A Multi Scale Strip-wise ConvNet for Infrared Image Stripe Removal

Wei You [1], Zhiqiang Bian [2], Xinbo Zhou [2], Yi Xu [1], Kaimin Sun [1]

[1] Wuhan University, State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing, Wuhan 430072, China - youwei2567@whu.edu.cn, 2023186190070@whu.edu.cn, sunkm@whu.edu.cn

[2] Shanghai Institute of Satellite Engineering, Shanghai 201109, China - bianzhiqiang2003@163.com, xiaoxiinbobo@163.com

Keywords: Image Destriping, Strip Convolution, Wavelet Transform, Dense Connection, Infrared Small Target Detection.

Abstract

Unmanned Aerial Vehicles (UAVs) often perform all-day all-weather missions such as nighttime surveillance and search-and-rescue, where visible-light sensors fail, motivating the use of long-wave infrared (LWIR) cameras. However, IR images suffer from horizontal or vertical stripe artifacts caused by sensor instability, which degrade image quality. Conventional methods struggle with non-periodic stripes due to their reliance on fixed frequency priors and stationary pattern assumptions, while learning-based methods often neglect structural features because of insufficient inductive bias. To further exploit the anisotropic geometry of stripe artifacts, we design a series of strip-wise convolution kernels with multiple kernel lengths. To address this, we propose multi-scale strip-wise convolution kernels that extend along stripe orientations, providing an expanded receptive field in the principal direction while limiting orthogonal interference. Multiple kernel sizes enable the network to capture both local distortions and long-range stripe structures, and dense connectivity promotes cross-scale feature interaction for effective stripe separation. This architecture, which we call Densely Connected Multi-Scale Strip-Conv (DCMS), constitutes the core of our proposed method. Furthermore, we conduct experiments on infrared image datasets and carry out ablation experiments to validate the efficacy of our innovative method and its modules. Experimental results demonstrate that our method achieves superior performance compared to state-of-the-art approaches in both quantitative metrics and visual quality.

1. Introduction

In Unmanned Aerial Vehicles (UAVs) applications, infrared sensors are widely used for detecting obstacles, vehicles, and humans, particularly in low-light or all-weather conditions. During the acquisition of infrared images, stripe artifacts may arise due to various factors, including temperature fluctuations, sensor instability, inconsistent response among receivers, and interference from dark current and so on. Inadequate suppression of stripes can lead to the loss of image details, which can have a negative impact on the subsequent tasks like detection or navigation, etc. Therefore, the process of removing stripes, which aims to reduce the differences between columns and improve image quality, is essential for these subsequent tasks. Many kinds of methods have been proposed to remove stripes in the few past decades. Existing methods can be roughly divided into four categories: statistical-based, filtering-based, optimization-based and deep learning-based. Regardless, these previous methods have pros and cons.

Infrared imagery can be categorized into different spectral ranges, including near-infrared (NIR), short-wave infrared (SWIR), mid-wave infrared (MWIR), and long-wave infrared (LWIR). In this work, we specifically focus on LWIR imagery, which is widely used in UAV-based thermal sensing. Unlike NIR or SWIR imaging, LWIR captures emitted thermal radiation and does not rely on external illumination. It is important to note that destriping methods designed for LWIR data may not directly generalize to other spectral bands due to differences in sensor characteristics and noise patterns.

Recent studies have explored statistical-based approaches and filtering-based methods for infrared image destriping, leveraging sparsity, directional priors, and optimization frameworks to suppress stripe artifacts. Collectively, these methods demonstrate the effectiveness of statistical priors and optimization strategies in mitigating stripe artifacts, particularly when conventional filtering approaches are insufficient. The filtering-based methods (Cao et al., 2016; Wang et al., 2019) suppress noise by selecting an appropriate form of filter to promote the uniformity of polluted images. With the arise of filtering algorithms in all kinds of image process, destriping are seen as a form of noise removal so that we can address this issue in both spatial and frequency domains. In recent years, with the proposal of Guided Filter (GF), Wang (Wang et al., 2019) and Cao (Cao et al., 2016) have incorporated it to extract stripe information. Nonetheless, the noise is inevitably coupled to the signal, which makes the filter-based method difficult to handle.

The optimization-model method regards stripe removal task as an ill-posed inverse problem within constraint terms to optimize step-by-step, such as total variation (Yan et al., 2023; Yi Chang et al., 2015), low-rank property (H. Zhang et al., 2022) or sparsity (Chang et al., 2014; Zhao and Yang, 2015). Optimization-based methods address the challenge of removing stripe noise from image by formulating the problem within an optimization framework. These methods are particularly adept at leveraging the incorporation of prior knowledge about images and noise characteristics into a constructed objective function, followed by an iterative solution process to estimate the original noise-free image. The primary advantage of these methods lies in their flexibility and adaptability to various types of stripe noise. By employing different regularization techniques such as total variation (Wang et al., 2019; Yan et al., 2023; Yi Chang et al., 2015), sparsity constraints (Jiang et al., 2021; Song and Huang, 2023), and low-rank approximations (Chang et al., 2016), they significantly enhance performance while preserving important image features. However, these methods also face challenges. The selection of appropriate regularization terms and their parameters is crucial and can significantly impact the results, requiring careful tuning. Moreover, the computational complexity associated with these methods, especially for large images or complex noise patterns, necessitates the development of efficient optimization algorithms and computational strategies.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-383-2026 | © Author(s) 2026. CC BY 4.0 License.

Learning-based methods (Chang et al., 2020; Fayyaz et al., 2022; Guan et al., 2019; Kuang et al., 2017; Li et al., 2022a; Xu et al., 2022) address stripe removal by leveraging the strong representation capability of convolutional neural networks.With substantial data, learning-based methods offer diverse approaches to address stripe noise in images. SNRCNN (Kuang et al., 2017), DLS-NUC (He et al., 2018), DMRN (Chang et al., 2019) use plain model to demonstrate the robust feature representation capabilities of convolutional networks while they are still insufficient in model construction. Compared with former methodologies, SNRWDNN (Guan et al., 2019), TSWEU (Chang et al., 2020) attempt to integrate wavelet transform into convolutional frameworks, but their performance is still limited in handling complex non-periodic stripe patterns. DINR (Fayyaz et al., 2022), DnRCNN (Guan et al., 2023) uses recurrent networks for iterative refinement, though at a computational cost. DMD-CNN (Xu et al., 2022) effectively targets noise at multiple scales, yet increases model size. MAM (Li et al., 2022b) tailors to specific materials, demanding careful parameter tuning. Despite their unique strengths and challenges, these methods significantly advance stripe noise reduction in imaging applications.

While learning-based approaches demonstrate flexibility and adaptability in addressing non-periodic fixed pattern noise (FPN), these methods have overlooked the structural information inherent in the stripes. This neglect may hinder the model's ability to fully leverage the texture and pattern information present in the imagery, thereby impacting the performance of denoising and restoration tasks. In light of the identified shortcomings of learning-based methods in capturing the structural information of stripes, we endeavor to integrate strip convolution into our convolutional model to address this limitation.

Strip convolution is an effective technique in computer vision (Cui et al., 2023; Cui and Knoll, 2024, 2023) for capturing long-range dependencies in images, particularly useful in semantic segmentation tasks involving elongated structures (Zhou et al., 2023). It operates by applying convolutions in a limited window that slides across the image, focusing on features within strips (Hou et al., 2020). This method enhances the network's ability to extract linear features and improves segmentation accuracy for objects like roads or rivers in satellite imagery (Zhou et al., 2023). The efficiency of strip convolution comes from its localized processing, which reduces computational load compared to full-image convolutions (Ni et al., 2024). Recent advances have integrated strip convolution into network architectures like DPSDA-Net (Zhao et al., 2023), demonstrating its potential to address the challenges of capturing narrow and elongated structures in high-resolution images. Overall, strip convolution provides a valuable tool for enhancing feature extraction and improving performance in specific computer vision applications.

Nevertheless, the aforementioned issues frequently include the following challenges:
1) The forms of stripe assumed in the aforementioned traditional methods are too ideal and simplistic to be useful in practical scenarios. Certain methods rely on prior assumptions that are specific to a particular type of stripe, such as periodicity, and may not effectively address mixed noise.
2) Variations in gains and offsets among adjacent sensors in images compromise local homogeneity, forming the structural characteristics of stripe noise. However, deep learning-based approaches fail to utilize it.
3) Due to the convolution's inductive bias, deep learning-based methods may disregard non-local information, despite the significance of long-range dependencies in extracting structural characteristics of stripes.

To address these problems, we make contributions as follows:
1) We present a U-shape model that employs wavelet transform and its inverse transform as the input and output. And we choose Haar Discrete Wavelet Transform (HDWT) as our wavelet for its convenience and efficiency in the image conversion to extract structural information of stripes.
2) In order to learn stripe structural characteristics, we design a novel feature extraction module that is formed by several strip-wise convolution layers with diverse kernel sizes. This kind of multi-scale strip-wise convolution blocks can capture both local and non-local information in the image.
3) To further improve the performance of our model, we densely connect these strip convolution layers to discern differences between reference image and corrupted image, thereby addressing the issues of gradient vanishing and enhancing feature representation.

2. Methodology

2.1 Overall Architecture

The overall architecture of our proposed model is illustrated in Fig 1. Before a degraded image $I_D \in \Re^{1\times H\times W}$. is given to model, it is transformed by HDWT to get the degraded wavelet features $I_H^T \in \Re^{4\times\frac{H}{2}\times\frac{W}{2}}$, where image I_D is transformed into 4 different sub-band images as their resolution reduce into a half one. Then I_H^T is fed into the model and use a single 3×3 convolution layer with activation to extract shallow features where C is the number of channels. Subsequently, the shallow features are fed into the four-scale encoder-decoder architecture to learn the in-depth features.

The encoder is designed with a down-sampling layer and a Densely Connected Multi-Scale Strip-Conv (DCMS) block, while the decoder comprises an up-sampling layer and a DCMS block. The difference is that the decoder should concatenate features from the encoder and the previous decoder. Before concatenating the features from the previous decoder, a 1×1 convolution layer is followed serving the purpose of reducing the channel number by half, aka transition layer. In each encoder, there is an increase in the number of channels accompanied by a decrease in resolution, whereas in each decoder, the number of channels decreases while the resolution increases.

Finally, a single 3×3 convolution layer is applied to high-resolution features to produce 4-channel residual features $F_R \in \Re^{4\times\frac{H}{2}\times\frac{W}{2}}$, to which the degraded wavelet features I_H^T is added to generate the restored wavelet features O_H^T via $O_H^T = F_R + I_H^T$. And then, we get the restored image O through inverse transform of HDWT.

2.2 Wavelet Transform Feature as Input

Wavelet transform is a mathematical technique that has found applications in various fields due to its ability to provide a time-frequency representation of signals. The 2D Discrete Wavelet Transform (DWT) is an extension of the 1D DWT to two dimensions, making it a powerful tool for image processing. It decomposes an image into its frequency components in a spatially localized manner.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-383-2026 | © Author(s) 2026. CC BY 4.0 License.

HDWT is the simplest orthogonal wavelet with a clear and straightforward structure, meaning that the transform is reversible without any information loss. Besides, the detail sub-bands highlight the edges in the image, making HDWT useful for the stripe removal task. As a result, we choose HDWT as our wavelet for these reasons: images can be easily compressed and rebuild, reduce the model complexity by reducing the input image size, contain stripe structural information hidden in the sub-band of the transformed image.

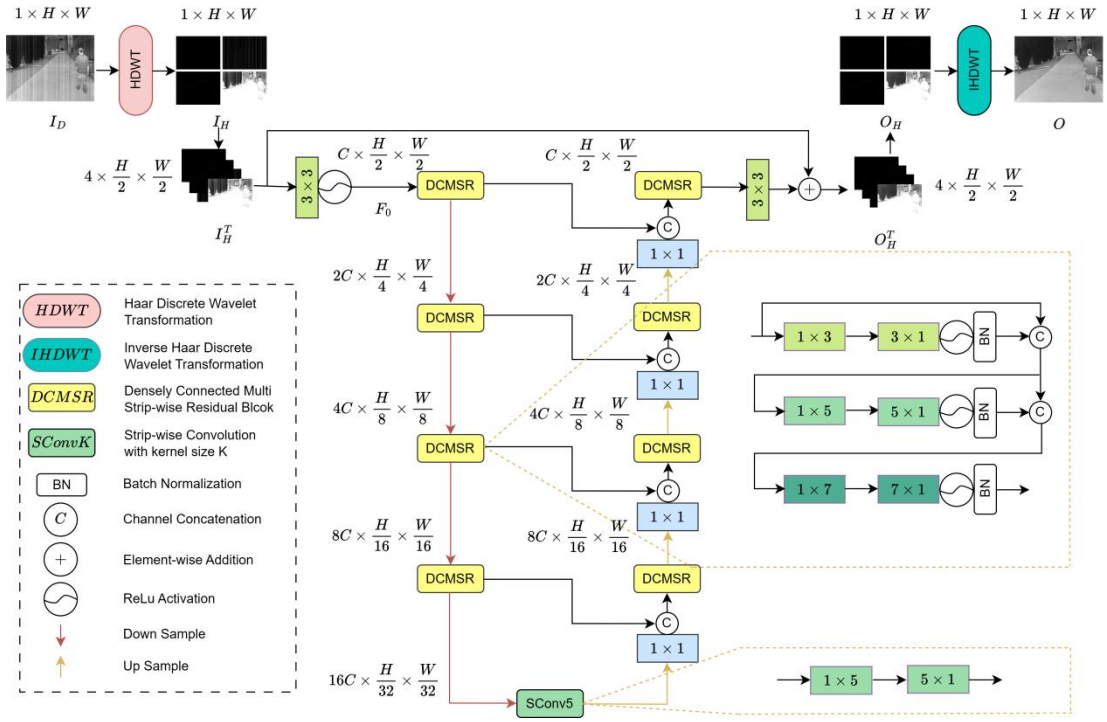

Figure 1 This is the overall architecture of our proposed method. We introduce the wavelet transform into our input and output of the U-Net, which can reduce the image size to half. Notably, the feature extraction module of our method is composed by these strip convolution layers with multiple kernel sizes in a dense-connected way.

We then explain how HDWT works in the input process module. When 2D DWT is introduced to process an image, we inevitably design four filters: a low-pass filter λ_{ll} and three high-pass filters $\lambda_{lh}, \lambda_{hl}, \lambda_{hh}$ to decompose the input image into an image with four sub-band. Concretely, Haar wavelet is defined as follow:

$$\lambda_{ll} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \lambda_{lh} = \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}, \\ \lambda_{hl} = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}, \lambda_{hh} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad (1)$$

The operation of HDWT is defined as:

$$I_H^T = \lambda \otimes I_D \downarrow 2, \lambda \in \{\lambda_{ll}, \lambda_{lh}, \lambda_{hl}, \lambda_{hh}\} \quad (2)$$

where $\otimes$ denotes the convolution operator, and $\downarrow 2$ represents downsampling with stride of 2.

2.3 Strip Convolution

The strip convolution block is stacking by a horizontal strip convolution layer and a vertical one, while the tradition convolution block is a convolution layer with normal square kernel. In the convolution process, we can clearly notice that the above structure is fit for extracting stripe information. Input image I_D is filtered by HDWT and transformed into I_H^T. The sub-band of I_H^T filtered by λ_{hl} has more clear information about stripe. If using normal convolution, much stripe structural information would be destroyed. Stripe noise can be characterized by the variations between columns. Thus, we choose to stack strip convolution layers like Figure 2. This is an illustration of how strip convolution layer works. Compared with the normal convolution layer, the strip convolution layer has less parameters and are more convenient for stripe extraction.

This type of strip convolution block has advantages: 1) reduce the parameters, 2) reduce the complexity, which consists of one horizontal strip convolution layer and one vertical strip convolution layer, denoted as:

$$\begin{cases} \text{Conv}_c^k(\mathbf{F}) = f^{k\times 1}(\mathbf{F}), \\ \text{Conv}_r^k(\mathbf{F}) = f^{1\times k}(\mathbf{F}), \\ \text{SConv}^k(\mathbf{F}) = \text{Conv}_c^k\left(\text{Conv}_r^k(\mathbf{F})\right) \end{cases} \quad (3)$$

Reduce the Parameters: For a normal convolution layer with square kernel, we define its kernel size as $k \times k$, while the number of input feature map channels is C_{in} and the number of output feature map channels is C_{out}. Consequently, the total parameters of a normal convolution layer are $C_{out} \times k \times k \times C_{in}$. For a strip convolution layer with strip kernel, we define its kernel size as $k_s \times 1, s \in \{h, w\}$, depends on the strip convolution direction. Similarly, C_{in}/C_{out} means the number of input/output feature map channels. Then, the total parameters of a strip convolution layer are $C_{out} \times k_s \times C_{in}$. In order to calculate conveniently, we assume the value of k_s is k. Therefore, compared with a normal convolution layer, a strip convolution layer can reduce parameters by $1/k$.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-383-2026 | © Author(s) 2026. CC BY 4.0 License.

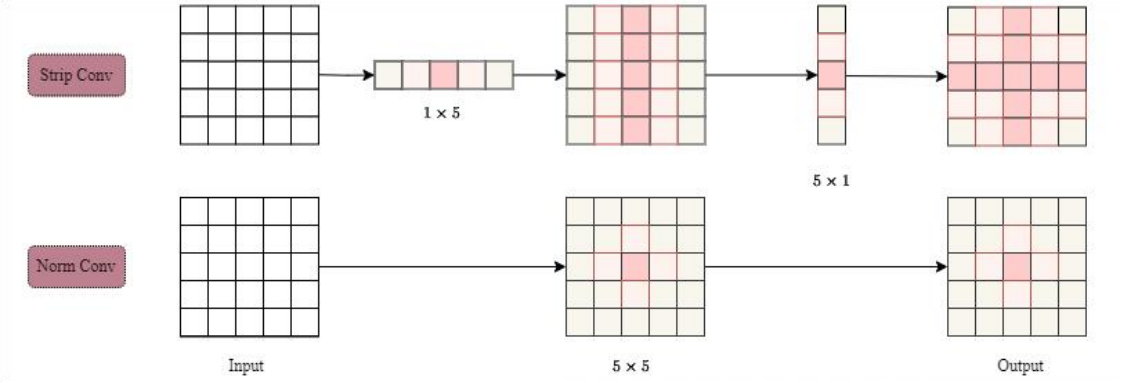

Figure 2 This is an illustration of how strip convolution layer works. Compared with the normal convolution layer, the strip convolution layer has less parameters and are more convenient for stripe extraction

Reduce the Complexity: For every square kernel, computing an element of the output feature map needs $k \times k$ multiplication and $k \times k - 1$ addition. Typically, the assessment of a model's complexity primarily focuses on multiplication operations, as these are regarded as more computationally intensive compared to other operations. As a result, the calculation complexity of a normal convolution layer is $C_{out} \times C_{in} \times k \times k \times O_h \times O_w$, where $O_h \times O_w$ means the output feature map size. Under the same assumption of strip convolution kernel size, the calculation complexity of a strip convolution layer is $C_{out} \times C_{in} \times k \times k \times O_h \times O_w$, which is reduced by $1/k$.

Feature Representation: In the Figure 2, we denote the $m \times n$ as the kernel size. Specifically, strip convolution kernel size is denoted as 1×5 and 5×1, while normal convolution kernel size is represented as 5×5. It is evident that the strip convolution kernel illustrated in the figure is the central row or column of the normal convolution kernel. Analyzing the output feature map generated by the strip convolution reveals that it effectively captures information within the criss-cross region surrounding the specified point. If we assume that this point lies within a stripe, it becomes apparent that the strip convolution possesses a distinct capability to accurately identify the stripe, a task that the normal convolution fails to accomplish.

2.4 Densely Connected Multi Scale Module

As a feature extraction module, SConv^k is activated by a sigmoid layer and then normalized by a batch normalization layer, named as SConvB^k:

$$\mathrm{SConvB}^k(\mathrm{F}) = \mathrm{BN}\left(\sigma\left(\mathrm{SConv}^k(\mathrm{F})\right)\right) \tag{4}$$

where BN means batch normalization and σ means the activation function.

In order to have a better local information aggregation, DCMS is set by three kinds of kernel size to stack these strip convolution layers in a densely connected way, which is demonstrated as below:

$$\mathrm{DCMS} = \mathrm{M}\left(\mathrm{SConvB}^{k_1}, \mathrm{SConvB}^{k_2}, \mathrm{SConvB}^{k_3}\right)(\mathrm{F}), k \in \{3,5,7\} \tag{5}$$

where $\mathrm{M}(*)$ represent the densely connected convolution aggregation and symbol $*$ means the aggregation of three strip convolution layers which can be described as follow:

$$\begin{cases} \mathbf{F}_{01} = \mathrm{Concat}\left(\mathrm{F}, \mathrm{SConvB}^3(\mathbf{F})\right), \\ \mathbf{F}_{12} = \mathrm{Concat}\left(\mathrm{F}_{01}, \mathrm{SConvB}^5(\mathbf{F}_{01})\right), \\ \mathrm{DCMS} = \mathrm{M}(*)(\mathbf{F}) = \mathbf{F}_{23} = \mathrm{Concat}\left(\mathbf{F}_{12}, \mathrm{SConvB}^7(\mathbf{F}_{12})\right) \end{cases} \tag{6}$$

2.5 Loss Function

We utilized L2 Loss, a commonly employed metric in image generation tasks, to assess the performance of the model. And then we add a constraint ε to promote convergence(Lai et al., 2018). Therefore, the final loss which is also known as Charbonnier loss, is articulated as follows:

$$L_{char} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\|\mathbf{O}_H^T(i) - \mathbf{I}_H^T(i)\|^2 + \varepsilon^2} \tag{7}$$

3. Experiments

To evaluate the effectiveness of the model, we conducted a series of experiments on the infrared image stripe removal task. We first delineate the experimental settings, after which we present the model's performance on the specified dataset. Following this, we carry out a series of ablation studies and downstream tasks.

3.1 Experiments Settings

Implementation Details: We conduct a series of comparison experiments on the platform with a single NVIDIA RTX 4080 GPU, an INTEL i7-12700 CPU and a 32GB RAM. Meanwhile, we train the model by setting the batch size and epoch number as 8 and 50 respectively. We optimize the model by ADAM optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The initial learning rate was set at 0.002, which is subsequently reduced at a rate of 0.0001 through the implementation of the exponential decay method with a warm-up strategy.

Training Datasets: The training images are collected from the CVC-09 dataset , which consists of LWIR thermal images captured in real-world scenarios. To simulate stripe noise, we generate synthetic degraded images by adding Gaussian noise with noise level $\sigma \in (0,0.2)$ and structured stripe patterns. The stripe noise is designed to mimic non-uniformity effects commonly observed in infrared sensors. Additionally, we implement data augmentation, including flipping and rotation. Following that, each image is paired with its corresponding corrupted image, and those pairs culminate in a comprehensive dataset consisting of 14528 pairs of clean and noisy images.

Baseline and Metrics: To assess the efficacy of our proposed method, we conduct a comparative analysis against leading techniques in the image stripe removal domain. This comparison encompasses traditional methods and learning-based approaches. The traditional methods evaluated include GF (Cao et al., 2016), LRSID (Chang et al., 2016) and SEID (Song and Huang, 2023). The learning-based methods examined consist of DLS-NUC (He et al., 2018), SNRWDNN (Guan et al., 2019), TSWEU(Chang et al., 2020), DMRN (Chang et al., 2019). For the evaluation of our method, we utilize both simulated and real stripe images. In the case of simulated images, we compute the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index Measure (SSIM). For real stripe images, we assess the performance using the Surface Roughness Index ρ and the Naturalness Image Quality Evaluator (NIQE).

3.2 Main Result on Synthesized Image

We randomly select 50 images containing pedestrian from the CVC-09 dataset as the test dataset. And for the image synthesis, we add three different types of noise with different noise level, gaussian distributed noise with level $\sigma \in [0.1, 0.2]$, normal distributed noise with level $\mu \in [25, 50]$ and periodical distributed noise with level $\mu \in [25, 50]$ and period $pr \in [10, 20]$. As can be seen from the Table 1 our method outperforms these state-of-the-art (SOTA) methods.

Quantitative Results: We conduct experimental evaluations of the SOTA methods on the aforementioned datasets in the context of the single-frame stripe removal task.
In Table 1, we have emphasized the methods achieving the highest and second-highest PSNR and SSIM scores by using bold and underlined formatting, respectively. Specifically, on our method surpasses DMRN, which is the second-best performing method, by margins ranging from 5 to 7 points. These results clearly demonstrate the superior efficacy of our method in comparison to other SOTA approaches.

Besides, our research has yielded additional insights: the SEID method is specifically designed for hyperspectral sequence images stripe removal, and its application to single-frame image stripe removal inherently limits its access to sequence information, which may be perceived as a disadvantage.

Surprisingly, traditional methods have demonstrated superior performance relative to certain deep learning approaches. For instance, the GF method outperform deep learning-based models such as DLS-NUC, DMRN, and TSWEU. Furthermore, the LRSID method surpasses all the other deep learning model. But it cannot be neglected regarding its large demand of time consumption which we reveal in the Table 3.

Visual Performance: We randomly select several images to display the synthesized stripe removal performance. These images are specifically Pedestrian, Motor, Rear Window, and Signboard, which are all depicted in Figure 3. It is apparent that various methods, have disparity throughout the entire image after removing stripes, while our method does not have such limitation.

Column Relevance: The efficacy of the stripe removal technique can be illustrated from an alternative perspective. By calculating the mean value of all pixels within each column of the image after stripe removal, we can designate this as the observed value, while the mean value of all pixels in each column of the original image serves as the true value. To elucidate this outcome, we quantify the differences between the true and observed values. This quantification involves calculating the absolute differences in column-wise pixel means, followed by averaging these differences across the row direction. This calculation aligns with the mathematical definition of Mean Absolute Error (MAE). The detailed definition of this metric is as follows:

$$\mathrm{MAE} = \frac{1}{mn}\sum_{j=0}^{n}\sum_{i=0}^{m}\left|p_{ij} - \hat{p}_{ij}\right| \tag{8}$$

where i denotes the column-wise pixel position, j denotes the row-wise pixel position, m represents the number of column-wise pixels, and n represents the number of row-wise pixels. As can be observed in the Table 2, the results of pixel mean differences for our method are the most optimal in Pedestrian, Motor, Sidewalk, and Car.

3.3 Main Results on Real Data

To assess the efficacy of the stripe removal technique in a practical environment, we present the results on from the ADMIRE dataset (Tendero and Gilles, 2012), which consists of LWIR infrared images affected by stripe noise caused by detector non-uniformity. As can be seen from the Table 5, our method outperforms SOTA methods.

Quantitative Results: Table 5 illustrates the quantitative metrics associated with this dataset, indicating that our approach demonstrates better results. It is noteworthy that our approach exhibits performance that closely aligns with LRSID and TSWEU, especially concerning the metrics ρ and NIQE. However, our approach exhibits reduced computational time as opposed to these two well-matched peers, and we will now offer an in-depth explanation below.

Visual Performance: We randomly select several images to display the real-world stripe removal performance. These images are specifically Building Window, Pedestrian, Car, and Rear Window, which are all depicted in Figure 4.

Time Consumption: Furthermore, we evaluate the computational time consumption of these methods, with the findings presented in Table 3, where Time/s denotes the average inference time per image. The inference time is measured on a single NVIDIA RTX 4080 GPU using images of resolution 640×480.The results indicate that our method remains highly competitive in terms of efficiency. the time consumption of LRSID is significantly greater than that of the other techniques.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-383-2026 | © Author(s) 2026. CC BY 4.0 License.

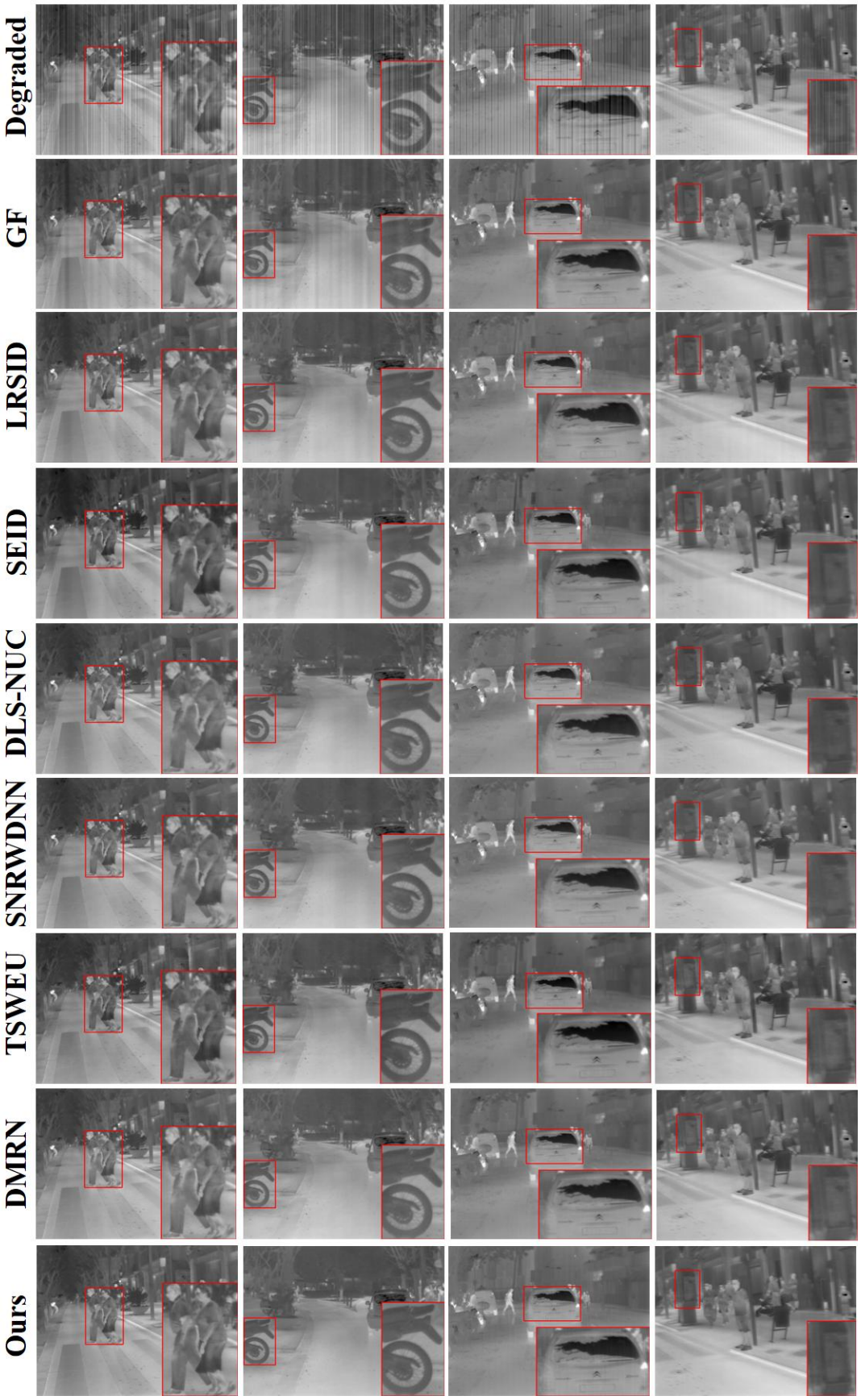

Figure 3 This is the visual performance of SOTA methods on synthesized images of CVC-09. We zoom in with red rectangles and the details in the pictures are exposed.

Category	Level	Index	GF	LRSID	SEID	DLS-NUC	SNRWDNN	TSWEU	DMRN	Ours
Gaussian	0.1	PSNR ↑	30.7243	32.8213	28.5728	29.6116	32.4068	30.5944	29.8537	**36.1363**
		SSIM ↑	0.8926	0.9614	0.8756	0.8813	0.9600	0.9730	0.9136	**0.9795**
	0.2	PSNR ↑	22.9413	26.5979	22.0361	20.1666	27.0991	20.4824	17.1623	**30.6635**
		SSIM ↑	0.5004	0.8874	0.7096	0.4405	0.9083	0.5792	0.1980	**0.9424**
Uniform	25	PSNR ↑	35.5154	35.4655	34.7263	38.7038	36.6650	30.2848	33.2073	**41.1306**
		SSIM ↑	0.9630	0.9691	0.9605	0.9715	0.9729	0.9667	0.9752	**0.9854**
	50	PSNR ↑	29.5668	31.8792	29.6805	33.0014	32.2104	33.1312	39.5435	**37.9987**
		SSIM ↑	0.8589	0.9576	0.8985	0.9131	0.9525	0.9749	0.9842	**0.9780**
Periodical	25,10	PSNR ↑	37.4736	34.4119	32.4390	37.2593	32.3276	29.6857	34.6268	**36.5205**
		SSIM ↑	0.9831	0.9699	0.9604	0.9714	0.9786	0.9662	0.9765	**0.9866**
	25,20	PSNR ↑	36.6603	35.1264	34.5948	37.8007	34.3245	30.3869	35.0627	**41.4424**
		SSIM ↑	0.9638	0.9697	0.9616	0.9714	0.9732	0.9672	0.9764	**0.9856**
Periodical	25,20	PSNR ↑	36.6603	35.1264	34.5948	37.8007	34.3245	30.3869	35.0627	**41.4424**
		SSIM ↑	0.9638	0.9697	0.9616	0.9714	0.9732	0.9672	0.9764	**0.9856**
	50,20	PSNR ↑	30.8129	32.7238	32.5851	32.5703	28.9865	32.8511	40.0103	**37.8857**
		SSIM ↑	0.8646	0.9662	0.9575	0.8963	0.9551	0.9751	0.9857	**0.9700**

Table 1 The average PSNR (dB) and SSIM of SOTA methods on four different test sets.

Image	Index	GF	LRSID	SEID	DLS-NUC	SNRWDNN	TSWEU	DMRN	Ours
Pedestrian	MAE↓	9.6090	5.7245	9.6780	11.0284	6.2269	9.8917	14.5268	**4.1659**
Motor	MAE↓	3.1045	2.2262	2.2681	1.4431	2.4543	4.7684	3.9032	**1.1434**
Sidewalk	MAE↓	2.6727	2.1935	1.9781	2.9148	5.3591	2.7652	2.7725	**1.3206**
Car	MAE↓	1.1589	1.2026	10.9617	2.2926	3.7528	6.6948	2.7243	**0.8819**

Table 2 Mean Absolute Error of Images from Different Types of Noises.

4. Discussion

4.1 Ablation Study

As shown in Table 6, he optimization of model performance, starting from a baseline PSNR of 31.1481 dB, is systematically detailed in the accompanying table. Initially, the implementation of identity mapping between the input and output resulted in an enhancement of 0.1544 dB in PSNR. The application of Wavelet Transform further contribute to a significant increase of 2.061 dB in PSNR. To augment feature representation capabilities, we integrate the dense connection structure, which lead to a PSNR enhancement of 0.8032 dB. Subsequently, we evaluate the impact of various kernel sizes in strip convolution, which produce PSNR variations of -0.0682 dB, 0.5829 dB, and 0.7695 dB, respectively. Based on these results, we amalgamate strip convolution layers with multiple kernel sizes , replacing the original DC module. This modification brings an increase of 0.9068 dB and finally culminate in a final model that achieved a PSNR of 35.4125 dB.

4.2 Wavelet Transform

When comparing the model input that incorporates wavelet transform to the model input that does not, it is evident that the complexity of the model utilizing wavelet transform is significantly reduced as HDWT changes the image size. It is important to emphasize that the model retains its consistency apart from the input processing method. Thus, the number of parameters remains unchanged despite variation in image size. From the Table 4, we can see that the parameter is the same, but the FLOPs for model employing HDWT is 27.3×10^9, in contrast to 109.2×10^9 for the model that does not utilize HDWT. However, we can easily find from the Table 4 that the model with Wavelet Input is better than the model without Wavelet Input. Concretely, it is an increase of 2.061 dB in PSNR but accompanied with a decrease of 4 times computational resources.

Method	Type	Time/s↓
GF	Filter-based	**0.0831**
LRSID	Optimization-based	9.9388
SEID	Optimization-based	0.5918
DLS-NUC	Learning-based	0.2124
SNRWDNN	Learning-based	**0.0829**
TSWEU	Learning-based	1.4217
DMRN	Learning-based	0.5486
Ours	Learning-based	0.3946

Table 3 Time consumption of different methods.

	Params/M↓	FLOPs/G↓
w HDWT	21.6462	109.2506
w/o HDWT	21.6479	27.3504

Table 4 Computation Complexity of Model w/o HDWT.

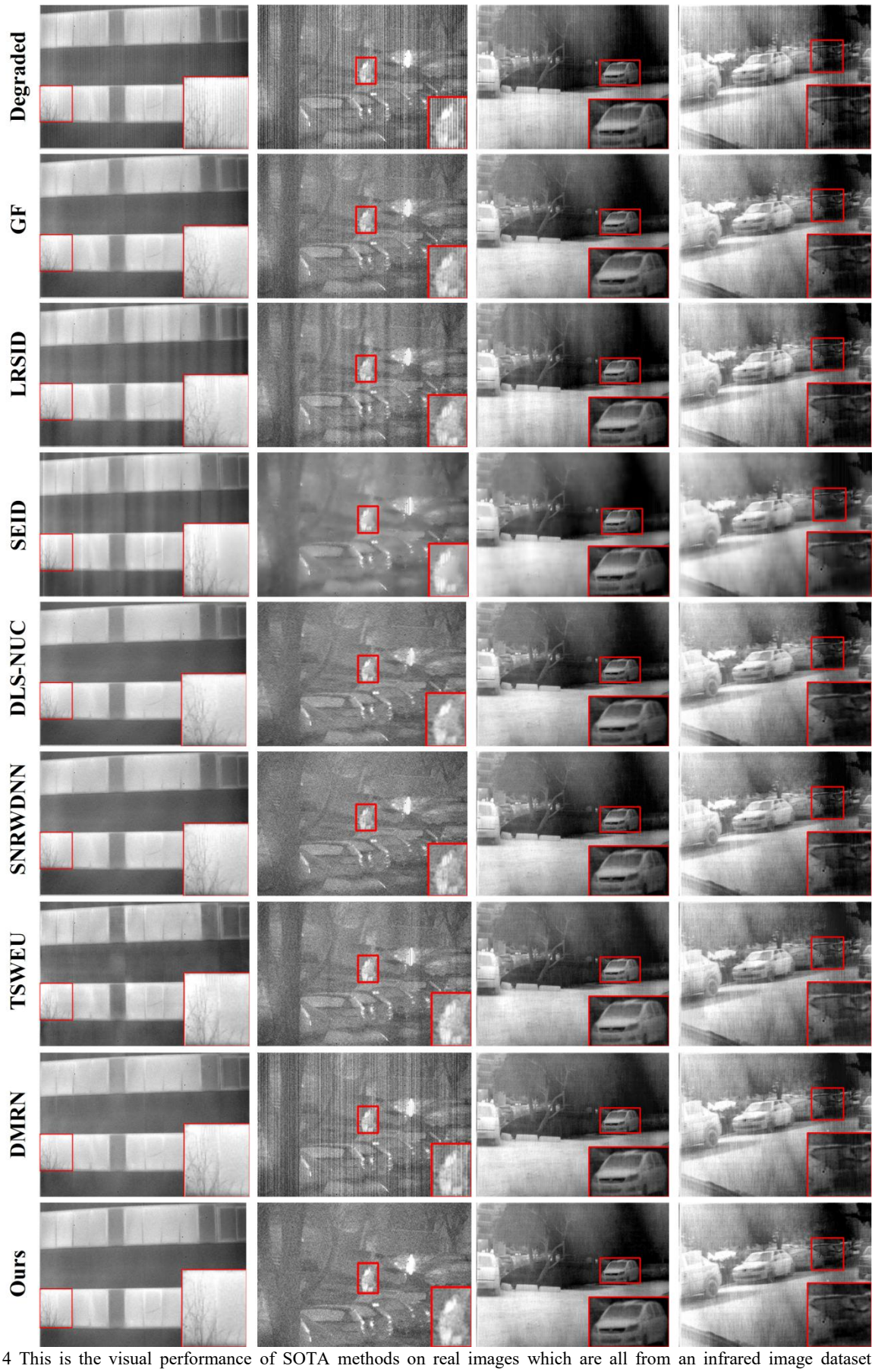

Figure 4 This is the visual performance of SOTA methods on real images which are all from an infrared image dataset called ADMIRE. We zoom in with red rectangles and the details in the pictures are exposed.

To incorporate strip convolution into our model, we gradually undertake a systematic enhancement of its design. Initially, we employ strip convolution for feature extraction, which we then subsequently augment the integration by the dense connection. We perform the ablation experiments for each variant and we observe a significant performance enhancement for our model associated with the dense connection of strip convolutions. Additionally, we introduce varying kernel sizes of convolution layer which can further aggregate different receptive fields. The results are illustrated through line charts, revealing that our method substantially improved the model's metrics with only a marginal increase in the number of parameters and computational demands relative to the baseline model.

Datasets	Index	GF	LRSID	SEID	DLS-NUC	SNRWDNN	TSWEU
ADMIRE	ρ↓	0.1980	0.1603	0.1621	0.1665	0.1658	0.1604
	NIQE↓	13.0830	10.8464	11.4819	12.9259	11.2061	11.4025

Table 5 The ρ and NIQE of different methods on test sets with real stripe noises.

U-Net	+Identiy Mapping	+Wavelet Input	+Dense Connection	+Strip Convolution	+Multi Scale	PSNR↑	SSIM↑
✓	×	×	×	×	×	31.1481	0.9631
✓	✓	×	×	×	×	31.3037	0.9678
✓	✓	✓	×	×	×	33.3647	0.9762
✓	✓	✓	✓	×	×	34.5057	0.9844
✓	✓	✓	✓	✓	×	34.4375	0.9840
✓	✓	✓	✓	✓	✓	35.4125	0.9887

Table 6 Ablation Study.

5. Conclusion

When signals are transmitted through infrared sensors, they are susceptible to generating FPN, which adversely impacts subsequent tasks. To address this prevalent issue, we propose an innovative integration of strip convolution into models to effectively reduce this noise. To further enhance the feature representation capabilities of strip convolution, we incorporate it into dense connection and fuse strip convolution layers with varying kernel sizes, which we designate as the DCMS Block. Additionally, we introduce HDWT technique into the input-output phase, which significantly reduces computational demands while markedly enhancing the model's performance. Based on these innovative concepts, we conduct comparative analyses with various SOTA methods to assess the efficacy of our approach in both synthetic and real images.

Acknowledgements

This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFB3902900, and in part by the Shanghai Aerospace Science and Technology Innovation Foundation under Grant SAST2023-046 as well as in part by the Fengyun Application Pioneering Project (FY-APP).

References

Cao, Y., Yang, M.Y., Tisse, C.-L., 2016. Effective Strip Noise Removal for Low-Textured Infrared Images Based on 1-D Guided Filtering. IEEE Transactions on Circuits and Systems for Video Technology 26, 2176 – 2188. https://doi.org/10.1109/TCSVT.2015.2493443.

Chang, Y., Chen, M., Yan, L., Zhao, X.-L., Li, Y., Zhong, S., 2020. Toward Universal Stripe Removal via Wavelet-Based Deep Convolutional Neural Network. IEEE Transactions on Geoscience and Remote Sensing 58, 2880 – 2897. https://doi.org/10.1109/TGRS.2019.2957153.

Chang, Y., Yan, L., Fang, H., Liu, H., 2014. Simultaneous Destriping and Denoising for Remote Sensing Images With Unidirectional Total Variation and Sparse Representation. IEEE Geoscience and Remote Sensing Letters 11, 1051 – 1055. https://doi.org/10.1109/LGRS.2013.2285124.

Chang, Y., Yan, L., Liu, L., Fang, H., Zhong, S., 2019. Infrared Aerothermal Nonuniform Correction via Deep Multiscale Residual Network. IEEE Geoscience and Remote Sensing Letters 16, 1120 – 1124. https://doi.org/10.1109/LGRS.2019.2893519.

Chang, Y., Yan, L., Wu, T., Zhong, S., 2016. Remote Sensing Image Stripe Noise Removal: From Image Decomposition Perspective. IEEE Transactions on Geoscience and Remote Sensing 54, 7018 – 7031. https://doi.org/10.1109/TGRS.2016.2594080.

Cui, Y., Knoll, A., 2024. Dual-domain strip attention for image restoration. Neural Networks 171, 429 – 439. https://doi.org/10.1016/j.neunet.2023.12.003.

Cui, Y., Knoll, A., 2023. Exploring the potential of channel interactions for image restoration. Knowledge-Based Systems 282, 111156. https://doi.org/10.1016/j.knosys.2023.111156.

Cui, Y., Tao, Y., Jing, L., Knoll, A., 2023. Strip Attention for Image Restoration. Presented at the Thirty-Second International Joint Conference on Artificial Intelligence, pp. 645 – 653. https://doi.org/10.24963/ijcai.2023/72.

CVC-09: FIR Sequence Pedestrian Dataset – Elektra, n.d. URL http://adas.cvc.uab.es/elektra/enigma-portfolio/item-1/ (accessed 6.3.25).

Fayyaz, Z., Platnick, D., Fayyaz, H., Farsad, N., 2022. Deep Unfolding for Iterative Stripe Noise Removal, in: 2022 International Joint Conference on Neural Networks (IJCNN). Presented at the 2022 International Joint Conference on Neural Networks (IJCNN), pp. 1 – 7. https://doi.org/10.1109/IJCNN55064.2022.9892708.

Guan, J., Lai, R., Li, H., Yang, Y., Gu, L., 2023. DnRCNN: Deep Recurrent Convolutional Neural Network for HSI Destriping. IEEE Transactions on Neural Networks and Learning Systems 34, 3255 – 3268. https://doi.org/10.1109/TNNLS.2022.3142425.

Guan, J., Lai, R., Xiong, A., 2019. Wavelet Deep Neural Network for Stripe Noise Removal. IEEE Access 7, 44544 – 44554. https://doi.org/10.1109/ACCESS.2019.2908720.

He, Z., Cao, Yanpeng, Dong, Y., Yang, J., Cao, Yanlong, Tisse, C.-L., 2018. Single-image-based nonuniformity correction of uncooled long-wave infrared detectors: a deep-learning approach. Appl. Opt., AO 57, D155 – D164. https://doi.org/10.1364/AO.57.00D155.

Hou, Q., Zhang, L., Cheng, M.-M., Feng, J., 2020. Strip Pooling: Rethinking Spatial Pooling for Scene Parsing, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4002 – 4011. https://doi.org/10.1109/CVPR42600.2020.00406.

Jiang, S., Zhi, X., Zhang, W., Wang, D., Hu, J., Chen, W., 2021. Remote sensing image fine-processing method based on the adaptive hyper-Laplacian prior. Optics and Lasers in Engineering 136, 106311. https://doi.org/10.1016/j.optlaseng.2020.106311.

Kuang, X., Sui, X., Chen, Q., Gu, G., 2017. Single Infrared Image Stripe Noise Removal Using Deep Convolutional Networks. IEEE Photonics Journal 9, 1 – 13. https://doi.org/10.1109/JPHOT.2017.2717948.

Kuang, X., Sui, X., Liu, Y., Liu, C., Chen, Q., Gu, G., 2018. Robust destriping method based on data-driven learning. Infrared Physics & Technology 94, 142 – 150. https://doi.org/10.1016/j.infrared.2018.09.015.

Lai, W., Huang, J., Ahuja, N., Yang, M.H., 2018. Fast and Accurate Image Super-Resolution with Deep Laplacian Pyramid Networks.

Li, J., Zeng, D., Zhang, J., Han, J., Mei, T., 2022a. Column-Spatial Correction Network for Remote Sensing Image Destriping. Remote Sensing 14, 3376. https://doi.org/10.3390/rs14143376.

Li, J., Zhang, J., Chen, F., Zhao, K., Zeng, D., 2022b. Adaptive Material Matching for Hyperspectral Imagery Destriping. IEEE Transactions on Geoscience and Remote Sensing 60, 1 – 20. https://doi.org/10.1109/TGRS.2022.3158901.

Ni, Z., Chen, X., Zhai, Y., Tang, Y., Wang, Y., 2024. Context-Guided Spatial Feature Reconstruction for Efficient Semantic Segmentation. https://doi.org/10.48550/arXiv.2405.06228.

Song, L., Huang, H., 2023. Simultaneous Destriping and Image Denoising Using a Nonparametric Model With the EM Algorithm. IEEE Transactions on Image Processing 32, 1065 – 1077. https://doi.org/10.1109/TIP.2023.3239193.

Tendero, Y., Gilles, J., 2012. ADMIRE: a locally adaptive single-image, non-uniformity correction and denoising algorithm: application to uncooled IR camera. pp. 83531O-83531O – 16. https://doi.org/10.1117/12.912966.

Wang, E., Jiang, P., Li, X., Cao, H., 2019. Infrared stripe correction algorithm based on wavelet decomposition and total variation-guided filtering. Journal of the European Optical Society-Rapid Publications 16, 1. https://doi.org/10.1186/s41476-019-0123-2.

Xu, K., Zhao, Y., Li, F., Xiang, W., 2022. Single infrared image stripe removal via deep multi-scale dense connection convolutional neural network. Infrared Physics & Technology 121, 104008. https://doi.org/10.1016/j.infrared.2021.104008.

Yan, F., Wu, S., Zhang, Q., Liu, Y., Sun, H., 2023. Destriping of Remote Sensing Images by an Optimized Variational Model. Sensors 23. https://doi.org/10.3390/s23177529.

Yi Chang, Luxin Yan, Houzhang Fang, Chunan Luo, 2015. Anisotropic Spectral-Spatial Total Variation Model for Multispectral Remote Sensing Image Destriping. IEEE Trans. on Image Process. 24, 1852 – 1866. https://doi.org/10.1109/TIP.2015.2404782.

Zhang, H., Cai, J., He, W., Shen, H., Zhang, L., 2022. Double Low-Rank Matrix Decomposition for Hyperspectral Image Denoising and Destriping. IEEE Transactions on Geoscience and Remote Sensing 60, 1 – 19. https://doi.org/10.1109/TGRS.2021.3061148.

Zhao, L., Ye, L., Zhang, M., Jiang, H., Yang, Z., Yang, M., 2023. DPSDA-Net: Dual-Path Convolutional Neural Network with Strip Dilated Attention Module for Road Extraction from High-Resolution Remote Sensing Images. Remote Sensing 15. https://doi.org/10.3390/rs15153741.

Zhao, Y.-Q., Yang, J., 2015. Hyperspectral Image Denoising via Sparse Representation and Low-Rank Constraint. IEEE Transactions on Geoscience and Remote Sensing 53, 296 – 308. https://doi.org/10.1109/TGRS.2014.2321557.

Zhou, Y., Zheng, X., Ouyang, W., Li, B., 2023. A Strip Dilated Convolutional Network for Semantic Segmentation. Neural Process Lett 55, 4439 – 4459. https://doi.org/10.1007/s11063-022-11048-5.