

Gradient-aware channel attention network for infrared small target image denoising before detection

LIN Zai-Ping¹, LUO Yi-Hang¹, LI Bo-Yang¹, LING Qiang¹, ZHENG Qing²,
YANG Jing-Yi³, LIU Li¹, WU Jing^{1*}

- (1. College of electronic science and technology, National University of Defense Technology, Changsha 410073, China;
2. Department of Military Representative Bureau of Aerospace Systems, Beijing 100000, China;
3. Shanghai Institute of Satellite Engineering, Shanghai 200000, China)

Abstract: Infrared small target denoising is widely used in military and civilian fields. Existing deep learning-based methods are specially designed for optical images and tend to over-smooth the informative image details, thus losing the response of small targets. To both denoise and maintain informative image details, this paper proposes a gradient-aware channel attention network (GCAN) for infrared small target image denoising before detection. Specifically, we use an encoder-decoder network to remove the additive noise of the infrared images. Then, a gradient-aware channel attention module is designed to adaptively enhance the informative high-gradient image channel. The informative target region with high-gradient can be maintained in this way. After that, we develop a large dataset with 3981 noisy infrared images. Experimental results show that our proposed GCAN can both effectively remove the additive noise and maintain the informative target region. Additional experiments of infrared small target detection further verify the effectiveness of our method.

Key words: infrared small target, denoising before detection, gradient-aware channel attention

基于梯度可感知通道注意力模块的红外小目标检测前去噪网络

林再平¹, 罗伊杭¹, 李博扬¹, 凌 强¹, 郑 晴², 杨晶贻³, 刘 丽¹, 吴 京^{1*}

- (1. 国防科技大学 电子科学学院, 湖南 长沙 410073;
2. 航天系统部装备部军事代表局, 北京 100000;
3. 上海卫星工程研究所, 上海 200000)

摘要: 红外图像去噪在军事及民用领域应用广泛。现有基于深度学习的图像去噪方法主要为可见光图像设计, 此类方法容易过度平滑图像细节, 从而导致弱小目标丢失, 为后续的检测任务带来困难。为了在去除噪声的同时保留好红外图像中的目标信息, 本文提出了一种基于梯度可感知通道注意力模块的红外弱小目标检测前去噪网络。该网络首先采用编码器-解码器结构来去除图像中的加性噪声, 然后通过梯度可感知通道注意力模块对图像高频区域进行自适应增强, 有效保持红外弱小目标的响应强度。此外, 本文提出了领域第一个包含 3 981 张含噪声的红外图像数据集。实验结果表明, 该网络能够在有效去除加性噪声的同时避免过度平滑, 很好地保留了红外图像中的目标信息, 最终实现了在含噪声环境下的高鲁棒性红外弱小目标检测。

关 键 词: 红外小目标; 检测前去噪; 梯度可感知通道注意力模块

中图分类号: TP753 文献标识码: A

Received date: 2023- 06- 29, revised date: 2023- 12- 21

收稿日期: 2023- 06- 29, 修回日期: 2023- 12- 21

Foundation items: Supported by the National Natural Science Foundation of China (62001478, 61972435), Aviation Science Foundation Project Contract (ASFC-20165188004), Shanghai Aerospace Science and Technology Innovation Fund (SAST2021-035), Independent Research Fund of Key Laboratory of Military Scientific Research

Biography: LIN-Zaiping (1982-), male, Zhejiang, PHD. Research area involves space-based information processing. E-mail: linzaiping@sina. com. LIN Zai-ping and LUO Yi-hang contributed equally to this work.

* **Corresponding author:** E-mail: jingwu@nudt. edu. cn

Introduction

With the rapid development of infrared imaging technology, the infrared imaging system has been widely used in marine resource utilization, high-precision navigation, and ecological environment monitoring^[1-6]. Since IR imaging device is generally applied to long-range imaging, the imaging quality of infrared imaging system is easily disturbed by terrible environment, which includes internal imaging-device environment (e. g., thermal noise of amplifiers and detectors) and external natural environment (e. g., clouds, low-light conditions and atmospheric perturbations)^[7]. Therefore, noises with different characteristics generally interact with each other and perform complex distribution in IR images. To simplify the mixed noise, one common assumption is that the noise in IR images is additive white Gaussian noise (AWGN) with standard deviation^[8]. As shown in Fig. 1(a1-a3), IR images in the same scene would be corrupted under different levels of noise caused by the varied conditions of the imaging device and the external environment. The detection results generated by DNANet^[9] under different levels of noise are shown in Fig. 1(b1-b3). It demonstrates that the additive noise not only introduces the decrease of image quality but also brings obvious performance decrease for the subsequent detection task. To our surprise, as shown in Fig. 1(c-d), our denoising method helps to make the noisy image recover to a clean one and thus alleviate the performance decrease of target detection task.

To alleviate the negative effect caused by the additive noise, numerous traditional methods have been proposed, including the filtering-based method^[10], sparse-representation-based methods^[11-12], and Low-rankness-based method^[13]. Although the above works have achieved promising image denoising results, they are essentially manually-designed methods, which heavily rely on prior knowledge and hand-crafted features. When the characteristics of images (e. g., signal-to-cluster ratio (SCR)) dramatically change, traditional methods can hardly handle such changeable scenarios with fixed hyper-parameters. More robust solutions should be introduced to tackle such challenges.

Different from the previous model-driven traditional methods, the convolutional neural network (CNN) can achieve high-performance image denoising in a data-driven manner and has yielded promising results in optical image denoising. Jain *et al.*^[14] proposed the first CNN-based denoising method. A four-layer, fully-connected CNN structure was designed to achieve significant improvements over traditional denoising methods. Due to the simple and shallow CNN structure, the denoising performance is limited. Then, Zhang *et al.* proposed a denoising convolutional neural network (DnCNN)^[15]. DnCNN can remove the latent clean image from noisy observation through a residual learning strategy. Thanks to the powerful representation ability introduced by much deeper CNN layers, DnCNN achieves better noise reduction than the optimal traditional method^[10] and previous CNN-based methods^[14-15]. After that, Liang *et al.* de-

signed a strong baseline model SwinIR^[21] for image restoration based on the Swin Transformer. Higher performance under real noisy scene is achieved. However, the performance improvement is based on the huge number of optical images. The capacity of IR datasets is limited and hard to drive the transformer-based network. Moreover, the IR imaging system is generally used for long-distance imaging to capture small and dim targets which are not easily perceived by optical devices. Therefore, direct transfer of the existing optical denoising method may over-smooth the small targets and thus lose the response of the small target, which is unacceptable for subsequent high-level target detection and recognition tasks.

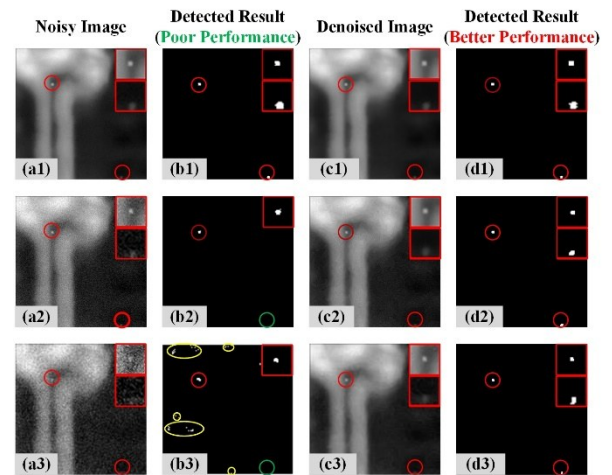


Fig. 1 (a1)-(a3) Visual results of noisy input images; (b1)-(b3) detected results without denoising; (c1)-(c3) denoised images by our method; (d1)-(d3) detected results with denoising
图1 (a1)-(a3)含噪声输入图像; (b1)-(b3)未去噪直接检测结果; (c1)-(c3)本文算法去噪后的图像; (d1)-(d3)去噪后检测结果

To both denoise IR images and maintain the response of small targets, we propose a novel infrared image denoising method named gradient-aware channel attention network (GCAN). We design an encoder-decoder-based network with residual connections to remove the additive noise of infrared images. Then, a gradient-based channel attention module (GCAM) is designed and embedded into the residual connection to adaptively enhance the informative high-gradient image channel and thus preserve the informative details. In this way, informative target regions with a high gradient can be preserved and additive noise of IR images is also removed.

The contributions of this paper can be summarized as follows:

1) An encoder-decoder denoising framework and a gradient-based channel attention module are proposed to remove the additive noise and adaptively enhance the informative image channels, respectively.

2) We develop an NUDT-IRSTDn dataset with various SCR ratios based on our previous NUDT-SIRST dataset. Both IR image denoising performance and corre-

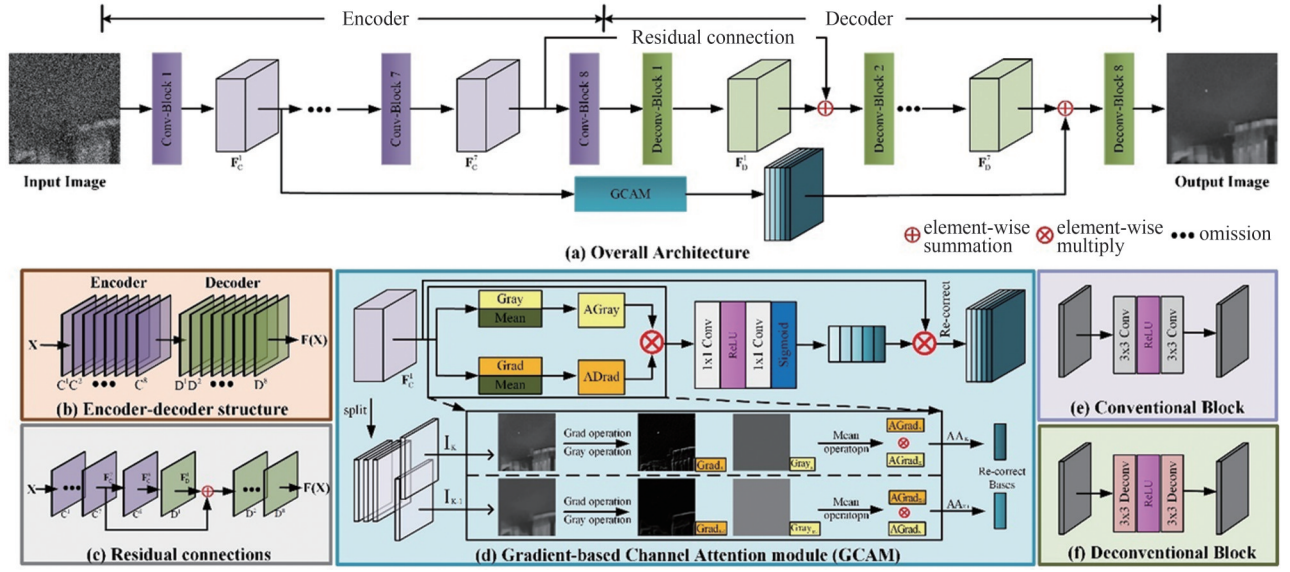


Fig. 2 An illustration of the proposed gradient-aware channel attention network (GCAN) for infrared small target image denoising before detection

图2 基于梯度可感知通道注意力模块的红外小目标检测前去噪网络示意图

sponding influence on subsequent target detection tasks can be evaluated.

3) The experimental results of both denoising and high-level object detection demonstrate that our GCAN can not only achieve high-performance of denoising compared to other state-of-the-art methods, but also effectively keep the performance of subsequent detection tasks stable under terrible imaging conditions.

1 Methodology

1.1 Denoise model

Assuming that $X \in \mathbb{R}^{m \times n}$ is a noise disturbance image and $Y \in \mathbb{R}^{m \times n}$ is a corresponding clean image, the relationship between them can be formulated as:

$$X = \delta(Y) \quad (1)$$

where $\delta: \mathbb{R}^{m \times n}$ denotes the complex degradation process involving internal and external IR imaging conditions.

The noise reduction process aims to recover the clean images from the degraded images. This process can be transformed to seek a function f to minimize the mse error between $f(x)$ and Y , which can be described as:

$$\arg \min_f \|f(X) - Y\|_2 \quad (2)$$

where f is regarded as the optimal approximation of δ^{-1} , and $f(X)$ denotes the recovered clean image.

1.2 Infrared image denoising network

1) Overall architecture: In this section, we introduce our infrared image denoising network (GCAN) in detail. First, we follow the encoder decoder-based architecture and combine with residual connections to remove the varied additive noise and initially pass image details to the top layers. It is worth noting that pooling layers and the ReLU layers are removed before the summation with residuals to avoid losing details. Then, we propose a gradient-based channel attention module to maintain the potential target regions (e. g., high-gradient region)

while denoising images. The overall architecture of the GCAN is shown in Fig. 2.

2) Encoder-decoder structure: The encoder-decoder structure consists of several stacked Conv-Blocks and Deconv-Blocks. The encoder part is designed to suppress image noise from low-level to high-level step by step while preserving informative information in the input images. As shown in Fig. 2(b), the preprocessed IR image X is first fed into sequential convolutional blocks (Conv-Block C^{th} ($th = 1, 2, \dots, N$)). After the stacked Conv-Blocks, the image X is transformed into a feature space, and the output of each Conv-Block is a feature map $F_c^{th} \in \mathbb{R}^{C^{th} \times H \times W}$ ($th \in \{1, 2, 3, \dots, N\}$). Then, the data flow through the Deconv-Blocks (D^{th} ($th=1, 2, \dots, N$)) follows the rule of FILO (First In Last Out). The feature from the last Conv-Block $F_c^N \in \mathbb{R}^{N \times H \times W}$ is fed to the first Deconv-Block to generate $F_d^1 \in \mathbb{R}^{C^1 \times H \times W}$. Finally, $F_c^1 \in \mathbb{R}^{C^1 \times H \times W}$ and $F_d^{N-1} \in \mathbb{R}^{C^{N-1} \times H \times W}$ are fed into the D^N to generate the recovered image $F(X)$. The output of C^{th} can be formulated as:

$$F_c^{th} = w_j * [\text{ReLU}(w_i * X_{th-1} + b_i) + b_j] \quad (3)$$

Each Deconv-Block is symmetric with the corresponding Conv-Block, and the output of D^{th} can be formulated as:

$$F_d^{th} = w'_j \otimes [\text{ReLU}(w'_i \otimes X_{th-1} + b'_i) + b'_j] \quad (4)$$

where th ($th \in 1, \dots, N$) is the number of Blocks. w_i and b_i denote the weights and biases in the i ($i \in 1, \dots, I$) convolutional layer, respectively. $*$ and \otimes represent convolution and deconvolution operator, respectively.

X_0 is the input image, and X_k ($k > 0$) is the extracted feature from the previous layers. $\text{ReLU}(X) = \text{Max}(0, X)$ is the activation function.

3) Residual connections: The residual connection is used to avoid gradients vanishing as the network goes deep, and also serves as a simple detail recovery structure that can connect matched Conv-Blocks and Deconv-

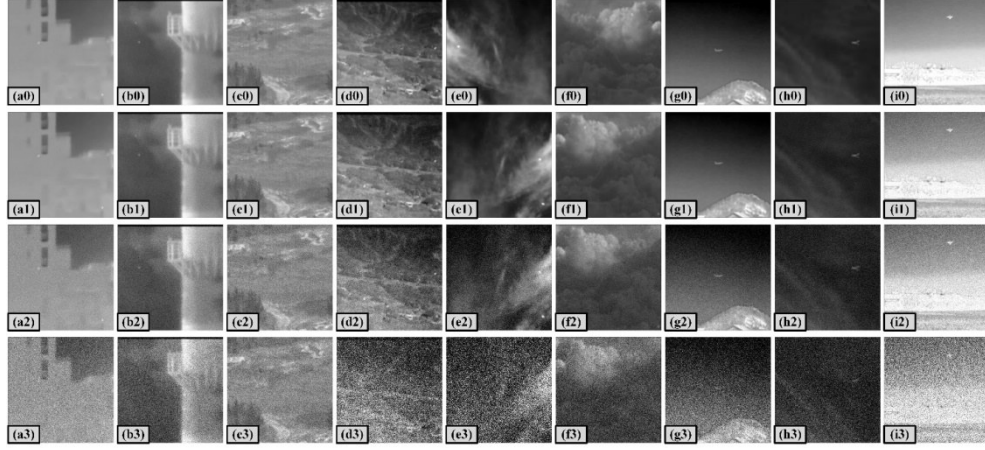


Fig. 3 Examples of the developed dataset, including (a0)-(i0) clean images; (a1)-(i1) level-1 noisy images; (a2)-(i2) level-2 noisy images; (a3)-(i3) level-3 noisy images
图3 数据集示例, (a0)-(i0) 无噪声图片; (a1)-(i1) 噪声等级为1的图片; (a2)-(i2) 噪声等级为2的图片; (a3)-(i3) 噪声等级为3的图片

Blocks to propagate the informative details from low-level to high-level features. As shown in Fig. 2(c), after the element-wise sum between the feature $F_d^1 \in \mathbb{R}^{C^1 \times H \times W}$ and $F_c^{N-1} \in \mathbb{R}^{C^{N-1} \times H \times W}$, the obtained map is fed into next Deconv-Block D^2 to generate the same scale feature map $F_d^2 \in \mathbb{R}^{C^2 \times H \times W}$.

4) Gradient-based channel attention module (GCAM): To avoid over-smooth the informative small target region, we design a GCAM as shown in Fig. 2(d) to adaptively enhance the informative image channel and enhance the target regions with high gradient. GCAM enhances details by the feature rescaling strategy. Inspired by no-reference image quality metrics, we use average gray to represent the amount of information in the feature map, and average gradient to describe the amount of high-intensity information. GCAM takes the output of first Conv-Block $F_c^1 \in \mathbb{R}^{C^1 \times H \times W}$ as input and computes $\text{Gray}_K \in \mathbb{R}^{C^1 \times H \times W}$ and $\text{Grad}_K \in \mathbb{R}^{C^1 \times H \times W}$ for the I_K channel of F_c^1 . The Gray operation and the Grad operation are calculated as follows:

$$\text{Gray}_K = \sum_{i=1}^M \sum_{j=1}^N \sqrt{\frac{(I_K(i+1,j) - I_K(i,j))^2 + (I_K(i,j+1) - I_K(i,j))^2}{2}}, \quad (5)$$

where M and N represent the length and width of the image, respectively. Then Gray_K is fed to a mean

operation to generate AGray_K , respectively. After element-wise multiplication, $\text{AA}_K = \text{AGray}_K \otimes \text{AGray}_K$, GCAM can adaptively enhance the input feature map along the channel dimension.

2 The NUDT-IRSTDn dataset

2.1 Motivation

The high-quality dataset is essential for data-driven CNN-based methods. However, existing denoising methods are essentially data-driven and evaluated on their in-house dataset^[19]. Inspired by the single frame infrared small target detection dataset (NUDT-SIRST^[9]), we designed a large-scale infrared image dataset (namely, NUDT-IRSTDn) with different levels of noise to further

explore the influence of different levels of noise on high-level tasks (e. g., target detection).

These noisy images are manually synthesized by adding Gaussian white noise on those clean long-wave band IR images, whose wavelength locates between 8 μm and 14 μm . As shown in Table 1, three kinds of noise level are chosen (i. e., $\sigma = 0.05, 0.09$, and 0.25 for Noise-v1, Noise-v2, and Noise-v3). The original clean images can be regarded as the ground truth. Noise-v3 subset has the highest noise intensity among the three groups.

2.2 Implementation details

To simulate IR images subject to complex noise interference scenarios and better comparison of the influence of different noise intensities on subsequent tasks. We did not directly add the same levels of noise to the initial image. The synthesis process of our dataset is shown in Fig. 4. We first used LSCR as a quantitative metric of detection complexity and set three sets of detection thresholds T_{dec} (i. e., 5, 3.5, and 2). Then, we adopted an adaptive noise level function to adjust noise levels δ and make sure that the LSCR of adding noise IR image is less than T_{dec} . LSCR is defined as follows:

$$\text{LSCR} = \frac{|\mu_b - \mu_t|}{\sigma_b}, \quad (6)$$

where μ_b, μ_t, σ_b are the local background gray mean, target gray level mean, and local background gray standard deviation. We set the local background of the target as a rectangle centered at the target position with fixed width and height of 20 pixels. To eliminate the influence of the target region, we exclude the target region inside the rectangle. Some examples of the developed dataset are shown in Fig. 3.

As shown in Table 1, compared with the original noise-free NUDT-SIRST dataset, our developed NUDT-IRSTDn dataset provides much more number of images (i. e., 3981 vs 1327) under varied LSCR value. The LSCR value of NUDT-IRSTDn locates in 0.402-5, 0.402-3.5, and 0.402-2 for Noise v1, Noise v2, and

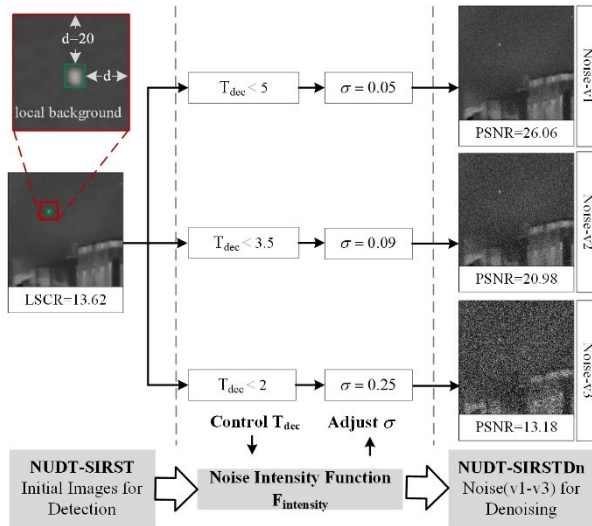


Fig. 4 Synthesis process of our dataset
图4 数据集合成步骤

Table 1 Main characteristics of NUDT-SIRST and NUDT-IRSTDn

表1 NUDT-SIRST和NUDT-IRSTDn数据集主要特性

Metrics	NUDT	NUDT-IRSTDn		
	SIRST	Noise. v1	Noise. v2	Noise. v3
<i>LSCR</i>	0.402~19.05	0.402~5	0.402~3.5	0.402~2
<i>LSCR'</i>	5.68	4.364	3.205	1.687
σ	—	0~0.06	0~0.1	0~0.5
σ'	—	0.013	0.04	0.154
<i>PSNR</i>	—	21.5~40.2	20.9~34.1	9.9~24.4
<i>PSNR'</i>	—	31.88	25.89	17.31
Number	1327	1327	1327	1327

Table 2 PSNR and SSIM values achieved by different denoising methods under varied noise-level dataset

表2 不同数据集下的多种去噪方法性能比较 (PSNR 和 SSIM)

Denoising Method	Noise. v1		Noise. v2		Noise. v3	
	PS NR	SSIM	PS NR	SSIM	PS NR	SSIM
BM3D	36.7	0.75	31.0	0.52	19.4	0.23
WNNM	34.6	0.38	33.1	0.36	30.3	0.28
K-SVD	35.2	0.62	34.0	0.43	31.2	0.27
RED CNN	36.8	0.87	35.6	0.82	29.5	0.74
DnCNN	44.3	0.93	40.3	0.91	33.6	0.87
SWINIR	44.9	0.92	41.7	0.97	34.3	0.87
GCAN	45.5	0.96	42.1	0.96	33.7	0.88

Noise v3, which are much smaller than that of NUDT-SIRST. Moreover, the average *LSCR* values (i. e., *LSCR'*) of NUDT-IRSTDn are 4.36, 3.20, and 1.68 for NUDT-IRSTDn with Noise v1, Noise v2, and Noise v3, respectively. More visually non-salient targets intro-

duce huge difficulty for precise detection.

3 Experiments

3.1 Experiment setting

1) Implementation Details: We conducted extensive experiments on the NUDT-IRSTDn dataset. To consist with the NUDT-SIRST dataset, we divided each group dataset into a training set and a test set with the ratio of 1:1. We resized all input IR images to 256×256 pixels. The batch size and learning rate in the process of network training were set as 8 and 1×e-5 respectively. We used the mean square error (MSE) as the loss function of our network. All models were implemented in PyTorch on a computer with an Intel Xeon Gold 5117 CPU and an Nvidia Tesla V100 GPU.

2) Evaluation Metrics: Following the previous works^[10, 15], we used PSNR and SSIM to evaluate the recovery image quality. We also adopted detection metrics (intersection over union (IoU), probability of detection (Pd) and false-alarm (Fa)) to evaluate the practical performance of denoising methods.

Table 3 Ablation study on our proposed GCAM module
表3 GCAM模块的消融实验结果

Method	#Params (dB)	FLOPs (G)	PSNR/SSIM (dB)
GCAN	1.848	83.89	44.6
w/o GCAM	2.345	157.30	45.5

3.2 Experimental results and analysis

1) Denoising results: To verify the superiority of our method, we compared our GCAN with state-of-the-art methods, including conventional model-based methods (BM3D^[10], WNNM^[13], and K-SVD^[11]) and CNN-based methods (RED CNN^[16] and DnCNN^[15]) on the NUDT-IRSTDn dataset. The proposed method and comparative methods are evaluated on the test set of the three subsets (i. e., Noise-v1, Noise-v2 and Noise-v3) of NUDT-IRSTDn. The results of PSNR and SSIM are presented in Table II. We can observe that our GCAN generates higher performance than the comparative three model-based methods and two learning-based methods in term of PSNR. Compared with DnCNN, GCAN has a much better denoising ability as shown in Table 2, our GCAN achieves much higher PSNR (i. e., 45.5 vs 44.3, 42.1 vs 40.3, and 33.7 vs 33.6 dB) than the DnCNN. It's worth noting that, 1 dB improvement of PSNR is high enough for the denoising task. It demonstrates that the superiority of our method to recover clean images. Meanwhile, the higher SSIM index also proves that our method has a stronger ability to recover accurate details and distinguish fine structure information from complex noise. The qualitative results are shown in Fig. 5. The zoomed images clearly show the regions of interest. It can be observed that GCAN suppresses different levels of noise and preserves the details of the target better. Compared to GCAN w/o GCAM, as shown in Table 3, our GCAN

achieves 0.9 dB performance increase (45.5 vs 44.6) in term of PSNR under NUDT-IRSTDn-v1 subset. That is because, our GCAM can adaptively enhance the input feature map along the channel dimension. More informative channel-dimension feature maps are enhanced, introducing better denoising results.

2) Effectiveness of Denoising for Detection: In this subsection, we evaluated the effectiveness of the denoising methods by comparing whether these methods can help the subsequent detection task maintain performance under a varied noisy environment.

Firstly, we evaluated the influence of additive noise on subsequent target detection. We selected five typical infrared small target detection methods (Top-hat^[17], RIPT^[18], ACM^[19], UNet^[20], and DNANet^[9]) to detect targets from the original image dataset and the corresponding three noise-level image datasets. The quantitative detection results on the four datasets are listed in Table 5. It can be observed that with the increase of noise intensity of the datasets (i. e., Oriset, Noise-v1, Noise-v2 and Noise-v3), the IoU value of the above five detection methods all gradually decreases. For example, after image denoising, the detection method (i. e., DNANet) achieves much better results (i. e., 1.6%, 1.6%, and 8.1×10^{-5} higher performance than DnCNN in term of IoU, Pd and FA on Noise-v1 subset). It is important for the infrared small target detec-

tion task under varied conditions of the imaging device and external environment.

Then, we compared the detection results on denoised images to evaluate the performance of denoising methods. We adopted Top-Hat^[17] and DNA-Net^[9] as the representatives of traditional and deep learning SIRST detection methods, respectively. As shown in Table 3, the improvements achieved by our GCAN over other denoising methods are obvious. It demonstrates that our GCAN achieves better performance on removing noise and retaining important details at different noise levels. Note that, the detection results on the denoised images with the WNNM method are even worse after denoising because of the over-smoothing of the target regions. Therefore, the denoising method for IR small target images needs to remove the noise while effectively retaining the details of the target region in the IR image, thus alleviating the degradation of detection performance under complex noise conditions.

3) Computational Efficiency: As shown in Table 6, GFLOPs, inference time (s), parameters, and PSNR performance of our GCAN are 157.30 GFLOPs, 0.206 s, 2.345 M, and 45.5 dB, respectively. Compared to three benchmark deep learning-based methods, our method achieves much better denoising performance in term of PSNR but introduces larger model size, longer inference time, and extra computation cost (i. e.,

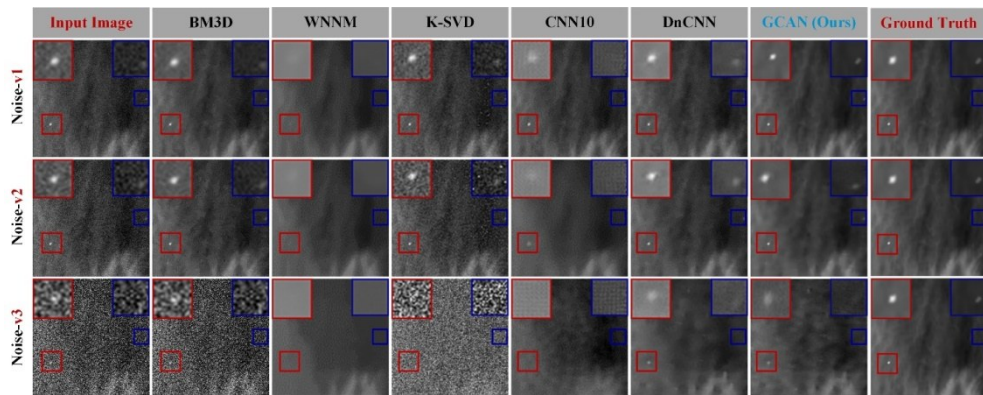


Fig. 4 Input image and corresponding denoising results of different methods on Noise-v1, Noise-v2 and Noise-v3 datasets respectively, for better visualization, the two regions of interest are enlarged and highlighted in red and blue, respectively, it can be observed that our GCAN can preserve target regions better

图4 不同去噪方法在三种噪声数据集上的输入图像和对应的去噪结果,图像中的重要区域采用红蓝框的形式进行了强调,可视化结果表明本文提出的去噪方法可以在有效去除噪声的同时保留好红外图像中的弱小目标信息

Table 4 IoU($\times 10^{-3}$), Pd($\times 10^{-3}$) and Fa($\times 10^{-4}$) values achieved by detection methods after pre-processing with noise reduction methods under varied noise-level dataset

表4 不同去噪方法去噪后,采用不同检测方法在三个数据集上交并比,检测率和虚警率的结果比较

Denoising Method	Noise. v1		Noise. v2		Noise. v3	
	Top-Hat ^[17]	DNANet ^[9]	Top-Hat ^[17]	DNANet ^[9]	Top-Hat ^[17]	DNANet ^[9]
BM3D ^[10]	23. 6/37. 5/1. 9	61. 1/72. 1/17. 7	13. 2/27. 4/3. 04	39. 4/49. 3/32. 9	5. 42/21. 3/128	5. 25/30. 8/18. 0
WNNM ^[13]	1. 89/6. 55/14. 5	1. 75/1. 58/1. 13	2. 11/7. 07/21. 55	2. 07/1. 90/0. 95	1. 13/3. 91/7. 82	0. 75/0. 63/0. 70
K-SVD ^[11]	21. 1/26. 3/12. 3	58. 9/67. 3/28. 1	13. 3/26. 2/45. 1	42. 1/51. 2/52. 0	5. 14/18. 5/86. 7	2. 12/32. 5/29. 1
RED-CNN ^[16]	13. 2/26. 9/39. 4	44. 5/58. 1/1. 91	5. 33/14. 8/3. 25	28. 1/28. 8/3. 92	1. 67/6. 61/3. 76	3. 57/10. 2/10. 0
DnCNN ^[15]	23. 9/39. 4/2. 05	72. 9/95. 1/1. 21	21. 1/35. 4/1. 96	60. 4/86. 2/1. 30	6. 29/18. 3/2. 75	15. 2/26. 2/5. 43
GCAN(ours)	24. 1/41. 7/1. 48	74. 5/96. 7/0. 40	22. 0/38. 4/1. 70	61. 6/87. 9/1. 00	8. 38/20. 2/2. 61	17. 5/29. 2/1. 07

Table 5 IoU($\times 10^{-2}$) values achieved by different detection methods under varied noise-level dataset

表5 不同检测方法在无噪和三种噪声等级数据集上的交并比性能

Detection Method	*Oriset	Noise. v1	Noise. v2	Noise. v3
Top-Hat ^[17]	25.8	23.6	13.0	5.21
RIPIT ^[18]	35.2	26.3	14.9	7.75
ACM ^[19]	44.1	39.1	20.7	1.19
UNet ^[20]	79.5	64.7	38.4	19.0
DNANet ^[9]	88.6	64.6	38.3	5.5

*Oriset indicates noise-free images in the image pairs.

Table 6 GFLOPs, Inference Time (s), Parameters, and PSNR performance of different denoising methods

表6 不同深度学习浮点运算量、推理时间、模型参数量和去噪性能比较

Denosing Method	GFLOPs (G)	Inference Time (s)	Params (M)	PSNR (dB)
RED-CNN ^[16]	83.89	0.156	1.848	44.6
DnCNN ^[15]	43.79	0.307	0.668	44.3
SWINIR ^[21]	49.64	0.271	11.80	44.9
GCAN	157.30	0.206	2.345	45.5

FLOPs). It may introduce inference delay under computational resources limited scenes, but is still affordable for the GPU-available scene.

4 Conclusion

In this paper, we propose a simple yet effective gradient-aware channel attention network (GCAN) for infrared small target image denoising before detection. To achieve this data-driven learning manner, we develop an infrared image denoising dataset, which contains 3 noise-level subsets. Then, we propose a novel infrared image denoising method (namely, GCAN) to achieve high-performance image denoising. Specifically, an encoder-decoder-based denoising network is used to initially remove the additive noise. Then, a residual connection structure and a gradient-based channel attention module (GCAM) are designed to preserve informative image details in IR images. Some conclusions can be summarized as follows:

(1) Compared to four benchmark denoising methods, GCAN achieves better denoising performance in terms of PSNR and SSIM. Better visually denoising performance is also achieved.

(2) The gradient-based channel attention module (GCAM) can avoid the over-smooth of IR images and effectively maintain the response of small target regions. Extensive experiments on five benchmark detection methods can verify the effectiveness of our method in terms of IoU, Pd and Fa.

(3) Although achieving better performance, larger model size and extra computation cost (i. e., FLOPs) are introduced, more light-weight computation operator

and simple network will be explored to increase the practicality under computational resources limited device in the future work.

References

- [1] Sun Y, Yang J, An W. Infrared dim and small target detection via multiple subspace learning and spatial-temporal patch-tensor model [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, **59** (5): 3737–3752.
- [2] Wu T, Li B, Luo Y. MTU-Net: Multilevel TransUNet for Space-Based Infrared Tiny Ship Detection [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, **61**(2): 1–15.
- [3] Li B, Wang Y, Wang L, et al. Monte Carlo Linear Clustering with Single-Point Supervision is Enough for Infrared Small Target Detection [C]// IEEE International Conference on Computer Vision. 2023, 455–468.
- [4] Liu T, Yang J, Li B, et al. Nonconvex tensor low-rank approximation for infrared small target detection [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, **60**(3): 1–18.
- [5] Li B, Guo Y, Yang J, et al. Gated recurrent multi-attention network for VHR remote sensing image classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, **60**(3): 1–13.
- [6] Liu T, Yang J, Li B, et al. Infrared small target detection via non-convex tensor Tucker decomposition with factor prior [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, **62**(4): 25–38.
- [7] Zhou J, Wang L, and Liu B. Analysis of the causes of non-uniformity in infrared images [J]. *Infrared and Laser Engineering*, 1997, **26** (3): 11–13.
- [8] Goyal. B, Dogra. A, Agrawal. S, et al. A. Image enoising review: From classical to state-of-the-art approaches [J]. *Information fusion*, 2020, **55**(1): 220–244.
- [9] Li B, Xiao C, Wang L, et al. Dense nested attention network for infrared small target detection [J]. *IEEE Transactions on image processing*, 2023, **32**(5): 1745–1758.
- [10] Dabov K, Foi A, Katkovnik V, et al. Image denoising by sparse 3-d transform-domain collaborative filtering [J]. *IEEE Transactions on image processing*, 2007, **16**(8): 2080–2095.
- [11] Aharon M, Elad M, Bruckstein A. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation [J]. *IEEE Transactions on signal processing*, 2006, **54**(11): 4311–4322.
- [12] Bouboulis P, Slavakis K., Theodoridis S. Adaptive kernel based image denoising employing semi-parametric regularization [J]. *IEEE Transactions on Image Processing*, 2010, **19**(6): 1465–1479.
- [13] Gu S, Xie Q, Meng D, et al. Weighted nuclear norm minimization and its applications to low level vision [J]. *International journal of computer vision*, 2017, **121**(2): 183–208.
- [14] Jain V, Seung H S. Natural Image Denoising with Convolutional Networks [C]//International Conference on Neural Information Processing Systems. 2008, 455–468.
- [15] Zhang K, Zuo W, Chen Y, et al. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising [J]. *IEEE transactions on image processing*, 2017, **26**(7): 3142–3155.
- [16] Chen H, Zhang Y, Zhang W, et al. Low-dose ct via convolutional neural network [J]. *Biomedical optics express*. 2017, **8**(2), 679–694.
- [17] Rivest J.-F. and Fortin R. Detection of dim targets in digital infrared imagery by morphological image processing [J]. *Optical Engineering*. 1996, **35**(7): 1886–1893.
- [18] Chirdchoo N, Soh W S, Chua K C. Ript: A receiver-initiated reservation-based protocol for underwater acoustic networks [J]. *IEEE Journal on Selected Areas in Communications*. 2008, **26**(9): 1744–1753.
- [19] Dai Y, Wu Y, Zhou F, et al. Asymmetric contextual modulation for infrared small target detection [C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2021: 950–959.
- [20] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation [C]// International Conference on Medical image computing and computer-assisted intervention. 2015, 234–241.
- [21] Liang J, Cao J, Sun G, et al. Swinir: Image restoration using swin transformer [C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 1833–1844.