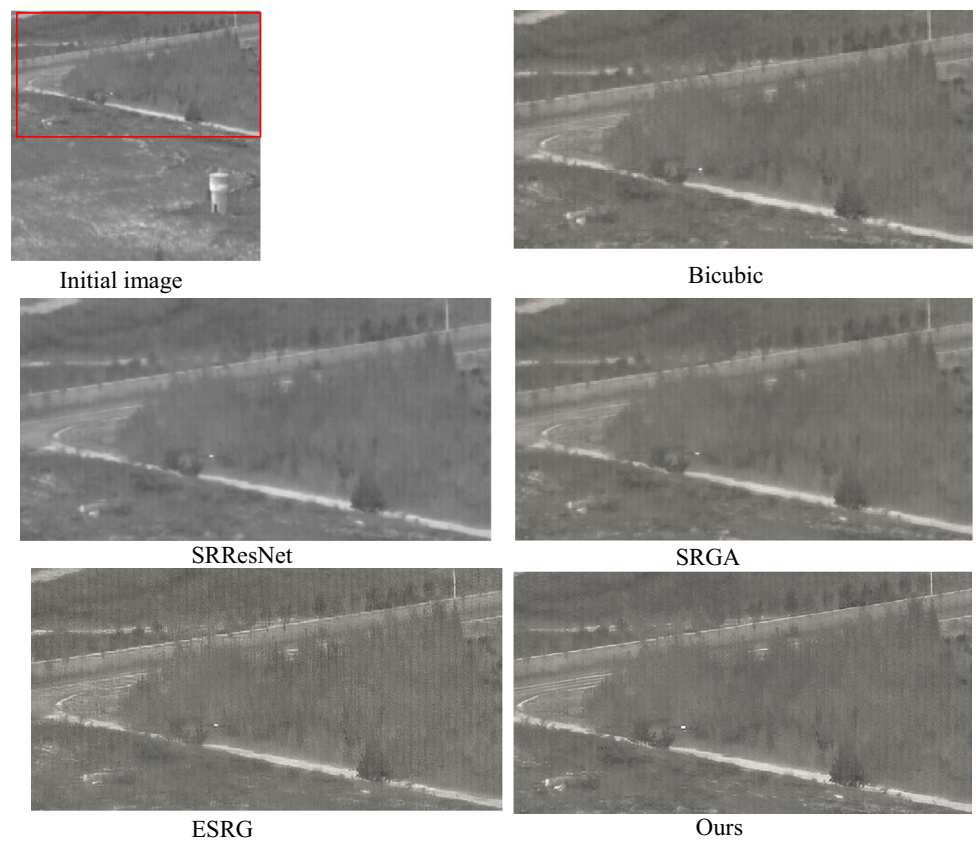


Fig. 12 A steel tower and a road in 35th frame of data19 [4 × super-resolution]

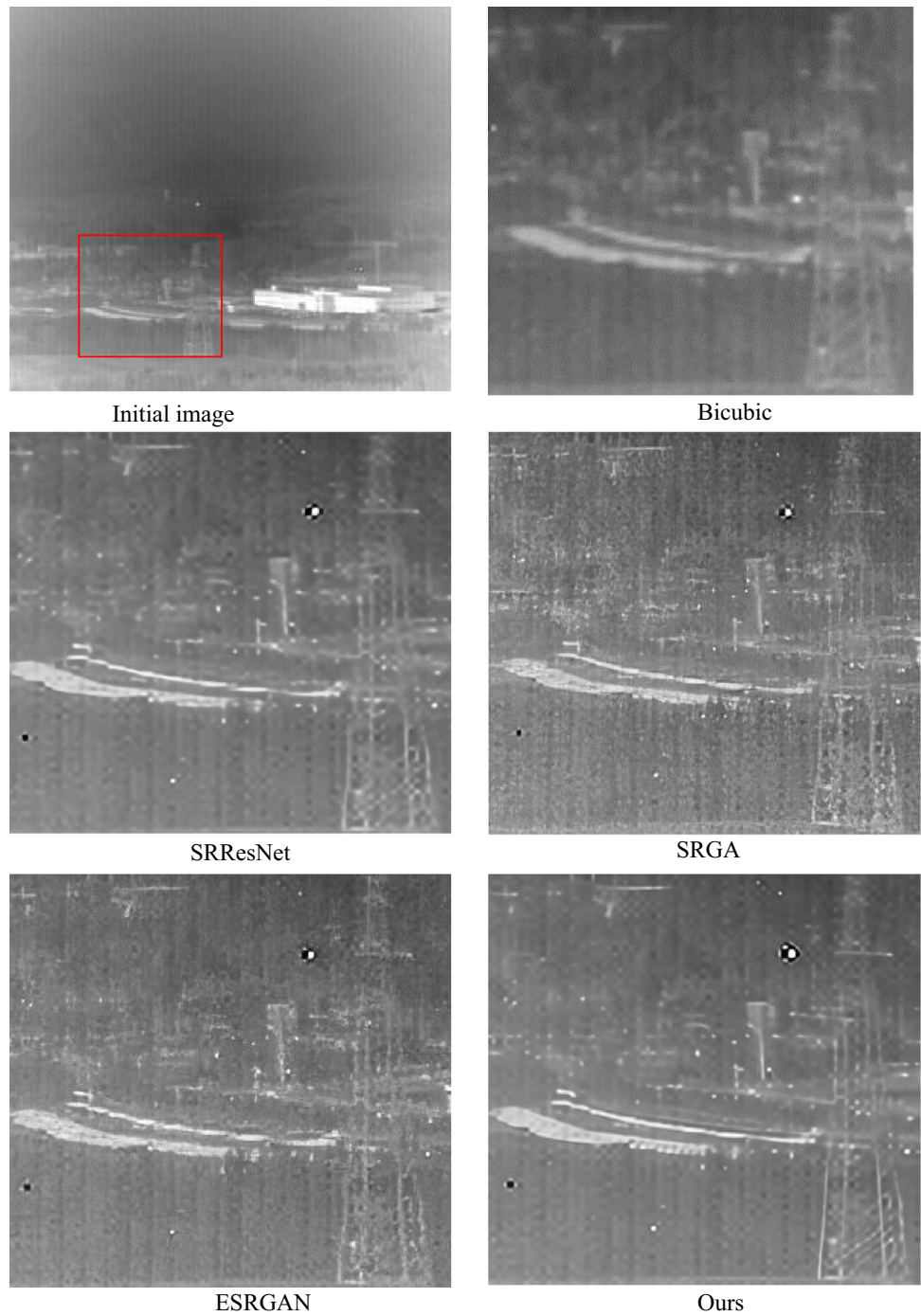


Table 1 Quantitative comparison performed on the public infrared image dataset

	Bicubic	SelfExSR	SRResNet	SRGAN	ESRGAN	Ours
avgPSNR	24.38	27.40	27.87	26.57	27.66	28.46
avgSSIM	0.6815	0.7518	0.8734	0.8269	0.8653	0.9012
avgPI	4.849	–	2.736	2.815	2.643	2.708

Table 2 Results of the ablation experiment

	Only image SR branch	Without haze removal	Without gradient SR branch	Without fusion block	Complete
PSNR	27.66	28.15	25.13	28.30	28.46
SSIM	0.8653	<i>0.8927</i>	0.7348	0.8846	0.9012
PI	2.643	2.723	2.901	2.832	2.708

PI) to quantitatively evaluate the SR performance. The best results for each indicator are shown in bolditalic, and the next best are in italic bold.

In the first experiment, we cut off the gradient branch and remove the fusion block. It is essentially a single-branch network that is ESRGAN. In the second experiment, we only remove the haze removal block. In the third experiment, we only remove the gradient SR branch. In the fourth experiment, we simply replace the fusion block with concat block, that is, the SR results of the two branches are directly added together. The last is the complete version of our method. The results of the ablation experiments are shown in Table 2.

When removing the haze removal module, the PSNR declined 0.31, the SSIM only slight declined 0.0085, and the PI raised 0.15. It proved that haze removal is beneficial for infrared image SR. When removing the gradient SR branch, the three indicators all became worse obviously. This may be because the gradient map is still LR while the image is HR. There will be a wrong correspondence between the LR gradient map and the HR image. When replacing the fusion block with concat block, the PSNR declined 0.16, the SSIM only slight declined 0.0166, and the PI raised 0.124. The fusion block is also beneficial for infrared image SR.

3.6 Computational cost analysis

Our experiment was carried out in the following environment: GPU 2080Ti, Intel(R) Xeon(R) CPU E5-2660 v2 2.20 GHz, RAM 32.0G; 64-bit Windows OS. When SR reconstructing an image from 256×256 pixels to 1024×1024 pixels, the average time required by ESRGAN is 0.233 s, while ours is 0.257 s. Though the computational cost of our method is slightly more expensive than the ESRGAN, it is still within the acceptable range.

4 Conclusions

The visible image SR methods do not perform well when directly used for infrared image SR. It is because that infrared image has weaker contrast and fewer details than visible image. To get high-quality SR from single-frame infrared images, this paper proposes a dual-branch deep neural network. The image SR branch reconstructs the SR image from

the initial infrared image using a basic structure similar to the enhanced SR generative adversarial network (ESRGAN). The gradient SR branch removes haze, extracts the gradient map, and reconstructs the SR gradient map. To obtain more natural SR image, a fusion block based on attention mechanism is adopted between these branches. To preserve the geometric structure, gradient L1 loss and gradient GAN loss are defined and added.

Experimental results on a public infrared image dataset demonstrate that, compared with the current SR methods, the proposed method is more natural and realistic, and can better preserve the structures.

In the future, we will study how to generate SR gradient images with only a single branch using gradient map as the guided filter.

Acknowledgements This work was supported by Key Laboratory Fund of Basic Strengthening Program (JKWATR-210503), Changsha Municipal Natural Science Foundation (kq2202067), and the Basic Science and Technology Research Project of the National Key Laboratory of Science and Technology on Automatic Target Recognition of Scientific Research under Grant (WDZC20205500209).

Declarations

Conflict of interest We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

References

1. Fangzhe, N., Yurong, Q., Yanni, X., et al.: Survey of single image super resolution based on deep learning. *Appl. Res. Comput.* **37**(02), 321–326 (2020)
2. Dong, C., Loy, C.C., He, K. et al.: Learning a deep convolutional network for image super-Resolution. *Computer Vision-ECCV 2014*. [S.I.]: Springer International Publishing, pp. 184–199 (2014)
3. Kim, J., Lee, J.K., Lee, K.M.: Accurate image super-resolution using very deep convolutional network, pp. 1646–1654 (2015)
4. Zhang, K., Zuo, W., Gu, S. et al.: Learning deep CNN denoiser prior for image restoration. In: *Proc of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2808–2817 (2017)
5. Tai, Y., Yang, J., Liu, X. et al.: MemNet: a persistent memory network for image restoration. In: *Proc of IEEE International Conference on Computer Vision*. [S.I.]: IEEE Computer Society, pp. 4549–4557 (2017)
6. Lai, W., Huang, J., Ahuja, N. et al.: Deep laplacian pyramid networks for fast and accurate super-resolution. In: *Proc of IEEE*

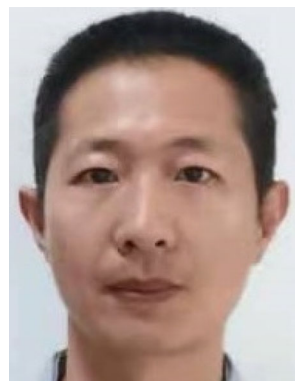
- International Conference on Computer Vision. [S.I.]: IEEE Computer Society, pp. 5835–5843 (2017)
7. Tai, Y., Yang, J., Liu, X.: Image super-resolution via deep recursive residual network. In: Proc of IEEE Conference on Computer Vision and Pattern Recognition. [S.I.]: IEEE Computer Society, pp. 2790–2798 (2017)
 8. Sumei, Li., Fan, Ru., Guoqing, L., et al.: A two-channel convolutional neural network for image super-resolution. *Neurocomputing* **275**, 267–277 (2018)
 9. Dong, C., Loy, C.C., Tang, X.: Accelerating the super-resolution convolutional neural network. In: Proc of European Conference on Computer Vision. Berlin: Springer, pp. 392–407 (2016)
 10. Xiaojiao, M., Chunhua, S., Yubin, Y.: Image restoration using convolutional auto-encoders with symmetric skip connections. *Res. Gate* **6**, 391 (2016)
 11. Shi, W., Caballero, J., Huszár, F. et al.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1874–1883 (2016)
 12. Ledig, C., Theis, L., Huszar, F. et al.: Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. 2016(9):105–114.
 13. Lim, B., Son, S., Kim, H. et al.: Enhanced deep residual networks for single image super-resolution. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE Computer Society, pp. 1132–1140 (2017)
 14. Li, L., Tang, J., Ye, Z., et al.: Unsupervised face super-resolution via gradient enhancement and semantic guidance. *Vis. Comput.* **37**, 2855–2867 (2021)
 15. Tong, T., Li, G., Liu, X. et al.: Image super-resolution using dense skip connections. In: 2017 IEEE International Conference on Computer Vision (ICCV). IEEE Computer Society, pp. 4809–4817 (2017)
 16. Amaranageswarao, G., Deivalakshmi, S., Ko, S.B.: Joint restoration convolutional neural network for low-quality image super resolution. *Vis. Comput.* **38**(1), 31–50 (2022)
 17. Danya, Z., Yepeng, L., Xuemei, Li., et al.: Single-image super-resolution based on local biquadratic spline with edge constraints and adaptive optimization in transform domain. *Vis. Comput.* **38**(1), 119–134 (2022)
 18. Zhang, Y., Tian, Y., Kong, Y. et al.: Residual dense network for image restoration. 2018(2):180.
 19. Zhou, F., Li, X., Li, Z.: High-frequency details enhancing dense net for super-resolution. *Neurocomputing* **290**, 34–42 (2018)
 20. Ledig, C., Theis, L., Huszar, F. et al.: Photo-realistic single image super-resolution using a generative adversarial network. Preprint <https://arxiv.org/abs/1609.04802> (2016)
 21. Wang, X., Yu, K., Wu, S. et al.: ESRGAN: enhanced super-resolution generative adversarial networks. In: The European Conference on Computer Vision Workshops (ECCVW), pp. 1–23 (2018)
 22. Mao, R.: Single Infrared Image Super-resolution and Enhancement Based on Fusion ESRGAN and Gradient Network. 2020, Xian University of Technology.
 23. Wang, X., Zhang, K., Yan, J., et al.: Infrared image complexity metric for automatic target recognition based on neural network and traditional approach fusion. *Arab. J. Sci. Eng.* **45**(4), 3245–3255 (2020)
 24. Ma, C., Rao, Y., Cheng, Y. et al.: Structure-Preserving Super Resolution with Gradient Guidance.2020, <https://arxiv.org/abs/2003.13063>
 25. Nayak, R., Balabantaray, B.K., Patra, D.: A new single-image super-resolution using efficient feature fusion and patch similarity in non-euclidean space. *Arab. J. Sci. Eng.* **45**(12), 10261–10285 (2020)
 26. Kaiming, H., Jian, S., Xiaoou, T.: Single Image Haze Removal Using Dark Channel Prior. *CVPR* (2009)
 27. Gautam, A., Singh, S.: Neural style transfer combined with EfficientDet for thermal surveillance. *The Visual Computer*, pp. 1–17 (2021)
 28. Songchen, H., Changxin, H., Wei, Li., et al.: An improved dehazing algorithm based on near infrared image. *Adv. Eng. Sci.* **50**(2), 347–356 (2018)
 29. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., et al.: Image quality assessment: from error visibility to structural similarity. *TIP* **13**(4), 600–612 (2004)
 30. Yochai, B., Roey, M., Radu, T., Tomer, M., Lihi, Z-M.: The 2018 pirm challenge on perceptual image super-resolution. In: *ECCV*, Springer, pp. 334–355 (2018)
 31. Huang, J.B., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston (2015)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



Huang Zhijian received the Ph.D. degree in information and communication engineering from the National University of Defense Technology, Changsha, China, in 2014. He is currently an Associate Professor with the School of Computer Science and Engineering, Changsha University. His current research interests include the infrared image processing and infrared target detection.



Hui Bingwei received the Ph.D. degree in information and communication engineering from the National University of Defense Technology, Changsha, China, in 2013. He is currently an Associate Research with the School of Computer Science and Engineering, Changsha University. His current research interests include the infrared target detection and the data science for ATR.



Shujin Sun received the B.S., M.S., and Ph.D. degrees in information and communication engineering from the National University of Defense Technology, Changsha, China, in 2009, 2012 and 2016, respectively. He is currently an Assistant Professor with the Hunan Province Key Laboratory of Industrial Internet Technology and Security. His current research interests include the development of algorithms for information extraction and classification remote sensing

images using machine learning algorithms.



Ma Yanxin received the Ph.D. degree in information and communication engineering from the National University of Defense Technology, Changsha, China, in 2018. He is currently an Associate Professor with the College of Meteorology and Oceanology, National University of Defense Technology. His current research interests include the infrared image processing and infrared target detection.