remote sensing

MDPI

Article

BCDA-Net: A Bottleneck-Free Channel Dual-Path Aggregation Network for Infrared Image Destriping

Lingzhi Chen [1,2], Feng Dong [1], Lingfeng Huang [1] and Yutian Fu [1,*]

[1] Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China; chenlingzhi23@mails.ucas.ac.cn (L.C.); dongfeng@mail.sitp.ac.cn (F.D.); huanglingfeng@mail.sitp.ac.cn (L.H.)
[2] University of Chinese Academy of Sciences, Beijing 101408, China
* Correspondence: yutianfu@mail.sitp.ac.cn

Highlights

What are the main findings?

- We developed BCDA-Net, a novel architecture that integrates a Dual-Path Channel Down-sampling (DCD) module and a Bottleneck-Free Aggregation Strategy (BFAS) to structurally block the propagation of low-resolution noise.
- By further coupling a Context-Guided Stripe Attention Block (CGSAB) with a Cascaded Dense Feature Aggregation (CDFA) module, the network achieves state-of-the-art visual quality and quantitative metrics on both simulated and real-world infrared datasets.

What are the implications of the main findings?

- The bottleneck-free design successfully overcomes the traditional trade-off between effective denoising and detail preservation by structurally blocking the propagation of low-resolution aliasing artifacts.
- This network provides a highly robust preprocessing framework for wide-swath high-resolution infrared remote sensing, significantly enhancing the reliability of subsequent downstream applications.

Abstract

The inherent non-uniformity of Infrared Focal Plane Arrays (IRFPA) inevitably results in stripe noise, which severely degrades image quality and hinders downstream applications. Existing deep learning methods often struggle to strike a balance between effective denoising and the preservation of fine thermal textures. To address this issue, we propose a Bottleneck-free Channel Dual-path Aggregation Network (BCDA-Net) based on a "Perception-Reconstruction" design principle. In the perception stage, the network jointly employs the Dual-Path Channel Down-sampling (DCD) module and the Context-Guided Stripe Attention Block (CGSAB). The DCD module utilizes a channel split strategy to simultaneously extract semantic features and preserve high-frequency textures, while the CGSAB performs global context modeling on these features to precisely perceive and locate global stripe noise patterns. In the reconstruction stage, we integrate the Cascaded Dense Feature Aggregation (CDFA) module with a Bottleneck-Free Aggregation Strategy (BFAS). The CDFA utilizes the perceived information to densely aggregate features and progressively reconstruct clean image details, whereas the BFAS structurally blocks the propagation of low-resolution noise during decoding, effectively mitigating aliasing artifacts induced by deep feature upsampling. Together, these components form a complete closed loop from accurate noise perception to high-fidelity reconstruction. Extensive experiments on public and real-world datasets demonstrate that BCDA-Net maximally preserves image details while removing non-uniform stripe noise. Both objective metrics and subjective visual quality outperform existing state-of-the-art methods.

Check for updates

Academic Editor: Qiangqiang Yuan

Received: 16 March 2026
Revised: 20 April 2026
Accepted: 23 April 2026
Published: 25 April 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: infrared image destriping; non-uniformity correction; detail preservation; deep learning; dual-path downsampling; bottleneck-free aggregation; contextual attention

1. Introduction

Infrared imaging systems play a pivotal role in numerous sectors, including military reconnaissance, ocean observation, weather forecasting, and agricultural monitoring [1–3], due to their all-weather and anti-interference capabilities. However, infrared focal plane arrays, known as IRFPAs, inevitably suffer from non-uniformity issues caused by limitations in manufacturing processes, material heterogeneity, and readout circuit design. Even under uniform radiation input, the output signals of different pixels exhibit significant deviation [4]. This deviation manifests as stripe noise in the captured images, which severely degrades image quality and obscures critical target information. The mathematical model of a degraded infrared image is typically expressed as:

$$Y = S + N \tag{1}$$

where Y represents the observed degraded stripy IR image, S denotes the clean infrared scene, and N represents the stripe noise. To remove this noise and restore high-quality images, early research primarily relied on calibration-based methods using blackbody reference sources [5]. Although these methods offer high precision, they require complex equipment and struggle to adapt to the drastic temperature changes encountered during in-orbit operations.

To overcome this reliance on hardware, scene-based correction methods emerged. These can be broadly categorized into spatial domain filtering, transform domain analysis, optimization-based variational methods, and statistical methods. Classical spatial domain approaches, such as guided filters [6,7], utilize edge-preserving smoothing properties to seek a balance between noise suppression and detail preservation. Transform domain methods employ tools like the discrete wavelet transform [8,9] or fourier transform [10,11] to decompose images and separate specific noise components in the frequency domain. Optimization-based approaches [12], such as total variation regularization [13,14] and methods based on the alternating direction method of multipliers [15,16], formulate destriping as an energy minimization problem by introducing structural priors, such as low-rank properties in the column gradient domain. Additionally, statistical methods based on grayscale statistics, such as improved median histogram equalization [17–19], avoid the need for reference sources. However, these traditional methods often exhibit limited generalization capabilities when processing real-world scenes with complex textures and dense targets, frequently leading to detail loss or ghosting artifacts.

In recent years, the rise of deep learning has led to Convolutional Neural Network (CNN) algorithms gradually outperforming and replacing traditional statistical and scene-based methods due to their powerful feature representation capabilities [20,21]. Early shallow networks and residual designs, such as DLS-NUC [22] and LSC-CNN [23], significantly improved noise removal by learning residual maps. Subsequent research introduced multi-scale dense connections [20] and attention mechanisms, such as CSCNet [24], CNN-FPNR [25], TS-NUC [26], SMNet [27] and SOCAM [28], to enhance column consistency and global uniformity. To effectively handle complex degradations, hybrid architectures have gained significant traction. For instance, S2OTAN [29] captures correlation-aware contextual information using hybrid attention within a Transformer block, while Translution-SNet [30] integrates CNNs with Transformers in a semi-supervised framework to robustly remove stripe noise. Specifically addressing the challenging mixture of fixed-pattern and

Gaussian noise, recent architectures like the Dual-Encoder UNet [31] leverage dual-scale residual attention encoders to independently perceive and fuse distinct noise characteristics for high-quality reconstruction. Furthermore, many methods have incorporated wavelet transforms into neural networks, such as ASCNet [32], TSWEU [33], SDNN [34] and SNR-WDNN [35], to achieve better spectral separation. Building upon this, recent studies have increasingly coupled wavelet transforms with advanced generative paradigms to leverage both structural priors and data-driven generalization. For instance, while DSCGAN [36] and MLFAN [37] achieve blind denoising through generative adversarial networks, the most recent KEDM [38] paradigm advances this trajectory by embedding wavelet-based stripe priors into a conditional diffusion model, further enhancing noise reduction and feature extraction capabilities. Recently, joint restoration frameworks have emerged to address compounded degradations in infrared images, such as combining destriping with deblurring to suppress artifacts and preserve edges [39,40]. While these unified approaches offer promising solutions for complex real-world scenarios, effectively disentangling stripe noise without compromising fine-grained high-frequency details remains a critical challenge. This highlights the necessity of architectures capable of seamlessly aggregating shallow information to ensure structural fidelity during the destriping process.

While recent general image restoration networks, such as the Adaptive Kernel Dilation Transformer (AKDT) [41], Restormer [42], the Hybrid Attention Transformer (HAT) [43], and CANS [44], achieve impressive performance in removing random noise, they are sub-optimal for infrared image destriping. Specifically, Transformer-based architectures like HAT and Restormer either disrupt the continuous spatial distribution of stripe noise through patch partitioning or over-smooth subtle thermal gradients. Furthermore, AKDT lacks the directional priors necessary to model the fixed-pattern noise (FPN) inherent in infrared sensors, while CANS suffers from severe information bottlenecks during deep feature extraction, leading to the irreversible loss of high-frequency thermal details. To address these limitations, our proposed BCDA-Net utilizes a Bottleneck-Free Aggregation Strategy (BFAS) that directly aggregates shallow information alongside deep features, effectively avoiding the degradation of fine-grained structural details.

In the task of infrared image destriping, fully exploiting and utilizing cross-scale contextual information is crucial for improving image reconstruction quality. Unlike the standard U-Net (as illustrated in Figure 1a), which is constrained by progressive layer-wise communication between adjacent levels, UNet3+ [45] (shown in Figure 1b) breaks the communication barrier between deep semantics and shallow details through its full-scale dense feature fusion mechanism. This provides a highly promising architectural foundation for high-quality image restoration.

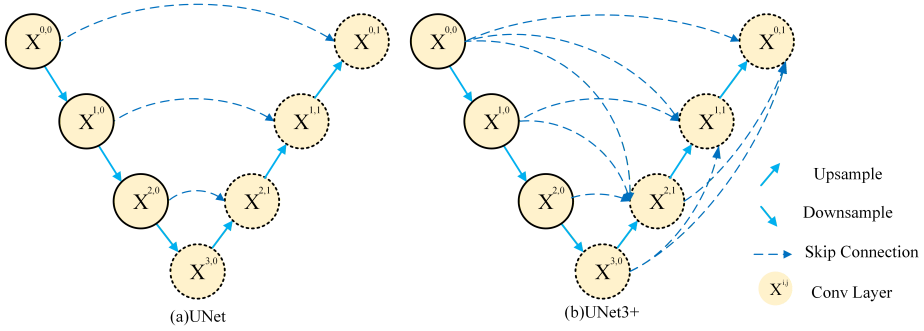


Figure 1. Structural comparison between (**a**) the standard U-Net with progressive layer-wise communication and (**b**) UNet3+ with full-scale dense feature fusion mechanism.

Meanwhile, existing mainstream advanced architectures also fail to perfectly reconcile feature interaction with physical priors. For example, although advanced convolutional

neural networks integrating wavelet transforms (e.g., ASCNet [32]) mitigate feature loss in the frequency domain, they still adhere to the layer-by-layer decoding mechanism of the standard U-Net, thereby restricting the direct flow of full-scale information. Furthermore, many methods for integrating the wavelet transform into networks rely on specific software libraries, making them difficult to deploy on certain edge devices. Unsupervised generative adversarial network-based methods (e.g., DSCGAN [36]) lack physical constraints tailored for non-uniformity degradation, making them susceptible to generating radiometric distortions or false artifacts in clean background regions. In addition, although recent general vision Transformer models (e.g., AKDT [41]) exhibit outstanding global modeling capabilities, their self-attention mechanisms are not only computationally expensive but also neglect the unique directional signatures of infrared stripes.

However, the degradation process of infrared focal plane arrays (FPAs) possesses strong directional. Directly applying the conventional dense fusion paradigm to the infrared destriping task exposes three critical architectural flaws: First, the large-scale over-upsampling of extremely low-resolution bottleneck layers is highly prone to inducing severe aliasing artifacts and edge blurring. Second, the indiscriminate hard concatenation of cross-scale features during the decoding stage fails to effectively filter out redundant information, easily leading to the re-introduction of shallow-layer noise. Third, conventional convolutions lack awareness of the characteristics of infrared stripes, making it difficult to achieve the precise decoupling of true thermal radiation and structured noise.

To address these challenges, this paper proposes a novel infrared destriping network based on a Dual-Path Channel Down-sampling and Bottleneck-Free Aggregation strategy, named BCDA-Net. First, we innovatively construct a Dual-Path Channel Down-sampling module, denoted as DCD module. By utilizing a channel splitting strategy to process texture preservation via a max-pooling path and complex semantic extraction via a convolutional path in parallel, this module achieves dual fidelity for both texture and semantics. Second, in the feature encoding–decoding stage, we integrate a Global Context-Guided Stripe Attention module, referred to as CGSAB, and a Cascaded Dense Feature Aggregation module, named CDFA. The CGSAB utilizes a global context branch to construct a noise mask for precisely locating and suppressing non-uniform stripes, while the CDFA reuses multi-level residual features to achieve coarse-to-fine detail reconstruction. Finally, to resolve the contradiction in multi-scale fusion, we design a Bottleneck-Free Aggregation Strategy, or BFAS. This strategy blocks the direct connection between low-resolution features from the bottleneck layer and the initial decoding stage, effectively suppressing the propagation of noise caused by excessive upsampling.

The main contributions of this paper are as follows:

1. We propose the DCD module to resolve the conflict between feature loss and parameter redundancy.
2. We design the BFAS to eliminate aliasing artifacts caused by the upsampling of deep features.
3. We construct a "Perception-Reconstruction" denoising network. By combining the global perception capability of CGSAB with the dense reconstruction ability of CDFA, the proposed network effectively removes non-uniform stripes of varying intensities. It maximally restores the original temperature distribution and detailed textures of infrared images, outperforming existing mainstream methods in both subjective and objective evaluations.

2. Methods

In Section 2.1, we first briefly introduce the overall network architecture of BCDA-Net. Subsequently, Section 2.2 details the DCD module within the network framework. The

BFAS are introduced in Section 2.3. The specifics of CGSAB and CDFA are elaborated in Sections 2.4 and 2.5, respectively. In Section 2.6, we describe the feature fusion module composed of multiple HFAN blocks.

2.1. Overall Network Architecture

The overall architecture of BCDA-Net is illustrated in Figure 2. Upon receiving a noisy input image I_{in}, the network first employs a convolutional layer (Double Conv) to extract the initial shallow feature $E0$. Then, the network systematically divides the restoration process into two interactive stages: the Perception Stage (Encoding) and the Reconstruction Stage (Decoding). The mapping from the input image to the target output feature map is learned through an asymmetric, full-scale connected U-shaped structure.

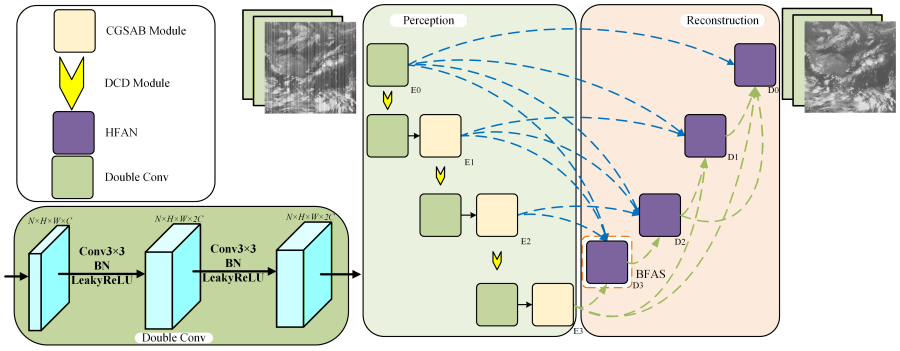


Figure 2. Overall network architecture of the proposed BCDA-Net, driven by a unified "Perception-Reconstruction" paradigm. The dashed blue arrows represent full-scale skip connections transferring multi-scale features from the perception stage and the dashed green arrows indicate bottom-up feature aggregation within the reconstruction stage. In the perception stage, the Dual-Path Channel Down-sampling (DCD) module and Context-Guided Stripe Attention Block (CGSAB) jointly decouple stripe noise from thermal textures. In the reconstruction stage, the Hierarchical Feature Aggregation Network (HFAN) progressively restores the image utilizing full-scale skip connections. Notably, the initial HFAN employs a Bottleneck-Free Aggregation Strategy (BFAS) to block noise propagation at the source, while subsequent HFANs utilize the Cascaded Dense Feature Aggregation (CDFA) module to calibrate upsampled deep features.

(1) The Perception Stage (Encoding): The primary goal of this stage is to accurately decouple noise patterns from clean semantic structures. Instead of relying on traditional max-pooling, the initial feature $E0$ is sequentially processed by three encoders utilizing the Dual-Path Channel Down-sampling (DCD) module and the Context-Guided Stripe Attention Block (CGSAB) to yield $E1$, $E2$, and $E3$. By splitting the channel dimensions, DCD simultaneously retains high-frequency thermal textures and extracts robust semantic representations while reducing feature dimensions. However, local features alone are insufficient for handling non-uniform stripes. Therefore, the Context-Guided Stripe Attention Block (CGSAB) is highly integrated to perform deep feature extraction and adaptive enhancement. Acting as a global observer, the CGSAB leverages the structured features from DCD to model long-range dependencies, successfully localizing and perceiving global stripe noise patterns across the spatial domain.

(2) The Reconstruction Stage (Decoding): Once the noise is perceived and decoupled, the reconstruction stage aims to progressively restore the high-fidelity image. To leverage both shallow details and deep semantics simultaneously, a full-scale skip connection strategy is adopted to fuse encoder and decoder features. The decoding process is driven by the Hierarchical Feature Aggregation Network (HFAN). Notably, the internal mechanism of HFAN adapts to its specific decoding stage, strategically governed by the Bottleneck-Free Aggregation Strategy (BFAS) and the Cascaded Dense Feature Aggregation (CDFA) module.

Specifically, in the initial decoding stage (the first HFAN), the network fuses the deepest encoded features with the shallow information without involving upsampling. Here, we implement the BFAS to operate as a structural gatekeeper. By establishing direct pathways without compression bottlenecks, BFAS structurally blocks the propagation of low-resolution noise aliases right at the source of reconstruction. Conversely, the subsequent HFAN layers additionally receive deep decoded features upsampled via bilinear interpolation. Since deep feature upsampling often inadvertently amplifies residual noise and introduces spatial misalignment, the CDFA module is employed immediately after the upsampling operation. CDFA utilizes the globally perceived context to densely weave the multi-scale features together, refining and calibrating the upsampled deep features. Subsequently, within each HFAN, the multi-source features (including the shallow information) are concatenated along the channel dimension and fused using a 3×3 convolution, followed by batch normalization (BN) and a ReLU activation function, thereby generating decoder features of specific dimensions.

Finally, through this unified "Perception-Reconstruction" closed loop, the fused decoder feature $D0$ with the highest resolution is passed through a final 3×3 convolutional layer and a Sigmoid activation function to produce the high-fidelity prediction map I_{out}.

2.2. Dual-Path Channel Downsampling (DCD) Module for Stripy Features in a UNet3+-Shaped Framework

To resolve the conflict between parameter redundancy and the loss of detailed features in traditional downsampling, while maximizing texture detail preservation during feature map size reduction, we designed a hybrid channel-split downsampling module based on a dual-path principle. Diverging from conventional methods that apply pooling or convolution uniformly to all input channels, this module introduces a Channel Split strategy [46,47]. The architecture of DCD module is illustrated in Figure 3. It is formulated as

$$h = f_{\mathrm{BN+ReLU}}(\mathrm{Concat}[\mathrm{Conv}_{3\times3}(X_{1:C/2}), \mathrm{Conv}_{1\times1}(\mathrm{MaxPool}(X_{C/2+1:C}))]) \quad (2)$$

where h denotes the output of the DCD module. $X_{1:C/2}$ and $X_{C/2+1:C}$ represent the first and second halves of the input channels, respectively. $\mathrm{MaxPool}(\cdot)$ denotes the max-pooling operation utilized for extracting salient features. $\mathrm{Conv}_{n\times n}(\cdot)$ represents a convolution operation with an $n \times n$ kernel and a stride of 1. Finally, $\mathrm{Concat}[\cdot]$ denotes the concatenation operation along the channel dimension, and $f_{\mathrm{BN+ReLU}}(\cdot)$ represents the sequential computations of Batch Normalization followed by the ReLU activation function.

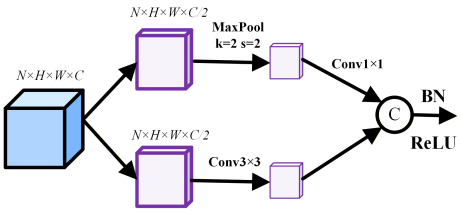


Figure 3. Structure of the Dual-Path Channel Downsampling(DCD) module.

Given an input feature $X \in \mathbb{R}^{C\times H\times W}$, the module first splits X along the channel dimension into two sub-feature maps, X_{pool} and X_{conv}, each with dimensions $\frac{C}{2} \times H \times W$. Salient Feature Path: X_{pool} passes through a Max-Pooling layer, aiming to preserve salient features and achieve translation invariance without introducing additional parameters. Then, the features are extracted through the $\mathrm{Conv}_{1\times1}$ operation. Extraction Path: X_{conv} passes through a Strided Convolution layer, which is responsible for reducing the feature map size while extracting semantic features and preserving texture structure.

Finally, the output features from the two paths are concatenated along the channel axis, restoring the total channel count to C. After batch normalization and non-linear activation, the final output $h \in \mathbb{R}^{C \times \frac{H}{2} \times \frac{W}{2}}$ is obtained.

2.3. Bottleneck-Free Aggregation Strategy (BFAS)

The standard UNet3+ architecture fuses features from all scales, including the deepest bottleneck layer, indiscriminately into the decoder via full-scale skip connections [45]. However, in the first decoding stage, the bottleneck features possess an extremely low spatial resolution and require substantial upsampling to align with the decoder layer dimensions. For infrared image destriping, this excessive and cascaded upsampling inevitably introduces interpolation artifacts, blurs fine spatial boundaries, and can even distort the structural patterns of stripe noise, thereby negating the benefits of deep semantic information.

To address this issue, we propose the Bottleneck-Free Aggregation Strategy (BFAS). This strategy prunes the upsampling path from the deepest feature bottleneck layer in the first decoder stage, integrating instead multi-scale features from each encoder layer. This strategy effectively prevents the cumulative errors caused by layer-by-layer upsampling, eliminates interference from interpolation artifacts, and provides a robust, artifact-free foundation for the subsequent decoding stages, ultimately achieving clearer edge preservation and enhanced image restoration capability.

2.4. Context-Guided Stripe Attention Block (CGSAB)

The architecture of CGSAB module is illustrated in Figure 4. Due to the stripe noise caused by the non-uniform response of the infrared focal plane array (FPA), which is characterized by strong directionality and non-uniform spatial distribution, a Context-Guided Stripe Attention Block (CGSAB) is designed. The CGSAB contains two branches: the Stripe noise and Local Thermal Detail Branch (SLD Branch) and the Global Context Branch (GC Branch). Inspired by the effective self-calibrated dual attention mechanism blocks for image deraining [48] and EMA [49], given an input feature $X \in \mathbb{R}^{C \times H \times W}$, our CGSAB process is defined as

$$\text{Yout} = \text{ReLU}(SLD(fX) \otimes GC(fX) + X) \tag{3}$$

where Yout denotes the output of CGSAB. $f(\cdot)$ denotes a convolution with a kernel size of 3×3 and a stride of one, followed by a ReLU activation function. $SLD(\cdot)$ denotes the output of the Stripe noise and Local Thermal Detail Branch, and $GC(\cdot)$ denotes the output of the Global Context Branch. $\text{ReLU}(\cdot)$ denotes the ReLU activation function.

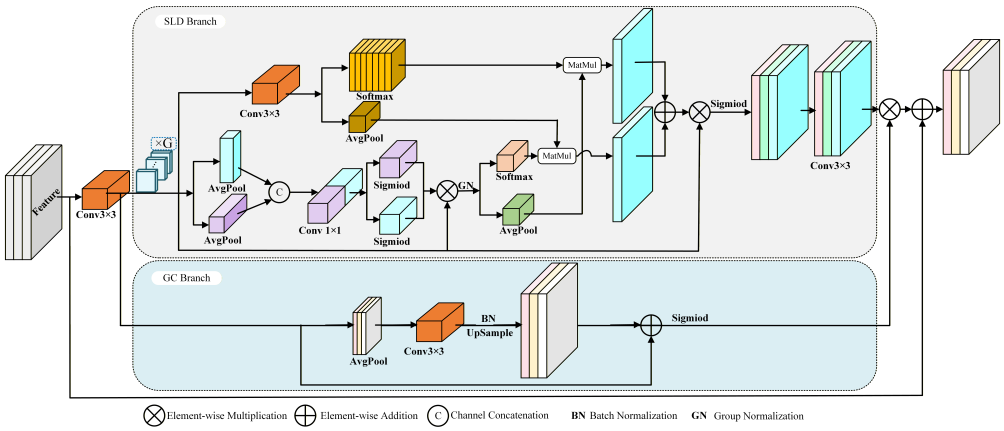


Figure 4. Structure of the Context-Guided Stripe Attention Block (CGSAB). The core modules of CGSAB are SLD Branch, which extracts stripe noise and local thermal detail, and GC Branch, which extracts global semantic context.

2.4.1. Stripe Noise and Local Detail Branch

Directional stripe prior encoding is first performed to capture the global distribution of non-uniformity. Since stripe noise exhibits strong directionality, we utilize 1D global average pooling to decompose the input features into vertical and horizontal encodings. Column-wise pooling (z^{col}) aggregates vertical information to model column FPN bias, while row-wise pooling (z^{row}) captures the consistency of the horizontal context:

$$\mathbf{z}^{col} = \mathrm{AvgPoolH}(X) \tag{4}$$

$$\mathbf{z}^{row} = \mathrm{AvgPoolW}(X) \tag{5}$$

where $\mathrm{AvgPoolH}(\cdot)$ and $\mathrm{AvgPoolW}(\cdot)$ denote the average pooling operations along the height and width dimensions, respectively.

These directional encodings are subsequently fused and transformed to generate the global stripe prior feature F_{stripe}:

$$F_{stripe} = \mathrm{GroupNorm}\Big(X \otimes \sigma(\phi(\mathbf{z}^{col})) \otimes \sigma(\phi(\mathbf{z}^{row}))\Big) \tag{6}$$

where F_{stripe} denotes the global stripe prior feature, and X represents the input feature map. The vectors $\mathbf{z}^{col}$ and $\mathbf{z}^{row}$ represent the column and row directional encodings, respectively. The function $\phi(\cdot)$ denotes a shared 1×1 convolution operation utilized for channel interaction. Specifically, $\mathbf{z}^{col}$ and $\mathbf{z}^{row}$ are concatenated and passed through this shared convolutional layer before being split. The function $\sigma(\cdot)$ represents the Sigmoid activation, and $\otimes$ indicates element-wise multiplication. $\mathrm{GroupNorm}(\cdot)$ denotes the Group Normalization operation. From a physical perspective, this entire step corresponds to a learnable 1D spatial correction process.

To prevent the smoothing of genuine structural details (such as edges of buildings or trees) during the destriping process, a convolution operation is employed to extract local feature variations, thereby preserving high-frequency texture information. The local detail feature, denoted as F_{local}, is mathematically formulated as

$$F_{local} = \mathrm{Conv}_{3\times3}(X) \tag{7}$$

where $\mathrm{Conv}_{3\times3}(\cdot)$ denotes a convolution with a kernel size of 3×3 and a stride of one.

Finally, a cross-spatial attention mechanism is used to enable mutual guidance between the global stripe context and local details. The final corrected feature is calculated as follows:

$$SLD = \mathrm{Conv}_{3\times3}\big(X \otimes \sigma\big(Softmax(F_{stripe})AvgP(F_{local}) + Softmax(F_{local})AvgP(F_{stripe})\big)\big) \tag{8}$$

where $\sigma(\cdot)$ represents the Sigmoid activation function. $Softmax(\cdot)$ and $AvgP(\cdot)$ represent the Softmax activation and average pooling operations, respectively. Note that the juxtaposition of the Softmax and AvgP terms (e.g., $Softmax(F_{stripe})AvgP(F_{local})$) explicitly denotes the matrix multiplication between them.

2.4.2. Global Context Branch

The stripe noise in infrared images is often superimposed on the low-frequency background thermal radiation. This branch utilizes an average pooling downsampling branch to construct a large receptive field for estimating the smooth global thermal distribution, assisting the network in distinguishing between high-frequency stripe noise and low-frequency background variations. Then, by fusing the global context features with local features and passing through a Sigmoid activation, an adaptive noise mask is generated. This mask acts as a soft gating signal to weight the features of the main branch, thereby suppressing

stripe interference while preventing the erroneous smoothing of genuine edges caused by temperature differences. It is expressed as

$$GC = \sigma\left(X + \mathrm{Up}\left(\mathrm{Conv}_{3\times3}\left(\mathrm{AvgPool}_{4\times4}(X)\right)\right)\right) \tag{9}$$

where $\mathrm{AvgPool}_{4\times4}(\cdot)$ is the mean filter of size 4×4. Here, we use average pooling for downsampling to increase the perceptual range, pooling spatial information within the kernel into a single pixel. $\mathrm{Up}(\cdot)$ is the bilinear interpolation operator.

2.5. Cascaded Dense Feature Aggregation (CDFA) Module

As shown in Figure 5, the architecture adopts a cascaded densely connected framework, composed of two cascaded CGSAB (Context-Guided Stripe Attention Block) units. It is designed to meticulously eliminate infrared strip noise while preserving details and reducing artifacts. Infrared images often suffer from low contrast and blurred edges; simple stacking of convolutional layers can easily lead to the smoothing out of small targets and textural details during the denoising process.

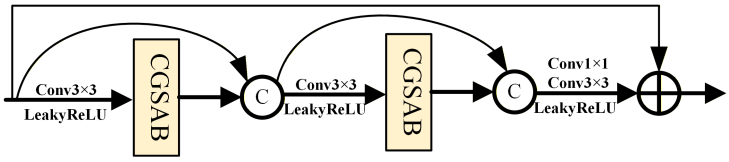


Figure 5. Structure of the Cascaded Dense Feature Aggregation (CDFA) module.

This module utilizes two cascaded CGSAB units for phased processing. The first stage removes prominent strong stripes, and the second stage leverages information from the previous stage to further eliminate faint residual noise, achieving a coarse-to-fine restoration. The strategy of Dense Connection is employed, directly concatenating the original residual features—rich in background information—with deeper features. This mechanism ensures that the thermal radiation distribution and textural details of the original image are transmitted losslessly to the deeper layers of the fusion module network, preventing feature blurring caused by multiple convolutions.

Finally, multi-branch features are fused using a 1×1 convolution, and a global residual connection is introduced. This primarily allows the module to learn a "pure noise map" rather than predicting the entire image, significantly reducing the optimization difficulty and enhancing the fidelity of the denoising results.

2.6. Hierarchical Feature Aggregation Network (HFAN)

To preserve delicate structural details during multi-scale fusion, our enhanced HFAN builds upon the full-scale skip connections of UNet3+ [45]. While standard full-scale fusion directly upsamples deep features, this introduces severe interpolation artifacts in infrared image destriping. To address this, we modify the baseline decoding process by explicitly integrating the previously introduced CDFA module immediately after the upsampling operation. The CDFA acts as a dedicated feature refiner to recover contextual information, repair interpolation artifacts, and perform fine-grained noise removal prior to multi-scale concatenation.

To illustrate the internal architecture, we take the generation of decoder node D_1 (Figure 6) as an example. This node aggregates four multi-scale branches: the downsampled shallow encoder feature (E_0), the same-scale encoder feature (E_1), the upsampled feature from the adjacent decoder stage (D_2), and the upsampled feature from the deepest bottleneck stage (D_3). Specifically, the features from deeper stages (D_2 and D_3) undergo 1-step bilinear upsampling $\mathrm{Up}(\cdot)$ and are immediately processed by independent CDFA

modules for artifact repair and fine-grained residual noise removal. After these respective spatial transformations, a 3×3 convolution ($\text{Conv}_{3\times3,16}$) is applied to each of the four branches to uniformly adjust their channel dimensions to 16. Finally, all aligned branches are concatenated and fused via a 3×3 convolution $\text{Conv}_{3\times3}$, Batch Normalization BN, and ReLU activation. The complete process is formulated as

$$\begin{aligned} D_1 = \text{ReLU}\Big(\text{BN}\Big(\text{Conv}_{3\times3}\Big(\Big[&\text{Conv}_{3\times3,16}(\text{Maxpool}(E_0)),\\ &\text{Conv}_{3\times3,16}(E_1),\\ &\text{CDFA}(\text{Conv}_{3\times3,16}(\text{Up}(D_2))),\\ &\text{CDFA}(\text{Conv}_{3\times3,16}(\text{Up}(E_3)))\Big]\Big)\Big)\Big) \end{aligned} \tag{10}$$

where $[\cdot]$ denotes channel concatenation, $\text{Maxpool}(\cdot)$ represents max pooling operations, and $\text{Conv}_{3\times3,16}(\cdot)$ adjusts the feature map to 16 channels.

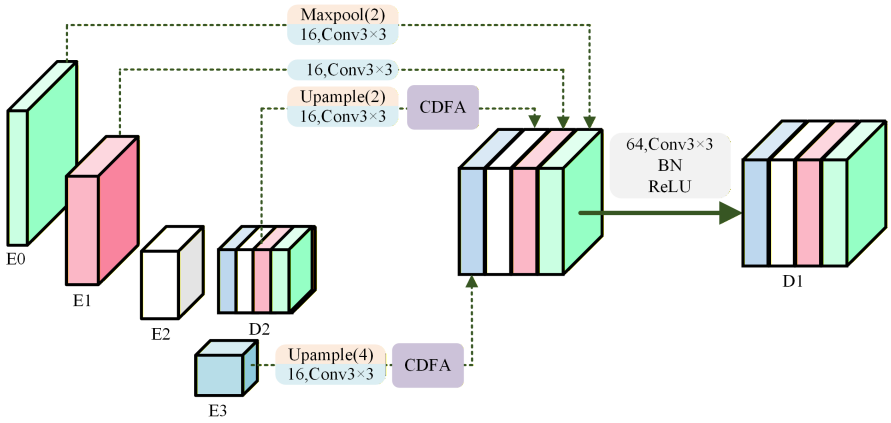


Figure 6. Structure of the Hierarchical Feature Aggregation Network (HFAN) for the decoder node D_1.The different colors of the feature blocks represent distinct hierarchical scales within the network: green, pink, white, and blue correspond to scales 0, 1, 2, and 3, respectively. In the concatenated tensor, these colors explicitly illustrate the comprehensive aggregation of multi-scale features derived from both the encoder nodes ($E0$, $E1$, $E3$)and the deeper decoder node ($D2$).

3. Experiments and Result

3.1. Configuration of Experiment

3.1.1. Construction of Dataset

To cover a wide range of scenarios, the experimental data used in this paper is composed of the following two existing infrared (IR) datasets.

- ICSRN [50]: This dataset contains 215 images of infrared cloud layers, with each image having a size of 579×572 pixels. From ICSRN, we selected 60 images for training, 20 images for validation, and used 20 images for testing.
- CVC09 [51]: This dataset consists of a large number of far-infrared (FIR) pedestrian images, each with a size of 640×480 pixels. We chose 200 images showcasing different scenes from CVC09 for training, reserved 50 images for validation, and used 45 images for testing.

We applied data augmentation techniques to the training data, including rotation, scaling, and flipping operations. A sliding window method with a window size of 64×64 and a stride of 40×40 was used to crop the augmented data. This process generated 181,752 image patches, which were then packaged into HDF5 format files for convenient

use during training [20,32]. Consequently, the dataset used in this paper comprises a total of 181,752 training image patches, 70 validation images, and 65 testing images.

3.1.2. Noise Setting

Previous studies have demonstrated that the cubic degradation model possesses strong generalization capabilities [20,22,32]. Therefore, we utilize cubic noise to train the model for stripe noise. Specifically, different cubic noise patterns are randomly generated and added dynamically in each training epoch, with the noise level for each order set to a mean of 0 and a random standard deviation ranging from 0 to 0.15. Considering the complexity of stripe noise, we test the model's performance on three typical stripe noise distributions—Gaussian, uniform, and periodical—to evaluate its robustness.

3.1.3. Loss Function

In this study, the Mean Square Error (MSE) loss function is adopted during the training process. For N pairs of images, this loss function can be expressed as follows:

$$L = \frac{1}{N}\sum_{i=1}^{N} \| I_i^O - I_i^C \|^2 \tag{11}$$

where I_i^C represents the i-th Ground Truth (GT) image, and I_i^O represents the corresponding i-th output image generated by the network.

3.1.4. Evaluation Metrics

Following previous infrared image destriping works [32], we calculate the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) to evaluate the destriping performance. Higher PSNR and SSIM values indicate a higher similarity between the destriped image and the clean image. The calculation methods are as follows:

$$PSNR = 20\log_{10}\left(\frac{MAX}{\sqrt{MSE}}\right) \tag{12}$$

where MAX represents the maximum possible pixel value of the image (e.g., 255 for an 8-bit image). MSE denotes the Mean Squared Error between the denoised image and the original ground truth.

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{13}$$

where x and y represent the ground truth image and the processed image, respectively. The parameters μ_x and μ_y denote the average pixel values (mean) of x and y, which estimate the luminance. σ_x^2 and σ_y^2 represent the variances of x and y, reflecting the image contrast. σ_{xy} denotes the covariance between x and y, measuring the structural correlation. Finally, C_1 and C_2 are small positive constants introduced to stabilize the division and prevent instability when the denominator approaches zero.

3.1.5. Implementation Details

The model optimization employs the Adam optimizer with an initial learning rate of 0.001. The training schedule follows a cosine annealing strategy over 430 epochs, which includes a 4-epoch warm-up phase. In the final stage, a fixed minimum learning rate of 5×10^{-5} is applied for 20 epochs. The total number of training epochs is set to 450, with a batch size of 256.

To enhance training stability and model generalization, orthogonal regularization [52] is applied to the convolutional kernels, encouraging orthogonality among filters to mitigate

issues such as gradient explosion or vanishing gradients. All experiments are conducted on a single NVIDIA GeForce 4090 GPU.

3.2. Simulated Image Destriping

3.2.1. Quantitative Results

We added the following three types of noise to the test set for validation:

- Gaussian: Stripe noise following a Gaussian distribution with noise level $\sigma \in [0, 0.15]$.
- Uniform: Stripe noise following a uniform distribution with noise level $\mu \in [-0.15, 0.15]$.
- Periodical: Randomly selected noise period $T \in \{6, 7, 8, 9\}$, with Gaussian-distributed stripe noise of level $\sigma \in [0, 0.15]$ added to period T.

To comprehensively evaluate our method, we compared it against 10 state-of-the-art approaches: GF [6], MIRE [19], WLS [53], ADOM [16], DLSNUC [22], SNRWDNN [35],CANS [44], AKDT [41], DSCGAN [36] and ASCNet [32]. Quantitative results are shown in the Tables 1 and 2. The best and second-best results are highlighted in bold and underlined, respectively. Evidently, our method outperforms state-of-the-art approaches.

As can be observed, the SSIM scores of DSCGAN, ASCNet, and the proposed BCDA-Net are highly comparable. This is primarily because the SSIM metric focuses more on evaluating the global structural similarity of an image. Since the spatial footprint of local stripe residuals is extremely small relative to the entire image, the globally computed score is easily dominated by the extensive, well-preserved background information. Consequently, high-performance networks tend to reach a saturation point in such global evaluation metrics. Nevertheless, a more detailed qualitative analysis in Section 3.2.2 will reveal significant differences in actual visual quality, demonstrating that our BCDA-Net holds a distinct advantage in the preservation and restoration of detailed information.

Table 1. Quantitative comparison (average PSNR/SSIM) of various destriping methods on the ICSRN dataset with simulated stripes. The best and second-best results are highlighted in **bold** and underlined, respectively. The upward arrow (↑) indicates that larger values correspond to better performance.

Type	Metric	Method											
		Degraded	GF	MIRE	WLS	ADOM	DLSNUC	SNRWDNN	CANS	AKDT	DSCGAN	ASCNet	BCDANet
Gaussian	PSNR ↑	22.3727	32.4095	18.3383	34.8442	33.6641	30.3831	34.1654	34.2162	37.9361	38.4576	39.0540	**39.3683**
	SSIM ↑	0.4754	0.9440	0.8263	0.9878	0.9862	0.9131	0.9785	0.9728	0.9889	0.9919	0.9936	**0.9942**
Uniform	PSNR ↑	31.8086	37.6903	17.7048	38.1673	36.0841	37.3339	39.5754	37.0106	38.7565	42.0705	41.0644	**42.0989**
	SSIM ↑	0.7442	0.9856	0.8107	0.9936	0.9878	0.9767	0.9884	0.9813	0.9923	0.9945	0.9952	**0.9953**
Periodical	PSNR ↑	24.1225	33.5137	17.6650	32.7782	29.6255	33.1818	33.3066	32.4061	33.1875	34.2659	34.7807	**37.5324**
	SSIM ↑	0.5304	0.9591	0.8115	0.9902	0.9813	0.9335	0.9828	0.9772	0.9879	0.9895	0.9919	**0.9934**

Table 2. Quantitative comparison (average PSNR/SSIM) of various destriping methods on the CVC09 dataset with simulated stripes. The best and second-best results are highlighted in **bold** and underlined, respectively. The upward arrow (↑) indicates that larger values correspond to better performance.

Type	Metric	Method											
		Degraded	GF	MIRE	WLS	ADOM	DLSNUC	SNRWDNN	CANS	AKDT	DSCGAN	ASCNet	BCDANet
Gaussian	PSNR ↑	22.8465	32.6036	23.6437	36.4861	32.8321	31.9411	35.7632	35.7599	37.4052	38.4771	39.0671	**39.2168**
	SSIM ↑	0.3146	0.9006	0.8953	0.9759	0.9742	0.8898	0.9638	0.9650	0.9808	0.9836	0.9856	**0.9870**
Uniform	PSNR ↑	31.2174	38.5301	21.5944	41.2431	35.3057	40.7098	40.5277	39.0351	40.0565	**42.0728**	40.8392	42.0723
	SSIM ↑	0.6099	0.9752	0.8997	0.9869	0.9774	0.9748	0.9784	0.9758	0.9830	0.9886	0.9882	**0.9889**
Periodical	PSNR ↑	26.2941	35.8377	22.5588	36.0662	30.6240	36.3454	35.7654	35.6437	36.0636	36.8413	36.7807	**38.6330**
	SSIM ↑	0.4080	0.9537	0.9158	0.9648	0.9723	0.9510	0.9729	0.9727	0.9782	0.9829	0.9865	**0.9877**

3.2.2. Visual Performance

The qualitative destriping results are shown in Figures 7–9. The three groups of test images correspond to Gaussian, uniform, and periodical noise, respectively. The

effectiveness of the network is evident, as it consistently delivers excellent results across all stripe scenarios.

Further local visual analysis demonstrates intrinsic differences between the networks in terms of detail fidelity and denoising capability. The network corresponding to Figure 7k tends to perform excessive smoothing on the image; although this strategy yields better results in global objective metrics, it severely suppresses faint details and exhibits limited effectiveness in handling high-intensity stripe noise. Specifically, a close inspection of Figure 7k reveals that the network incorrectly eliminates genuine thermal radiation information generated by tire friction as noise, while simultaneously leaving a visible faint stripe noise residual. This indicates that the method fails to effectively distinguish between authentic high-frequency details and stripe noise. In contrast, our proposed method (Figure 7l) not only thoroughly removes stripe noise of various intensities and restores the original luminance distribution but also precisely and clearly preserves key physical thermal features, such as tire thermal tracks. In Figure 8, our method successfully recovers the image details around the tree trunk edges, while other methods introduce artifacts and blurring.

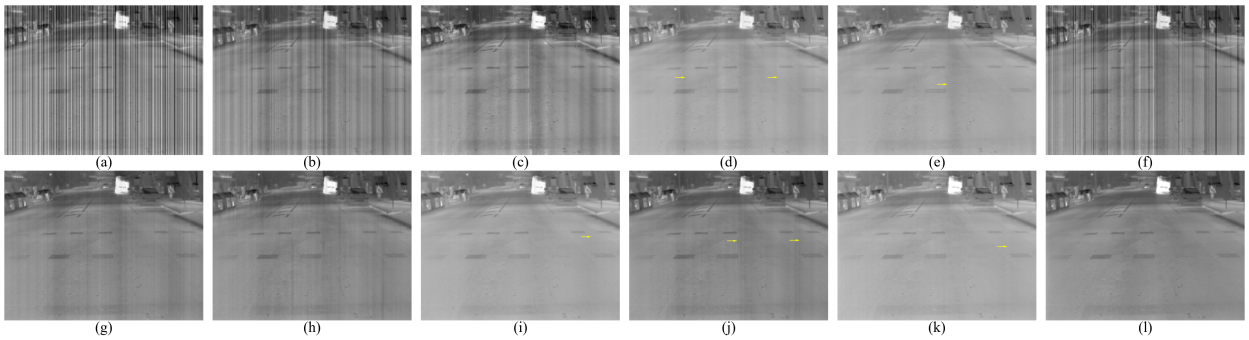


Figure 7. Image destriping results of different methods on CVC09 with Gaussian-distributed simulated noises. Yellow arrows and circles represent the stripe noise residue and blurred image details. (**a**) Degraded. (**b**) GF. (**c**) MIRE. (**d**) WLS. (**e**) ADOM. (**f**) DLS-NUC. (**g**) SNRWDNN. (**h**) CANS. (**i**) AKDT. (**j**) DSCGAN. (**k**) ASCNet. (**l**) BCDANet.

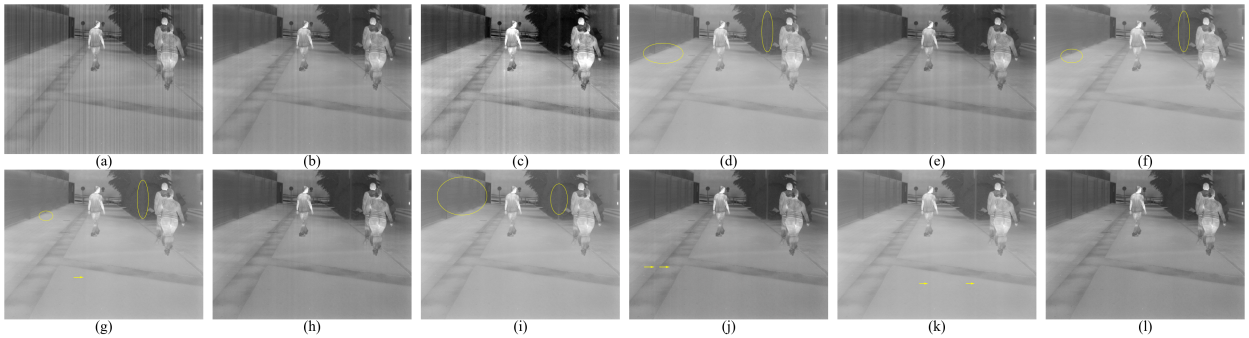


Figure 8. Image destriping results of different methods on CVC09 with uniform-distributed simulated noises. Yellow arrows and circles represent the stripe noise residue and blurred image details. (**a**) Degraded. (**b**) GF. (**c**) MIRE. (**d**) WLS. (**e**) ADOM. (**f**) DLS-NUC. (**g**) SNRWDNN. (**h**) CANS. (**i**)AKDT. (**j**) DSCGAN. (**k**) ASCNet. (**l**) BCDANet.

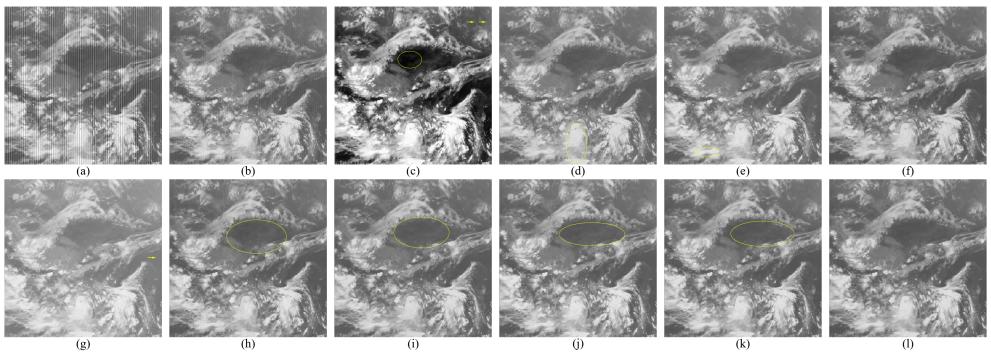


Figure 9. Image destriping results of different methods on ICSRN with periodical-distributed simulated noises. Yellow arrows and circles represent the stripe noise residue and blurred image details. (**a**) Degraded. (**b**) GF. (**c**) MIRE. (**d**) WLS. (**e**) ADOM. (**f**) DLS-NUC. (**g**) SNRWDNN. (**h**) CANS. (**i**) AKDT. (**j**) DSCGAN. (**k**) ASCNet. (**l**) BCDANet.

3.2.3. Anti-Noise Experiment

Figure 10 demonstrates the visual results of the network under Gaussian stripe noise with different intensities. Even under strong noise intensity, BCDA-Net still exhibits excellent image restoration capability.

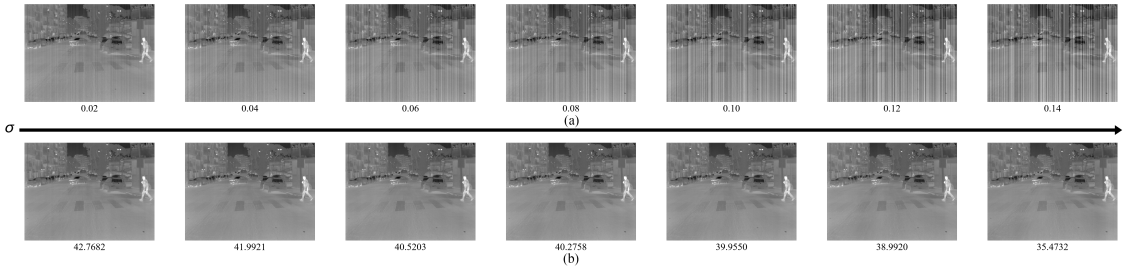


Figure 10. Exemplar stripy images and image destriping results of BCDANet. (**a**) Stripy image with different Gaussian-distributed noise levels. (**b**) Corresponding destriping results of ASCNet.

3.3. Real Image Destriping

To demonstrate the generalization capability of our network across various real-world scenarios, we conducted experiments on 20 distinct scenes. These scenes were sourced from Tiangong-1 satellite infrared images, self-captured UAV infrared images, and the public DLS-NUC [22] infrared dataset and FLIR [54] infrared dataset. These images encompass a rich diversity of stripe noise patterns.

3.3.1. Quantitative Metrics

Considering the absence of reference images (ground truth) for real striped data, and following previous works [32,55], we employ BRISQUE and NIQE to quantitatively evaluate the performance of various destriping methods.

These metrics are particularly suitable for our task because they do not rely on specific distortion types, making them robust against complex, real-world infrared stripe noise. Both BRISQUE and NIQE quantify the statistical deviation or distance of the evaluated image from a model built upon pristine, high-quality natural scenes. Consequently, a lower score implies that the statistical features of the processed image are closer to those of natural, undistorted images, indicating fewer unnatural artifacts (such as residual stripes) and superior human perceptual quality.

As shown in Table 3, BCDA-Net demonstrates exceptional performance, achieving the best results in both NIQE and BRISQUE metrics. This underscores its superior visual perceptual quality.

Table 3. No-reference image quality assessment (NIQE and BRISQUE) on real striped images. Lower is better. The best results are highlighted in **bold** and the second-best results are underlined. The downward arrow (↓) indicates that lower values correspond to better performance

Metric	Method											
	Orgin	GF	MHE	WLS	ADOM	DLSNUC	SNRWDNN	CANS	AKDT	DSCGAN	ASCNet	BCDANet
NIQE ↓	71.95	23.57	16.71	17.99	10.18	10.07	15.78	11.26	12.70	9.89	11.01	**9.65**
BRISQUE ↓	120.77	76.01	32.84	63.27	20.74	55.11	52.79	17.04	17.71	16.89	16.74	**12.58**

3.3.2. Visual Results

Figures 11–14 demonstrate the visual results in real-world scenarios. The regions highlighted in red and green boxes correspond to zoomed-in views for detailed inspection. As can be observed, the proposed method achieves the best destriping performance across various real noise patterns. For instance, only our method successfully avoids introducing artifacts and effectively suppresses strong noise in the space-based imagery in Figure 11. In Figures 12–14, our proposed method strikes an excellent balance by completely removing noise while meticulously preserving the fine structures of the original image.

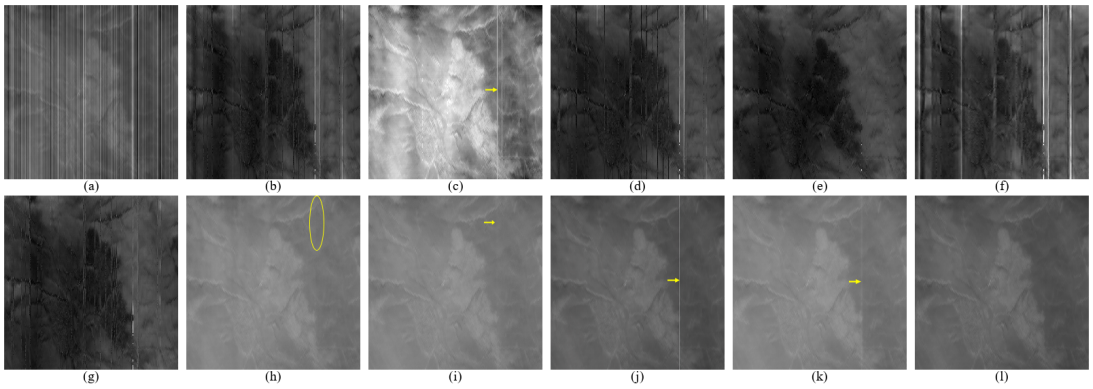


Figure 11. The proposed method was compared visually on a Tiangong-1 satellite dataset. Yellow arrows and circles represent the stripe noise residue or blurred image details. (**a**) Degraded. (**b**) GF. (**c**) MIRE. (**d**) WLS. (**e**) ADOM. (**f**) DLS-NUC. (**g**) SNRWDNN. (**h**) CANS. (**i**) AKDT. (**j**) DSCGAN. (**k**) ASCNet. (**l**) BCDANet.

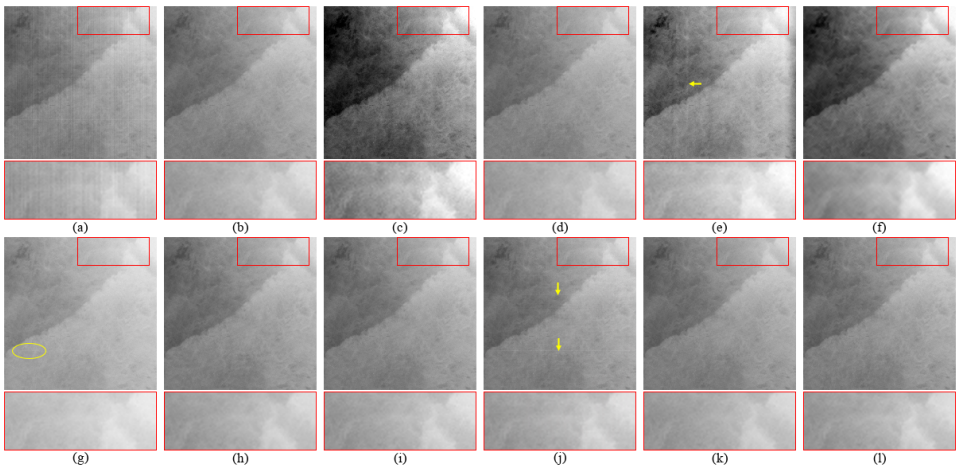


Figure 12. The proposed method was compared visually on a self-captured UAV infrared images dataset. For ease of observation, the texture of the details in the original image is enlarged to compare the denoising effects of different algorithms. Yellow arrows and circles represent the stripe noise residue or blurred image details. (**a**) Degraded. (**b**) GF. (**c**) MIRE. (**d**) WLS. (**e**) ADOM. (**f**) DLS-NUC. (**g**) SNRWDNN. (**h**) CANS. (**i**) AKDT. (**j**) DSCGAN. (**k**) ASCNet. (**l**) BCDANet.

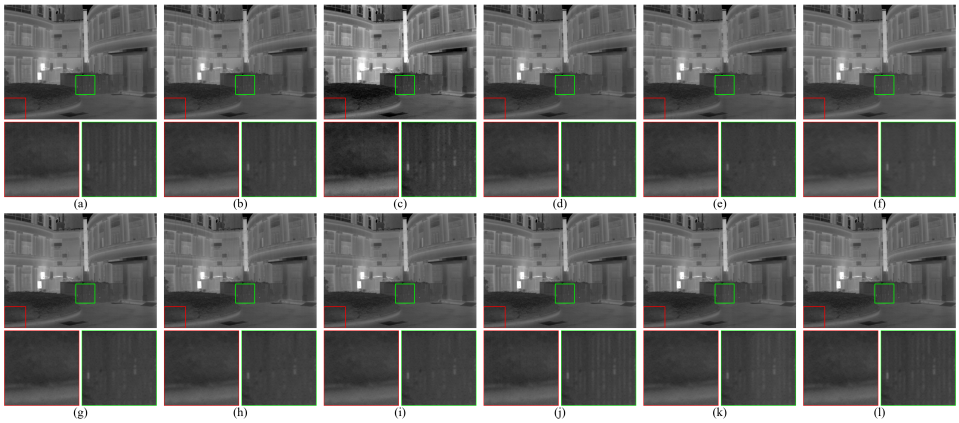


Figure 13. The proposed method was compared visually on a DLS-NUC dataset. For ease of observation, the texture of the details in the original image is enlarged to compare the denoising effects of different algorithms. (**a**) Degraded. (**b**) GF. (**c**) MIRE. (**d**) WLS. (**e**) ADOM. (**f**) DLS-NUC. (**g**) SNRWDNN. (**h**) CANS. (**i**) AKDT. (**j**) DSCGAN. (**k**) ASCNet. (**l**) BCDANet.

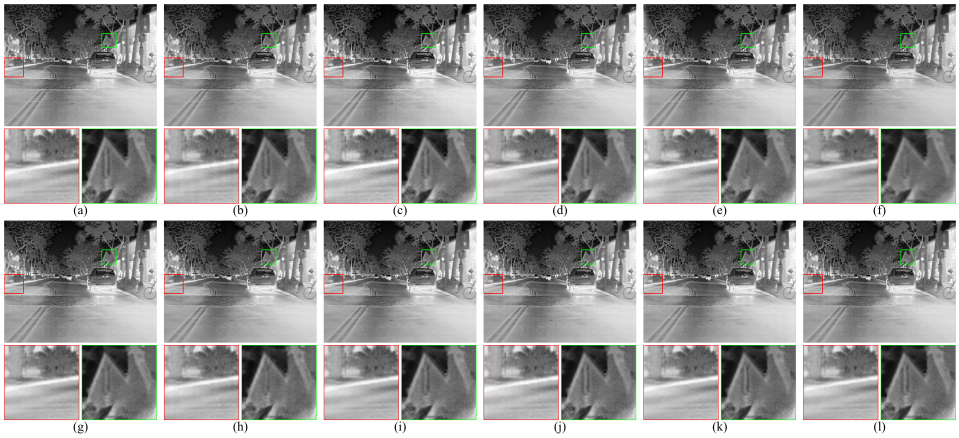


Figure 14. The proposed method was compared visually on a FLIR dataset. For ease of observation, the texture of the details in the original image is enlarged to compare the denoising effects of different algorithms. (**a**) Degraded. (**b**) GF. (**c**) MIRE. (**d**) WLS. (**e**) ADOM. (**f**) DLS-NUC. (**g**) SNRWDNN. (**h**) CANS. (**i**) AKDT. (**j**) DSCGAN. (**k**) ASCNet. (**l**) BCDANet.

3.4. Ablation Study

3.4.1. Ablation Study on Core Modules

To validate the effectiveness and synergistic interactions of the proposed modules (DCD, BFAS, CDFA, and CGSAB) within the overall architecture, we conducted comprehensive ablation studies. By systematically integrating a cumulative addition strategy, pairwise combinations, and leave-one-out experiments, we evaluated the entire test set. The quantitative results are summarized in Table 4.

As shown in Table 4, the network utilizing solely the UNet3+ (Figure 1b) architecture serves as our baseline (Model 1). Following a progressive refinement logic, we first evaluated the cumulative contribution of the proposed modules. As shown in the cumulative track (Models 1, 2, 3, 6 and 8), the network's performance exhibits a consistent and steady upward trend. Starting from the baseline (Model 1), each sequential introduction of the DCD, BFAS, CDFA, and CGSAB modules yields an incremental gain in both PSNR and SSIM. This stepwise improvement effectively demonstrates that every component provides a unique and positive contribution to the final optimal destriping performance.

Table 4. Comprehensive ablation study of the proposed BCDA-Net. ✓indicates the integration of a specific module. The best results are highlighted in **bold**.

Method	DCD	BFAS	CDFA	CGSAB	PSNR (dB)	SSIM
Model 1 (Baseline)					38.0278	0.9831
Model 2	✓				38.2629	0.9847
Model 3	✓	✓			38.3611	0.9853
Model 4	✓		✓		38.2792	0.9855
Model 5	✓			✓	38.4733	0.9847
Model 6	✓	✓	✓		38.5153	0.9862
Model 7	✓	✓		✓	38.8013	0.9875
Model 8 (BCDA-Net)	✓	✓	✓	✓	**39.2634**	**0.9892**

3.4.2. Analysis of Module Interactions

Compared to the pure baseline, incorporating the DCD module (Model 2) yields significant improvements, indicating its effectiveness in extracting multi-scale edge features.

Building upon the DCD backbone, pairwise combination experiments demonstrate strong complementary effects among the modules. For instance, the addition of BFAS (Model 3) further improves performance by effectively blocking the direct propagation of low-resolution features and focusing on high-resolution shallow information. Similarly, the combination with CDFA (Model 4) aids in recovering high-frequency details through its cascaded dense connection design, while adding CGSAB (Model 5) significantly enhances global stripe perception.

Crucially, to explicitly address the indispensable role of the attention mechanism, we analyzed the performance when the CGSAB module is removed (Model 6). Compared to the fully configured network, removing CGSAB causes a severe performance drop of 0.7481 dB in PSNR. This substantial decline proves that without CGSAB to extract precise contextual attention priors, the network struggles to differentiate between complex noise and background details, highlighting its irreplaceable function in feature recalibration. Similarly, removing the CDFA module (Model 7) limits the network's ability to fully utilize hierarchical features for dense reconstruction.

Finally, our complete architecture, BCDA-Net (Model 8), reaches the optimal performance. The results thoroughly demonstrate that these four modules exhibit a highly beneficial synergistic effect, and all are indispensable for the final optimal infrared image destriping.

3.4.3. The Different G Parameters of the SLD Branch of CGSAB

In the SLD branch of the CGSAB, we divide the input feature along the channel dimension into G sub-features to learn different semantics. In this section, we study the optimal value of the parameter G. In Table 5 and Figure 15, we can observe that when G is 2, an optimal balance is achieved between feature decoupling and maintaining sufficient intra-group information capacity, enabling the SLD branch to generate the most accurate multi-scale guidance map. Therefore, in the final model architecture, we empirically set the G of the SLD branch in the CGSAB to 2.

Table 5. Experimental results of different G parameters of the SLD branch of CGSAB on entire test set with Gaussian noise level $\sigma \in [0, 0.15]$. The upward arrow (↑) indicates that larger values correspond to better performance. The best results are highlighted in **bold**.

G	1	2	4	8
PSNR ↑	38.7563	**39.2634**	39.1254	38.6121
SSIM ↑	0.9866	**0.9892**	0.9876	0.9876

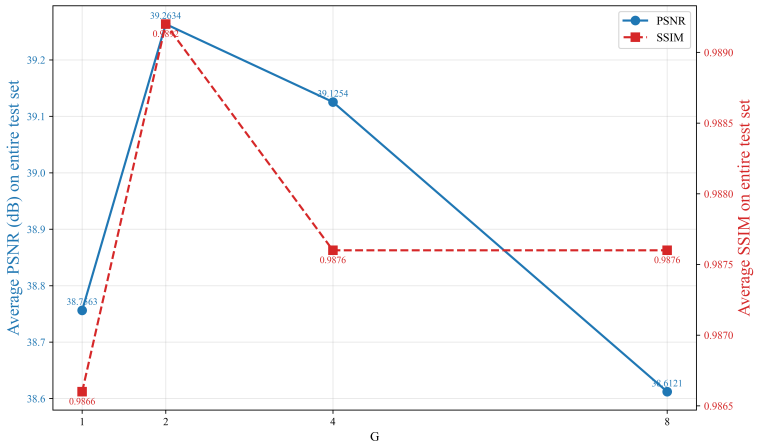


Figure 15. Trend graph of PSNR and SSIM with respect to G value on entire test set with Gaussian noise level $\sigma \in [0, 0.15]$.

3.4.4. Different Structures of CGSAB

As shown in Figure 16, we designed three different branch structures for CGSAB (a: skip connection containing both SLD and GC branches, b: skip connection containing only the SLD branch, c: only the SLD and GC branches). Taking $G = 2$ as the baseline, we conducted experiments on these three structures. The quantitative analysis of the experimental results is presented in Table 6. We observed that on the test set, the a structure outperformed b and c by 1.1869 dB and 1.1959 dB, respectively. The absence of the GC branch fails to accurately capture the contextual information of the entire image, while the absence of skip connections directly learns the output of the entire feature map instead of focusing only on detail features and stripe noise. Consequently, neither can achieve more refined and accurate outputs. Therefore, the CGSAB employing skip connections that include both the SLD and GC branches achieves the best performance.

Table 6. Experimental results of different structures of CGSAB on entire test set with Gaussian noise level $\sigma \in [0, 0.15]$. The upward arrow (↑) indicates that larger values correspond to better performance. The best results are highlighted in **bold**.

Methods	a	b	c
PSNR ↑	**39.2634**	38.0765	38.0675
SSIM ↑	**0.9892**	0.9863	0.9806

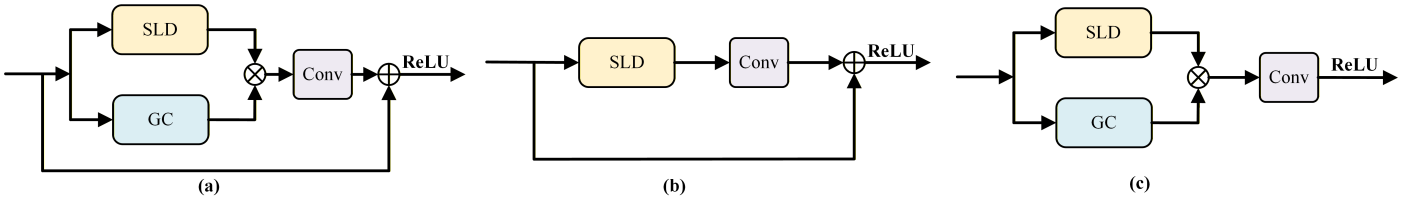


Figure 16. Different structures of CGSAB. (**a**) Skip connection containing both SLD and GC branches. (**b**) Skip connection containing only the SLD branch. (**c**) Only the SLD and GC branches.

3.4.5. Cross-Dataset Validation and Generalization

During the training phase, our model was optimized using 181,752 patch images. To enhance the network's robustness, orthogonal regularization was utilized, and the most complex cubic noise simulation was specifically employed and injected at the beginning of each training epoch to rigorously challenge the network. While extensive testing on two simulated datasets and four real-world scene datasets has already demonstrated a

solid degree of generalization, we conducted an additional zero-shot cross-dataset validation to further rigorously evaluate the model's robustness against scene-level domain shifts. Specifically, we randomly selected 20 representative infrared images from the publicly available RGBT-Tiny [56] dataset to serve as external clean backgrounds. To control variables and isolate the effect of background variation, we injected synthetic stripe noise following the exact same Gaussian distribution as our internal test set. The quantitative results are presented in Table 7.

Table 7. Comparison of cross-dataset generalization performance. The upward arrow (↑) indicates that larger values correspond to better performance, whereas the downward arrow (↓) denotes the opposite. The best results are highlighted in **bold**.

Model	Internal Test PSNR ↑/SSIM ↑	External OOD Test PSNR ↑/SSIM ↑	Performance Drop Δ PSNR ↓/Δ SSIM ↓
Baseline	38.0278/0.9831	34.8713/0.9818	3.1565/0.0013
ASCNet (SOTA)	39.1323/0.9884	38.9770/0.9881	0.1553/**0.0003**
BCDA-Net (Ours)	**39.2634**/**0.9892**	**39.1282**/**0.9887**	**0.1354**/0.0005

The experimental results reveal three key findings:

1. Severe Background Overfitting in Baseline: Despite facing the same noise distribution, the Baseline suffers a massive 3.1565 dB PSNR drop on the RGBT-Tiny backgrounds. This indicates a failure to decouple noise, causing the network to overfit to the specific thermal textures of the training set.
2. Superior Absolute Performance: BCDA-Net achieves the highest absolute PSNR and SSIM on both internal and external datasets, comprehensively outperforming the SOTA ASCNet.
3. Exceptional Cross-Domain Robustness: BCDA-Net demonstrates the most stable PSNR performance with the smallest degradation (0.1354 dB). While ASCNet exhibits a marginally smaller SSIM drop (0.0003 vs. 0.0005), BCDA-Net remains superior in absolute structural preservation (maintaining an SSIM of 0.9887) and proves its overarching robustness in unseen thermal scenarios.

In conclusion, BCDA-Net leverages the CGSAB module to effectively decouple stripes from thermal backgrounds. This ensures robust denoising and superior high-frequency detail preservation, even when applied to entirely new infrared scenarios.

3.5. Hyperparameter Analysis and Model Efficiency

While the fundamental training configurations (e.g., the optimizer, learning rate schedule, and orthogonal regularization) have been introduced in Section 3.1.5, ensuring the full reproducibility of our proposed method requires specifying the architectural details. We provide a comprehensive summary encompassing both the training settings and the key architectural hyperparameters of BCDA-Net in Table 8. Specifically, the table explicitly details the channel dimensions across different stages of the encoder–decoder structure and the number of groups G utilized in the SLD branch.

Based on the architectural configurations detailed above, we further evaluate the computational complexity of BCDA-Net against other state-of-the-art destriping methods. The evaluation strictly focuses on the spatial and computational overhead, specifically the number of trainable parameters (Params) and the computational cost measured in floating-point operations (FLOPs).

As illustrated in Figure 17, our optimized BCDA-Net (marked with a red star) consistently occupies the optimal top-left region in both subfigures (a) and (b), which plot PSNR versus the number of parameters (M, Log Scale) and computational cost (FLOPs, G,

Log Scale), respectively. This positioning demonstrates that BCDA-Net achieves highly competitive restoration performance while maintaining a significantly lower computational footprint. For instance, although models like DLSNUC have fewer parameters, their FLOPs are substantially higher due to the lack of downsampling, making them less suitable for real-time processing.

Table 8. Comprehensive summary of training configurations and architectural hyperparameters of BCDA-Net.

Category	Parameter/Configuration	Value/Setting
Training Settings	Optimizer	Adam
	Learning Rate Scheduler	Cosine Annealing
	Initial Learning Rate	1×10^{-3}
	Batch Size	256
	Total Epochs	450
	Hardware	NVIDIA GeForce 4090
Architectural	Base Channel Width (C)	16
	Encoder Channels	16(E0), 32(E1), 64(E2), 128(E3)
	Decoder Channels	16×4(D3), 16×4(D2), 16×4(D1), 16×4(D0)
	Number of Groups (G) in SLD	2

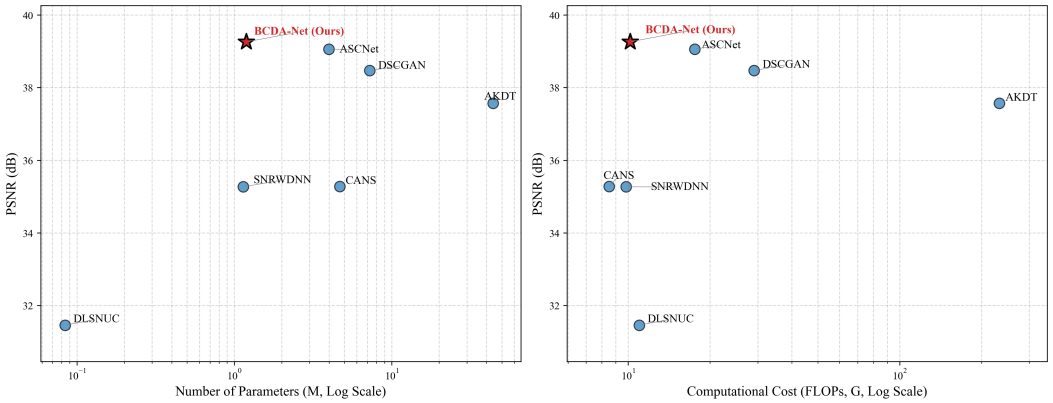


Figure 17. Comprehensive evaluation of model complexity and efficiency. (**a**) Comparison of destriping performance (PSNR) versus the number of parameters. (**b**) Comparison of destriping performance versus computational cost (FLOPs). Note that the horizontal axes in both subplots are presented in logarithmic scale to accommodate the large variation in model sizes.

In contrast, our BCDA-Net model effectively preserves fine-grained details and edges while keeping the total parameter count and FLOPs strictly controlled. Ultimately, these results demonstrate that BCDA-Net yields an optimal trade-off between model complexity and performance, which is a critical advantage for wide-swath high-resolution infrared remote sensing on-orbit real-time processing.

4. Discussion

While the comprehensive experimental results, including ablation studies and zero-shot cross-dataset validation, have quantitatively and qualitatively demonstrated the superiority, robustness, and high efficiency of the proposed BCDA-Net, several broader implications and practical considerations regarding its application in infrared remote sensing warrant further discussion.

Preservation of Fragile Infrared Features via Bottleneck-Free Architecture. A primary challenge in infrared remote sensing is that critical information, such as thermal boundaries and small target signatures, is extremely fragile due to low contrast. Conventional encoder–decoder architectures often suffer from an “informational bottleneck”, where suc-

cessive upsampling operations cause irreversible loss of high-frequency spatial details. The BCDA-Net addresses this by introducing the Bottleneck-Free Aggregation Strategy (BFAS). Specifically, the first Hierarchical Feature Aggregation Network (HFAN) is designed without an upsampling operator. Instead, it directly integrates the shallowest structural features from the first stage of the encoder. This "bypass" mechanism ensures that the most fine-grained radiometric details skip the deep informational bottleneck and participate directly in the reconstruction process, physically guaranteeing the sharp retention of thermal edges.

Complementary Feature Extraction through Dual-Channel Downsampling. The effectiveness of BCDA-Net also stems from its ability to disentangle complex stripe noise from actual thermal radiation signals. In our Dual-Channel Downsampling (DCD) module, we employ a complementary strategy: one half of the feature channels undergo max pooling, while the other half utilize strided convolution. This design is specifically tailored to the physical characteristics of infrared imagery. The max-pooling path is highly sensitive to local intensity peaks, effectively preserving salient thermal hotspots and distinct geometric boundaries. Conversely, the strided convolution path provides a learnable spatial transformation, meticulously capturing the anisotropic structural correlations of the stripe noise and the subtle gradients of the background. By aggregating these decoupled features, the network achieves a superior balance between noise suppression and detail preservation that single-path downsampling cannot match.

Pathways to On-Orbit Real-Time Deployment and Limitations. Although our efficiency analysis confirms that BCDA-Net achieves an optimal trade-off between performance and computational complexity (FLOPs and parameter count) on standard GPU platforms, deploying deep learning models for wide-swath high-resolution infrared remote sensing on-orbit real-time processing remains a challenging frontier. Spaceborne hardware platforms, such as Field Programmable Gate Arrays (FPGAs) or embedded Neural Processing Units (NPUs), have strict constraints on power consumption and memory bandwidth. A current limitation of our work is that it has not yet been optimized for hardware-level deployment. Future research will focus on extreme model compression, such as 8-bit quantization and hardware–software co-design, to bridge the gap between algorithmic simulation and actual on-orbit edge computing. Furthermore, exploring the model's blind destriping capability under entirely unpredictable, dynamically changing sensor degradation represents another crucial direction.

5. Conclusions

In this paper, we propose a Bottleneck-free Channel Dual-path Aggregation Network (BCDA-Net) to address the nonuniform stripe noise prevalent in infrared focal plane array imaging and the limitations of existing denoising methods in texture preservation.

First, to tackle the high-frequency information loss during downsampling, we design a Dual-path Channel Downsampling (DCD) module. By employing a channel splitting strategy, this module achieves parameter-free preservation of texture features and effective extraction of semantic features. Second, to eliminate aliasing artifacts introduced by deep feature upsampling, we propose a Bottleneck-Free Aggregation Strategy (BFAS), which architecturally blocks the propagation of low-resolution noise. Furthermore, by combining the Context-Guided Stripe Attention Block (CGSAB) with the Cascaded Dense Feature Aggregation (CDFA) module, the network successfully achieves precise perception of global nonuniform stripes and progressive reconstruction of weak thermal radiation targets.

Experimental results show that compared with state-of-the-art methods, BCDA-Net not only achieves significant improvements in objective metrics such as PSNR and SSIM but also effectively removes stripe noise of various intensities in subjective visual quality, without introducing edge blurring caused by over-smoothing.

Author Contributions: Conceptualization, L.C.; methodology, L.C.; software, L.C.; validation, L.C.; formal analysis, L.C.; investigation, L.C.; resources, Y.F., F.D. and L.H.; data curation, L.C.; writing—original draft preparation, L.C.; writing—review and editing, L.C., Y.F., F.D. and L.H.; visualization, L.C.; supervision, Y.F. and F.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Infrared/Visible Multi-channel Scanner Project, grant number JT02408150006.

Data Availability Statement: The original contributions presented in this study are included in the article; further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Hou, F.; Zhang, Y.; Zhou, Y.; Zhang, M.; Lv, B.; Wu, J. Review on Infrared Imaging Technology. *Sustainability* **2022**, *14*, 11161. [CrossRef]
2. Wei, J.; Wang, M. Mapping ocean surface algal blooms with SWIR-derived satellite remote sensing reflectance. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *131*, 103921. [CrossRef]
3. Chen, Q.; Lin, Q.; Xuan, Y.; Han, Y. Investigation on the thermohaline structure of the stratified wake generated by a propagating submarine. *Int. J. Heat Mass Transf.* **2021**, *166*, 120808. [CrossRef]
4. Milton, A.F.; Barone, F.R.; Kruer, M.R. Influence of Nonuniformity on Infrared Focal Plane Array Performance. *Opt. Eng.* **1985**, *24*, 245855. [CrossRef]
5. Chen, N.; Zhang, J.; Zhong, S.; Mao, W.; Hu, D.; Yao, L. Two-Dimensional Calibration for Fixed-Pattern Noise Reduction of Thermal Images. In *Proceedings of the 2019 IEEE International Symposium on Circuits and Systems (ISCAS)*; IEEE: New York, NY, USA, 2019; pp. 1–5. [CrossRef]
6. Cao, Y.; Yang, M.Y.; Tisse, C.L. Effective Strip Noise Removal for Low-Textured Infrared Images Based on 1-D Guided Filtering. *IEEE Trans. Circuits Syst. Video Technol.* **2016**, *26*, 2176–2188. [CrossRef]
7. Zhou, X.; Zhang, Y.; Liu, J.; Hu, Y. A Novel Two-Stage Destriping Algorithm Based on MWIR Energy Separation and Image Guidance (MES-IG). *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5001316. [CrossRef]
8. Chen, B.; Feng, X.; Wu, R.; Guo, Q.; Wang, X.; Ge, S. Adaptive Wavelet Filter with Edge Compensation for Remote Sensing Image Denoising. *IEEE Access* **2019**, *7*, 91966–91979. [CrossRef]
9. Wang, E.; Jiang, P.; Hou, X.; Zhu, Y.; Peng, L. Infrared Stripe Correction Algorithm Based on Wavelet Analysis and Gradient Equalization. *Appl. Sci.* **2019**, *9*, 1993. [CrossRef]
10. Zeng, Q.; Qin, H.; Yan, X.; Yang, T. Fourier Domain Anomaly Detection and Spectral Fusion for Stripe Noise Removal of TIR Imagery. *Remote Sens.* **2020**, *12*, 3714. [CrossRef]
11. Zeng, Q.; Qin, H.; Yan, X.; Zhou, H. Fourier Spectrum Guidance for Stripe Noise Removal in Thermal Infrared Imagery. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 1072–1076. [CrossRef]
12. Zhao, J.; Zhou, Q.; Chen, Y.; Liu, T.; Feng, H.; Xu, Z.; Li, Q. Single image stripe nonuniformity correction with gradient-constrained optimization model for infrared focal plane arrays. *Opt. Commun.* **2013**, *296*, 47–52. [CrossRef]
13. Wang, E.; Jiang, P.; Li, X.; Cao, H. Infrared stripe correction algorithm based on wavelet decomposition and total variation-guided filtering. *J. Eur. Opt. Soc.-Rapid Publ.* **2019**, *16*, 1. [CrossRef]
14. Cao, W.; Chang, Y.; Han, G.; Li, J. Destriping Remote Sensing Image via Low-Rank Approximation and Nonlocal Total Variation. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 848–852. [CrossRef]
15. Li, M.; Nong, S.; Nie, T.; Han, C.; Huang, L.; Qu, L. A Novel Stripe Noise Removal Model for Infrared Images. *Sensors* **2022**, *22*, 2971. [CrossRef]
16. Kim, N.; Han, S.S.; Jeong, C.S. ADOM: ADMM-Based Optimization Model for Stripe Noise Removal in Remote Sensing Image. *IEEE Access* **2023**, *11*, 106587–106606. [CrossRef]
17. Horn, B.K.; Woodham, R.J. Destriping LANDSAT MSS images by histogram modification. *Comput. Graph. Image Process.* **1979**, *10*, 69–83. [CrossRef]
18. Wegener, M. Destriping multiple sensor imagery by improved histogram matching. *Int. J. Remote Sens.* **1990**, *11*, 859–875. [CrossRef]
19. Tendero, Y.; Landeau, S.; Gilles, J. Non-uniformity Correction of Infrared Images by Midway Equalization. *Image Process. Line* **2012**, *2*, 134–146. [CrossRef]
20. Xu, K.; Zhao, Y.; Li, F.; Xiang, W. Single infrared image stripe removal via deep multi-scale dense connection convolutional neural network. *Infrared Phys. Technol.* **2022**, *121*, 104008. [CrossRef]

21. Fayyaz, Z.; Platnick, D.D.; Fayyaz, H.; Farsad, N. Deep Unfolding for Iterative Stripe Noise Removal. In *2022 International Joint Conference on Neural Networks (IJCNN)*; IEEE: New York, NY, USA, 2022; pp. 1–7. [CrossRef]
22. He, Z.; Cao, Y.; Dong, Y.; Yang, J.; Cao, Y.; Tisse, C.L. Single-image-based nonuniformity correction of uncooled long-wave infrared detectors: A deep-learning approach. *Appl. Opt.* **2018**, *57 18*, D155–D164. [CrossRef]
23. Li, T.; Zhao, Y.; Li, Y.; Zhou, G. Non-Uniformity Correction of Infrared Images Based on Improved CNN with Long-Short Connections. *IEEE Photonics J.* **2021**, *13*, 7800213. [CrossRef]
24. Li, J.; Zeng, D.; Zhang, J.; Han, J.; Mei, T. Column-Spatial Correction Network for Remote Sensing Image Destriping. *Remote Sens.* **2022**, *14*, 3376. [CrossRef]
25. Guan, J.; Lai, R.; Xiong, A.; Liu, Z.; Gu, L. Fixed pattern noise reduction for infrared images based on cascade residual attention CNN. *Neurocomputing* **2020**, *377*, 301–313. [CrossRef]
26. Dong, Y.; Zhao, B.; Yang, J.; Cao, Y.; Cao, Y. Two-stage convolutional neural network for joint removal of sensor noise and background interference in lock-in thermography. *NDT&E Int.* **2023**, *137*, 102816. [CrossRef]
27. Hu, X.; Luo, S.; He, C.; Wu, W.; Wu, H. Infrared thermal image denoising with symmetric multi-scale sampling network. *Infrared Phys. Technol.* **2023**, *134*, 104909. [CrossRef]
28. Li, Z.; Luo, S.; Chen, M.; Wu, H.; Wang, T.; Cheng, L. Infrared thermal imaging denoising method based on second-order channel attention mechanism. *Infrared Phys. Technol.* **2021**, *116*, 103789. [CrossRef]
29. Xiao, Z.; Qin, H.; Yang, S.; Yan, X.; Zhou, H. Spatial–Spectral Oriented Triple Attention Network for Hyperspectral Image Denoising. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5503717. [CrossRef]
30. Wang, C.; Xu, M.; Jiang, Y.; Zhang, G.; Cui, H.; Li, L.; Li, D. Translution-SNet: A Semisupervised Hyperspectral Image Stripe Noise Removal Based on Transformer and CNN. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5533114. [CrossRef]
31. Wang, M.; Yuan, P.; Qiu, S.; Jin, W.; Li, L.; Wang, X. Dual-Encoder UNet-Based Narrowband Uncooled Infrared Imaging Denoising Network. *Sensors* **2025**, *25*, 1476. [CrossRef]
32. Yuan, S.; Qin, H.; Yan, X.; Yang, S.; Yang, S.; Akhtar, N.; Zhou, H. ASCNet: Asymmetric Sampling Correction Network for Infrared Image Destriping. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5001815. [CrossRef]
33. Chang, Y.; Chen, M.; Yan, L.; Zhao, X.L.; Li, Y.; Zhong, S. Toward Universal Stripe Removal via Wavelet-Based Deep Convolutional Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 2880–2897. [CrossRef]
34. Zhang, S.; Sui, X.; Yao, Z.; Gu, G.; Chen, Q. Research on nonuniformity correction based on deep learning. In *Proceedings of the AOPC 2021: Infrared Device and Infrared Technology*; Gong, H., Shi, Z., Lu, J., Eds.; International Society for Optics and Photonics, SPIE: Bellingham, WA, USA, 2021; Volume 12061, p. 120610G. [CrossRef]
35. Guan, J.; Lai, R.; Xiong, A. Wavelet Deep Neural Network for Stripe Noise Removal. *IEEE Access* **2019**, *7*, 44544–44554. [CrossRef]
36. Yang, S.; Qin, H.; Yuan, S.; Yan, X.; Rahmani, H. DestripeCycleGAN: Stripe Simulation CycleGAN for Unsupervised Infrared Image Destriping. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 5039414. [CrossRef]
37. Yang, P.; Wu, H.; Cheng, L.; Luo, S. Infrared image denoising via adversarial learning with multi-level feature attention network. *Infrared Phys. Technol.* **2023**, *128*, 104527. [CrossRef]
38. Li, L.; Wang, X.; Huang, D.; He, Y.; Zhong, Z.; Xia, Q. KEDM: Knowledge-Embedded Diffusion Model for Infrared Image Destriping. *IEEE Photonics J.* **2025**, *17*, 7800309. [CrossRef]
39. Wang, N.; Huang, L.; Li, M.; Zhou, B.; Nie, T. Joint Deblurring and Destriping for Infrared Remote Sensing Images with Edge Preservation and Ringing Suppression. *Remote Sens.* **2026**, *18*, 150. [CrossRef]
40. Zhang, J.; Zhou, X.; Li, L.; Hu, T.; Fansheng, C. A Combined Stripe Noise Removal and Deblurring Recovering Method for Thermal Infrared Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5003214. [CrossRef]
41. Brateanu, A.; Balmez, R.; Avram, A.; Orhei, C. AKDT: Adaptive Kernel Dilation Transformer for Effective Image Denoising. In *Proceedings of the 20th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications—Volume 3: VISAPP*; SciTePress: Setúbal, Portugal, 2025; pp. 418–425. [CrossRef]
42. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H. Restormer: Efficient Transformer for High-Resolution Image Restoration. *arXiv* **2022**, arXiv:2111.09881. [CrossRef]
43. Chen, X.; Wang, X.; Zhang, W.; Kong, X.; Qiao, Y.; Zhou, J.; Dong, C. HAT: Hybrid Attention Transformer for Image Restoration. *arXiv* **2025**, arXiv:2309.05239. [CrossRef]
44. Ma, T.; Ma, L.; Li, Z.; Wang, Y.; Liu, J.; Xu, C.; Liu, R. Rethinking Reconstruction and Denoising in the Dark: New Perspective, General Architecture and Beyond. In *Proceedings of the 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; Computer Vision Foundation: New York, NY, USA, 2025; pp. 2323–2332. [CrossRef]
45. Huang, H.; Lin, L.; Tong, R.; Hu, H.; Zhang, Q.; Iwamoto, Y.; Han, X.; Chen, Y.W.; Wu, J. UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation. *arXiv* **2020**, arXiv:2004.08790. [CrossRef]
46. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. *arXiv* **2014**, arXiv:1409.4842. [CrossRef]

47. Yu, C.; Wang, J.; Peng, C.; Gao, C.; Yu, G.; Sang, N. BiSeNet: Bilateral Segmentation Network for Real-time Semantic Segmentation. *arXiv* **2018**, arXiv:1808.00897. [CrossRef]
48. Hao, Z.; Gai, S.; Li, P. Multi-Scale Self-Calibrated Dual-Attention Lightweight Residual Dense Deraining Network Based on Monogenic Wavelets. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *33*, 2642–2655. [CrossRef]
49. Ouyang, D.; He, S.; Zhang, G.; Luo, M.; Guo, H.; Zhan, J.; Huang, Z. Efficient Multi-Scale Attention Module with Cross-Spatial Learning. In Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4–10 June 2023; IEEE: New York, NY, USA, 2023; pp. 1–5. [CrossRef]
50. Xiao, P.; Guo, Y.; Zhuang, P. Removing Stripe Noise From Infrared Cloud Images via Deep Convolutional Networks. *IEEE Photonics J.* **2018**, *10*, 7801114. [CrossRef]
51. Socarrás, Y.; Ramos, S.; Vázquez, D.; López, A.M.; Gevers, T. Adapting Pedestrian Detection from Synthetic to Far Infrared Images. In Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW), Sydney, Australia, 2–8 December 2013; Volume 3, pp. 1–3.
52. Zhang, K.; Zuo, W.; Zhang, L. FFDNet: Toward a Fast and Flexible Solution for CNN-Based Image Denoising. *IEEE Trans. Image Process.* **2018**, *27*, 4608–4622. [CrossRef]
53. Li, F.; Zhao, Y.; Xiang, W. Single-frame-based column fixed-pattern noise correction in an uncooled infrared imaging system based on weighted least squares. *Appl. Opt.* **2019**, *58*, 9141–9153. [CrossRef]
54. FLIR Systems. Free FLIR Thermal Dataset for Algorithm Training. 2018. Available online: https://oem.flir.com/solutions/automotive/dataset (accessed on 27 February 2026).
55. Tang, Y.; Min, C.; Miao, R.; Lu, J. Self-Supervised Infrared Image Denoising via Adaptive Gradient-Perception Network for FPN Suppression. *Electronics* **2025**, *14*, 4334. [CrossRef]
56. Ying, X.; Xiao, C.; An, W.; Li, R.; He, X.; Li, B.; Cao, X.; Li, Z.; Wang, Y.; Hu, M.; et al. Visible-Thermal Tiny Object Detection: A Benchmark Dataset and Baselines. *IEEE Trans. Pattern Anal. Mach. Intell.* **2025**, *47*, 6088–6096. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.