Mathematical Problems of Computer Science 62, 126–144, 2024.

doi:10.51408/1963-0126

UDC 004.932.4

Mamba-based Thermal Image Dehazing

Sargis A. Hovhannisyan

Yerevan State University, Yerevan, Armenia
e-mail: sargis.hovhannisyan@ysu.am

Abstract

Atmospheric phenomena such as rain, snow, urban, forest fires, and artificial disasters can degrade image quality across various applications, including transportation, driver assistance systems, surveillance, military, and remote sensing. Image dehazing techniques aim to reduce the effects of haze, dust, fog, and other atmospheric distortions, enhancing image quality for better performance in computer vision tasks. Haze not only obscures details but also reduces contrast and color fidelity, significantly impacting the accuracy of computer vision (CV) models used in object detection, image classification, and segmentation. While thermal infrared (TIR) imaging is often favored for long-range surveillance and remote sensing due to its resilience to haze, atmospheric conditions can still degrade TIR image quality, especially in extreme environments.

This paper introduces MTIE-Net, a novel Mamba-based network for enhancing thermal images degraded by atmospheric phenomena like haze and smoke. MTIE-Net leverages the Enhancement and Denoising State Space Model (EDSSM), which combines convolutional neural networks with state-space modeling for effective denoising and enhancement. We generate synthetic hazy images and employ domain-specific transformations tailored to thermal image characteristics to improve training in low-visibility conditions. Our key contributions include using the Mamba architecture with 2D Selective Scanning for thermal image enhancement, developing a specialized Enhancement and Denoising module, and creating a labeled thermal dataset simulating heavy haze. Evaluated on the M3DF dataset of long-range thermal images, MTIE-Net surpasses state-of-the-art methods in both quantitative metrics (PSNR, SSIM) and qualitative assessments of visual clarity and edge preservation. This advancement significantly improves the reliability and accuracy of critical systems used in remote sensing, surveillance, and autonomous operations by enhancing image quality in challenging environments.

Keywords: Thermal Image, Image Dehazing, Thermal Image Enhancement, Mamba.
Article info: Received 10 October 2024; sent for review 19 October 2024; accepted 26 November 2024.
Acknowledgments: This work was partly supported by the Advance Research Grants from the Foundation for Armenian Science and Technology, funded by Sarkis and Nune Sepetjians.
I would also like to acknowledge SOLARON LLC for providing the real-world solar panel images used in this study. Additionally, I thank Professor S. Agaian for his invaluable guidance and support throughout this project

1. Introduction

Images captured in hazy or foggy weather conditions often suffer from significant degradation, making it difficult for imaging systems to identify objects and their features. The reduction in contrast and color shift further complicates imaging applications. Haze causes reduced contrast, faint surfaces, color distortion, and blurred intensity, which diminishes visibility and object perception [1]. To address this, image dehazing algorithms improve scene clarity [2,3,4,5,6]. However, this task is challenging because haze formation depends on various factors like unknown scene depth, the density and size of atmospheric particles, and the wavelength of light [7]. The effect of haze varies across different parts of the electromagnetic spectrum: shorter wavelengths in the visible (VIS) are more affected, while longer wavelengths, such as those in the thermal infrared (TIR), are less affected, but factors like atmospheric absorption and scattering can degrade the quality and accuracy of thermal images, which is shown in Fig. 1. This degradation poses significant challenges for vision algorithms [8] such as pedestrian detection and segmentation [9], military target detection [10], surveillance and security [11], and remote sensing [12]. Despite its benefits, thermal imaging suffers from drawbacks like low contrast and blurred details, which limit the observation of infrared targets and hinder the development of infrared imaging applications. Thermal Image Enhancement (TIE) techniques aim to improve visibility, clarity, and overall image quality for automated processing applications [13]. These methods vary across fields and objectives, encompassing both traditional approaches and Convolution Neural Networks (CNN). Table 1 highlights the advantages and challenges of various thermal image enhancement methods, emphasizing the need for more robust techniques and exposing the need to develop more robust IIE techniques. Recent developments have introduced State Space Models (SSMs), especially the Mamba model [14], which captures global contextual information with a linearly lower complexity for input tokens. The Structured State-Space Sequence model (S4) [15] was the first to highlight the potential of SSMs, offering a novel alternative to CNNs and Transformers for handling long-range dependencies. Followed by the S5 layer [16], which introduced MIMO SSM and efficient parallel scanning. SSM-based architectures have gained significant attention in various fields [17]. For example, Mamba4KT [18] is specifically developed for knowledge tracing in intelligent education systems, utilizing the Mamba model to effectively capture long-term relationships between exercises and students' knowledge levels. The latest Mamba architecture, Mamba-2 [19], introduces an enhanced core layer within the Mamba selective SSM, achieving speeds 2 to 8 times faster while maintaining competitive performance with Transformers in language modeling.

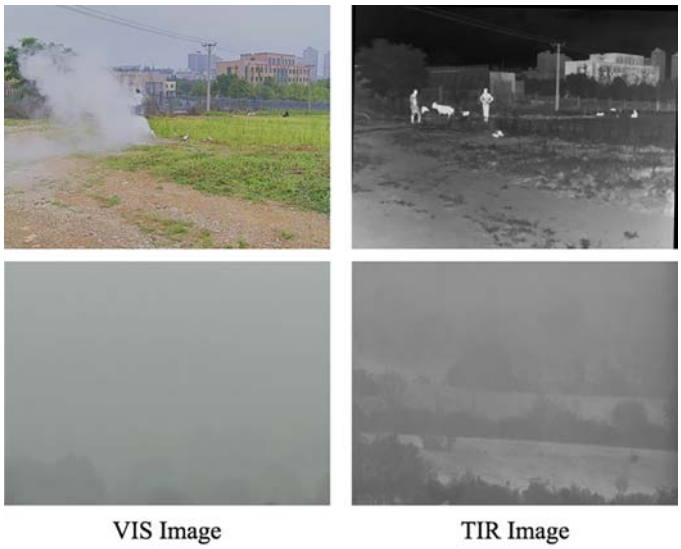

Fig. 1. Haze visibility is shown across two spectra. Heavy fog with larger particles affects both spectra compare with smoke in the first row [36, 32].

This paper introduces the **Mamba-Based Thermal Image Enhancement Network (MTIE-Net),** a novel approach for addressing thermal image dehazing. **MTIE-Net incorporates an Enhancement and Denoising State Space Model (EDSSM),** which integrates convolutional neural networks (CNNs) with state-space modeling to denoise and enhance thermal images effectively. Additionally, the framework utilizes advanced techniques, including synthetic data generation using the Atmospheric Scattering Model (ASM) and domain-specific transformations

tailored to thermal image characteristics. These enhancements improve the model's robustness and accuracy under severe haze conditions. Our main contributions are:

1. **Introducing Mamba for thermal image enhancement** for the first time, utilizing SS2D as the computational backbone to ensure efficient and scalable dehazing.
2. **Proposing an Enhancement and Denoising (ED) module** that optimizes image enhancement and noise reduction processes, effectively addressing challenges posed by atmospheric disturbances, which is its backbone to achieve linear computational efficiency. In other words, this makes the dehazing process faster and more effective, particularly for long-range thermal imaging tasks.
3. **Creating a labeled thermal dataset** simulating heavy haze conditions will enable more realistic training and testing of dehazing models.

Table 1. Advantages and challenges of various thermal image enhancement methods.

Method	Advantages	Challenges
Histogram Equalization [22]	- Enhances global contrast, is simple, and is fast to implement.	- May lead to noise amplification, it can cause an unnatural appearance and loss of detail
Contrast Adjustment [37]	- Flexible control over brightness and contrast. Can enhance specific regions	- May not work well with images with uneven illumination. Risk of over/under-enhancement Can cause loss of details in smooth regions.
Wavelet-based Methods [23]	- Effective in multi-scale analysis, good at edge and detail preservation	- It requires parameter selection/tuning and can be computationally intensive. If not properly managed, it may introduce artifacts.
Top-hat Transform Morphological Operation [24]	- Useful for highlighting small features. Effective in background suppression	- May not perform well with complex backgrounds, sensitive to structuring element selection and noise
Gradient Field Equalization [25]	- Enhances edges and local contrast. Suitable for highlighting object boundaries	- May introduce artifacts and noise. It can be sensitive to high-frequency noise. Computational intensity: making real-time processing challenging.
IE-CGAN [42]	- Produces visually appealing results with higher contrast and details	- It requires large datasets and computational resources, potentially overfitting and artifacts. Tends to produce unnatural sharpness in degraded images.
BBCNN [41]	- Maintain a natural scene appearance	- It struggles in lighter areas and with subtle contrast differences, obscuring key details.

To validate the effectiveness of our proposed MTIE-Net, we conducted extensive evaluations using the **M3DF dataset**, which consists of long-range thermal infrared (TIR) images affected by varying degrees of atmospheric distortion. Experimental results show that MTIE-Net outperforms existing state-of-the-art methods in quantitative metrics, such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) [20], and qualitative measures, including visual clarity and edge preservation.

The M**TIE-Net** framework advances the field of thermal image dehazing by providing a scalable and highly effective solution for improving image quality in real-world applications such as surveillance, autonomous driving, and environmental monitoring. These performance

improvements have broad practical implications: MTIE-Net can enhance the reliability and accuracy of critical systems, contributing to better overall performance and safety, whether in enhancing surveillance accuracy or improving the safety of autonomous vehicles.

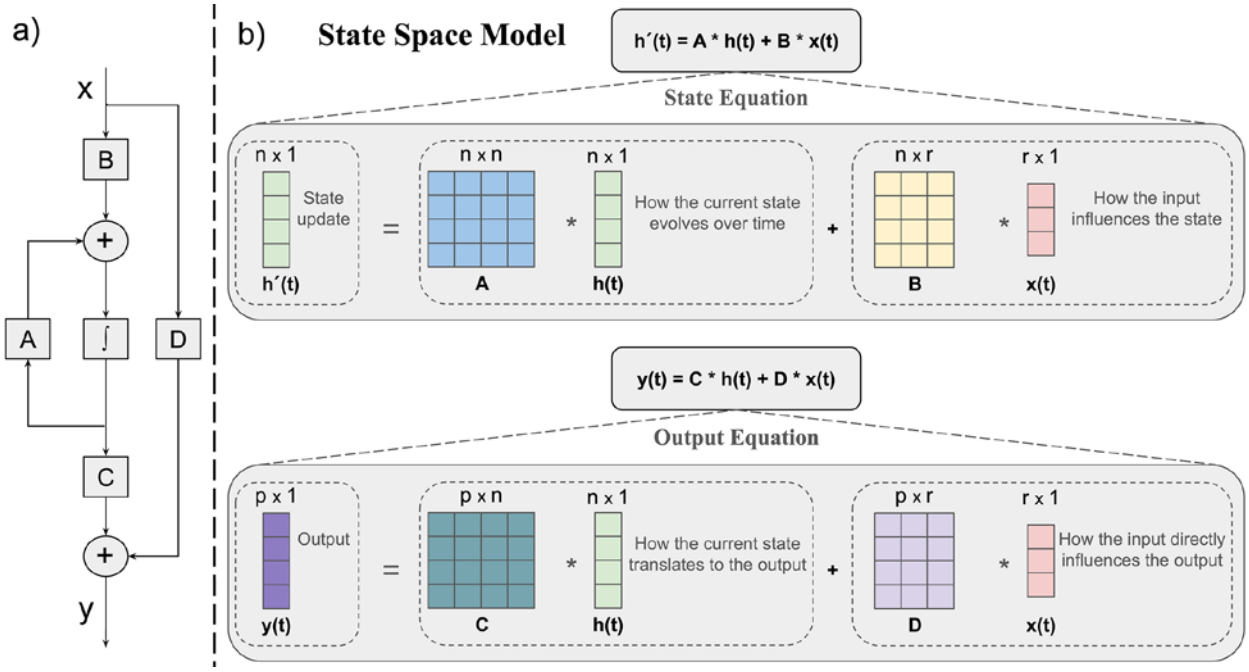

Fig. 2. Mathematical expression of the State-Space Model.

2. Background

2.1 Enhancing Thermal Imaging: Challenges and Advances

Thermal imaging cameras convert thermal energy (heat) into visible light, typically in grayscale or color scales. However, the spatial resolution of thermal images is often limited by diffraction effects in thermo-reflectance, and the signal-to-noise ratio (SNR) can affect image quality and application. Recent advancements aim to enhance thermal image quality for various applications, including geoscience, resource exploration, military surveillance, astronomy, and humanitarian missions, all of which demand high-quality, accurate thermal images. Creating high-resolution thermal images is particularly challenging for thermal cameras used on satellites or aerial platforms. These challenges include low dynamic range, lack of detail clarity, and blurred edges. Additionally, thermal images often suffer from low luminance due to factors like heat radiation intensity, object distance, and reflection angles, which can degrade the performance of vision-based systems.

Thermal image enhancement algorithms fall into two main categories: traditional and learning-based. Traditional methods [21], see Table 1, such as Histogram Equalization [22], Wavelet-Based methods [23], Top-Hat Transform [24], and Gradient Field Equalization [25], focus on improving contrast and reducing noise. While effective in some scenarios, they often struggle in complex situations, sometimes over-enhance, leading to noise amplification and brightness distortion.

In contrast, learning-based methods utilize neural networks to enhance image quality, effectively addressing issues like low contrast, noise, and blurred details, making thermal images more suitable for analysis. These methods also face challenges, including high computational demands, reliance on large datasets, training instability, risk of overfitting, domain adaptation difficulties,

noise sensitivity, and limited real-time processing capabilities. Ongoing research aims to mitigate these issues, improving the robustness and practicality of learning-based approaches.

2.2 State Space Models (SSM)

A State Space contains the minimum number of variables describing a system. SSMs were initially developed in control theory [26] to model dynamic systems and have since been adapted into deep learning for their ability to handle sequential data effectively [17, 27]. In control theory, SSMs represent systems where the current state is influenced by prior states and external inputs, making them well-suited for capturing temporal dynamics. This framework allows for real-time tracking of system evolution over time, which is essential in dynamic processes. When SSMs were introduced into deep learning, they brought a distinct advantage: their powerful capability to model long-range dependencies in sequential data, outperforming traditional recurrent architectures like RNNs and LSTMs [28]. Unlike these traditional models, which often struggle with vanishing gradients and computational inefficiency, particularly in long sequences, SSMs are designed to maintain linear computational complexity. This makes SSMs far more efficient, allowing for the processing of long sequences without the significant computational cost typically incurred by other architectures. As a result, SSMs can handle both short-term and long-term dependencies in the data, making them highly versatile.

State-Space Models for Linear Systems: SSMs provide a mathematical framework for representing physical systems using inputs, outputs, state variables, and differential equations. An SSM dynamically describes a system's behavior and is constructed using two types of equations: the state equation and the output (or observation) equation.

- The **state equation** defines the temporal evolution of the system's state as a dynamical system.
- The **output equation** defines how the internal state is observed or measured through outputs.

The SSM order, or the number of differential equations required to represent a physical model, depends on the system's input and output variables. SSM-based control, a fundamental tool, is crucial in analyzing linear and non-linear systems with multiple inputs and outputs. Fig. 2(a) illustrates a block diagram of a state-space model for a linear system with a feedback control loop connected to the inputs and outputs. This model dynamically describes the system's behavior using state variables $h(t)$, inputs $x(t)$, and outputs y(t). The following equations govern the state-space model:

$$\textbf{State Equation: } h'(t) = A \cdot h(t) + B \cdot \mathrm{x(t)}$$

This equation defines how the state evolves over time through the state transition matrix A, which captures the influence of the current state on the next state. Matrix B describes the impact of the input x(t) on the state changes. Here, h(t) represents the latent state at time t, and x(t) is the input at the same time step. This equation demonstrates how the system changes based on both its current state and the input it receives.

$$\textbf{Output Equation: } y(t) = C \cdot h(t) + D \cdot x(t)$$

The output equation converts the state into an observable output. Matrix C controls how the latent state h(t) is translated into the output y(t), while matrix D determines the direct influence of the input x(t) on the output. Matrices A, B, C, and D are all learnable parameters that can be adjusted to optimize the model's predictive accuracy. Figure 2(b) provides a visualization of both the state and output equations. It shows how the state equation governs state evolution, with matrix A capturing the influence of the current state on the future state and matrix B capturing how the input x(t) affects state transitions.

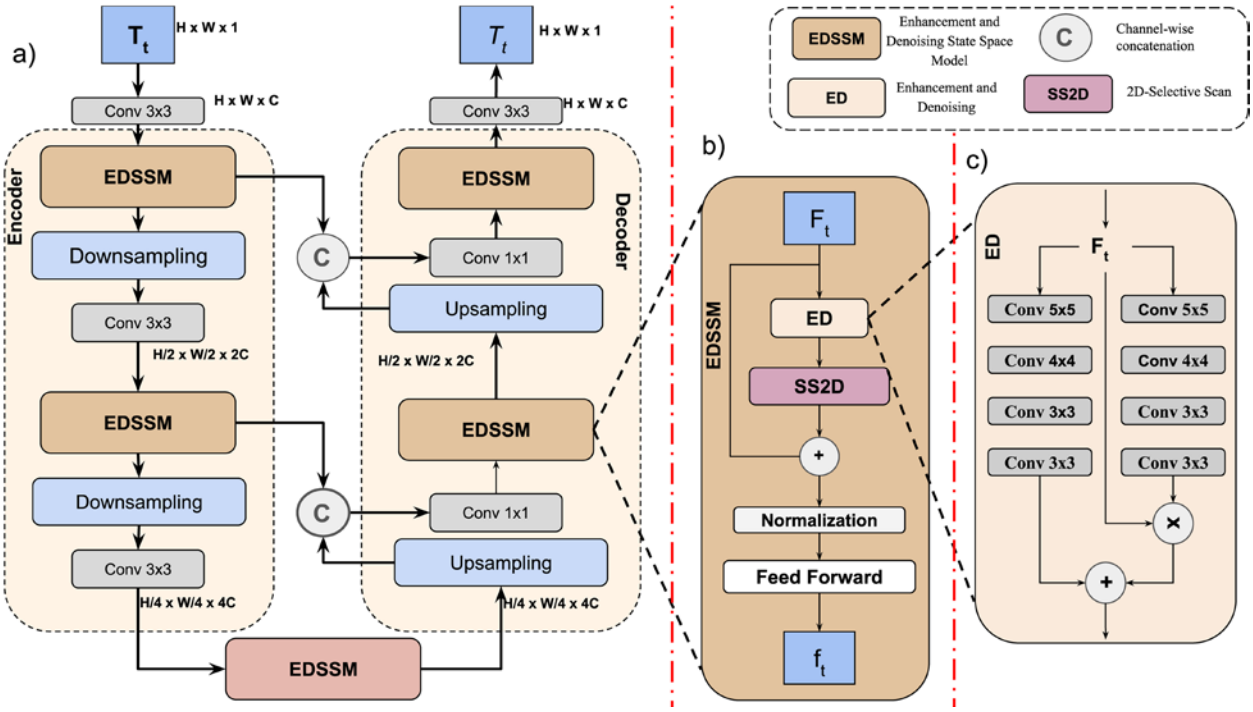

Fig. 3. Overall architecture of MTIE-Net (a), Enhancement and Denoising State Space Model (b), Enhancement and Denoising Module (c).

Solving these equations aims to uncover the statistical relationships that predict the system's future state based on the observed data (input sequences) and the previous state. This approach is fundamental in control theory and signal processing, allowing for the modeling and analysis of complex dynamic systems.

The efficiency and scalability of SSMs have drawn significant attention in areas that require analyzing sequential data, such as natural language processing, time-series forecasting, and, increasingly, computer vision tasks. In particular, SSMs have proven valuable in tasks that involve long-range dependencies, such as video analysis and image enhancement techniques like dehazing. By leveraging the strengths of SSMs, modern deep learning architectures can more effectively model both spatial and temporal relationships, improving performance in tasks where temporal and spatial consistency is critical. This ability to capture complex dependencies with high efficiency positions SSMs as a key technology in advancing deep learning capabilities in both time-dependent and spatially structured data domains.

2.3 Visual Mamba

The Mamba model, a dynamic state space model (SSM) with efficient selection mechanisms, is gaining traction in computer vision due to its ability to handle long-range dependencies in data while maintaining linear complexity. This significantly contrasts with traditional transformers,

which face limitations due to their quadratic complexity as image sizes increase. Mamba has demonstrated promising results in various visual tasks, including image classification, feature enhancement, and multimodal fusion [29]. Its efficiency positions it as a strong contender to replace CNNs and transformers as a foundational architecture for visual tasks.

Fig. 4. This figure shows original image from M3DF, corresponding VIS image, depth map and generated hazy image.

Since 2024, Mamba has been successfully applied to diverse applications such as image enhancement, video analysis, and object detection [30]. These applications showcase their versatility and potential to significantly improve computer vision systems' accuracy and efficiency. By effectively managing long-range dependencies and maintaining computational efficiency, Mamba offers a robust solution for modern visual tasks, paving the way for advancements in the field. We adopted the integration of SSM into visual tasks, following the approach outlined in [31]. The SS2D module consists of three main operations: Scan Expanding, S6 blocks, and Scan Merging. In our model, the input images first undergo the Scan Expanding operation, systematically unfolding the image from its four corners toward the center. This process rearranges the spatial structure of the image, effectively allowing the model to capture features from different spatial regions. The height and width of the image are then flattened into a token length, turning the two-dimensional image into a one-dimensional sequence. Once the image has been flattened, each sequence is input into the S6 module, responsible for feature extraction. The S6 module operations can be expressed as:

$$h_t = Ah_{t-1} + Bx_t,$$

$$y_t = Ch_t,$$

where, x is the input variable, y is the output, and A, B and C are all learnable parameters. The outputs from the four directions of extracted features are summed and merged, and the dimensions of the merged output are readjusted to match the input size. Then, after processing the S6 blocks,

the Scan Merging operation restores the spatial structure by reorganizing the flattened sequence back into its original two-dimensional form. This combination of scan operations enables the SS2D module to capture local and global features in the image effectively, improving feature extraction for our visual tasks.

Fig. 5. This figure shows original image from M3DF, generated hazy image and generated label image.

3. Network Structure and Training

This section describes the proposed network and its training strategy for thermal image enhancement.

3.1 Structure

Fig. 3(a) illustrates our MTIE-Net framework, which consists of an encoder-decoder structure, capturing both local and long-range contextual features. Both processes are symmetric and divided into two levels. Each downsampling level consists of an EDSSM (see Fig. 3(b)), and a convolutional layer with kernel size of 3 × 3 (stride=2). Similarly, upsampling involves two levels, each including a upsampling operation, a 1 × 1 convolution applied to the merged features from the corresponding downsampling layer, and an EDSSM. Finally, a 3 × 3 convolution is applied to the image to reduce dimensionality and restore it to grayscale with a single channel. EDSSM block includes an Enhancement and Denoising (ED) module (see Fig. 3(c)) consisting of two branch processing: the first branch outputs a single-channel feature map, which is element-wise multiplied with the input image, enhancing contrast and preserving important edge information. The second branch processes the input, and its output is added to the multiplied result from the first branch, effectively removing noise and artifacts from the image.

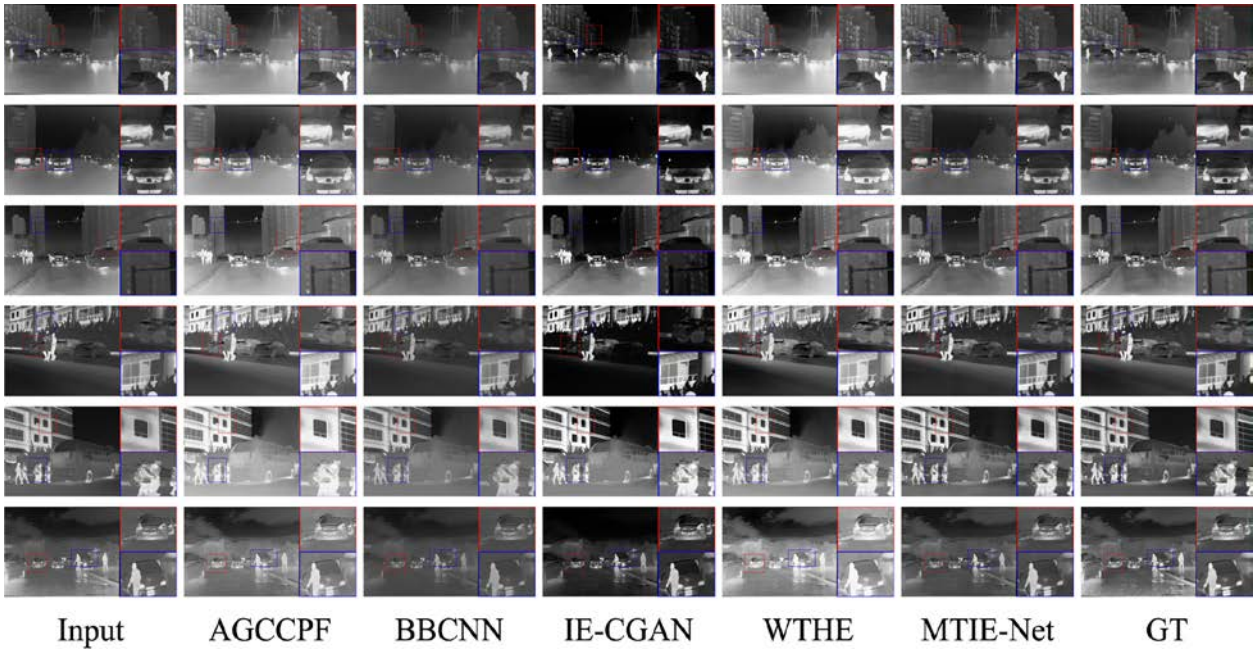

Fig. 6. Qualitative comparison of MTIE-Net.

3.2 Training and Dataset

We selected the M3DF [32] dataset for training due to its unique provision of pixel-aligned TIR and VIS image pairs. The dataset was captured using a synchronized system consisting of one binocular optical camera and one binocular infrared sensor, ensuring accurate alignment between the TIR and VIS images. The scenes feature diverse environments, including campus areas and roads, with a total of 4,200 image pairs as well as an additional 300 pairs from independent scenes, which we used as the test set. We employed the atmospheric scattering model (ASM) to create synthetic hazy images widely used in dehazing methods [33]. ASM models a hazy image *I(x)* as

$$I(x) = J(x) \cdot t(x) + A \cdot \big(1 - t(x)\big),$$

where $J(x)$ is the haze-free image, A the atmospheric light and

$$t(x) = e^{-\beta\, d(x)}$$

is the transmission matrix [34] with β being the atmospheric scattering coefficient and $d(x)$ the depth map. For this approach, both the image and its corresponding depth map are required. We utilized corresponding VIS images to generate depth maps for the TIR images, as many depth estimation algorithms work well for visible images. We used the DIFFNet [35] method for generating depth maps, results shown in Fig. 4. The scattering coefficient β was selected uniformly from the range [0.6, 1.8], while A from [0.2, 0.6], following [36]. Finally, we added Gaussian noise to make the synthetic TIR images more realistic. For the ground truth label images, we applied the Contrast Limited Adaptive Histogram Equalization [37] technique to enhance the input images, using a clip limit 1.2.

Since thermal images inherently contain noise, we incorporated the Non-local Means Denoising algorithm [38] to reduce noise levels while preserving important features. To retain essential details that might be lost during denoising, we merged the denoised image with the enhanced image using a coefficient of 0.4. We further combined this result with the original input

image using a coefficient of 0.2 (as shown in Fig. 5). This multi-step merging process ensures that the final output maintains the critical features of the original input while effectively denoising and enhancing the image. The network was trained for 300 epochs using the Adam optimizer (Adaptive Moment Estimation) with an initial learning rate of 1×10−4. Input images were cropped to 256×256 pixels to ensure consistent processing and to optimize the training efficiency.

3.3 Loss Function

The network training process minimizes the loss between the predicted images and the corresponding high-quality ground truth images. We use the Mean Squared Error (MSE) metric, which calculates the average of the squared differences between predicted and actual values. MSE assigns greater weight to larger errors, making it sensitive to outliers and effective for minimizing the overall error in the model. The loss function is defined as follows:

$$L_{MSE} = MSE(I_{pred}, I_{gt})$$

where

$$MSE(I_{pred}, I_{gt}) = \sum_{i=1}^{M} \sum_{j=1}^{N} (I_{pred_{ij}} - I_{gt_{ij}})^2 \quad (1)$$

where I_{pred} is the predicted image, I_{gt} is the ground truth image M, N are the width and height of the image, respectively. The MSE ensures consistency between the network's predictions and the reference images, guiding the model to improve its accuracy.

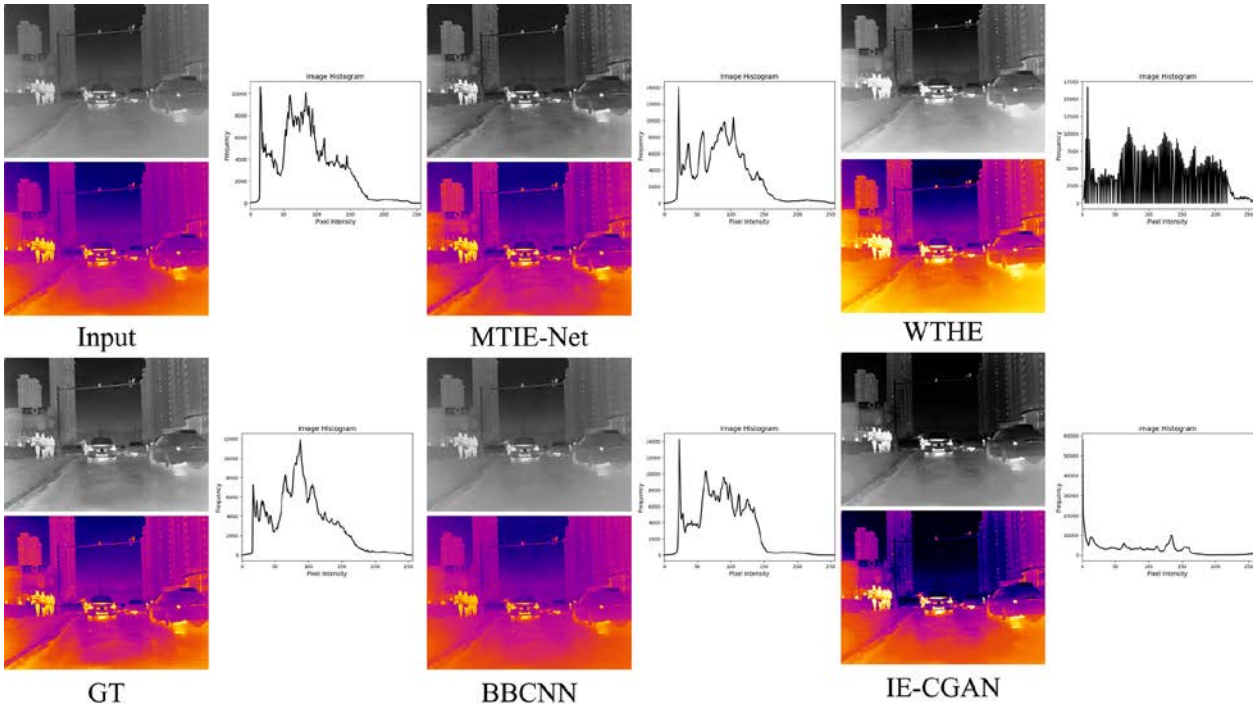

Fig. 7. This figure shows our ability to recover both high-level and low-level features compared with BBCNN, IE-CGAN and WTHE.

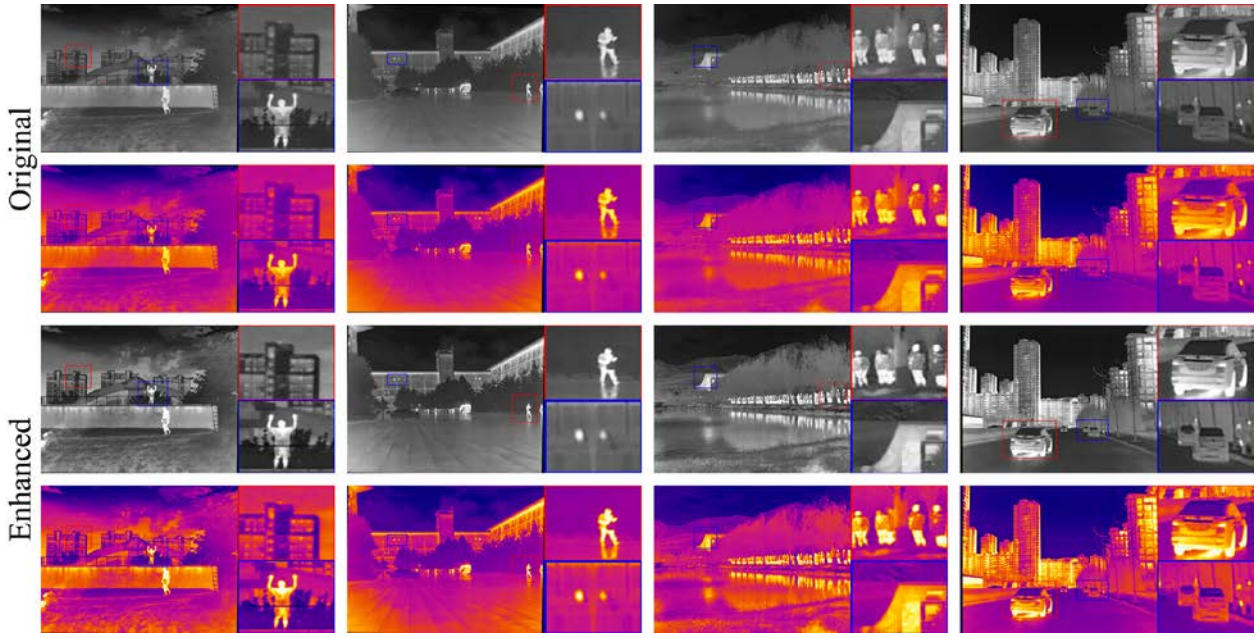

Fig. 8. Performance of MTIE-Net on M3DF dataset.

4. Experimental Results

This section presents the experimental results of our proposed approach in comparison with several existing methods, including AGCCPF [39], WTHE [40], BBCNN [41], and IE-CGAN [42]. Through this comparative analysis, we aim to highlight the strengths and weaknesses of each method and evaluate the effectiveness of MTIE-Net against these techniques. MTIE-Net provides a more robust foundation for object detection, tracking, and classification applications by maintaining optimal visibility across various image degradation scenarios.

4.1 Qualitative Comparison

Figure 6 compares MTIE-Net qualitatively against other thermal image enhancement techniques. As seen in the figure, IE-CGAN successfully reduces haze but makes the image too dark, leading to the loss of crucial details. For example, in the third row, it fails to recover the traffic light and loss of car information, which is critical for applications like pedestrian or vehicle detection. AGCCPF and WTHE tend to over-enhance road textures, losing essential details, particularly in lighter (hazy) areas. However, WTHE performs better in handling contrast in extremely dark regions. BBCNN often produces unnatural sharpness and introduces halo artifacts, negatively affecting overall image quality. In contrast, MTIE-Net provides a balanced enhancement of hazy regions while preserving essential details, maintaining natural contrast and depth information as shown in Fig. 7. It restores image details without introducing halo artifacts, making it more applicable for real-world applications. Although to understand the generalizability of our approach, we ran the algorithm on the original M3DF dataset, showing its performance on real-world thermal images (see Fig. 8). MTIE-Net effectively enhances thermal images captured in natural conditions, where haze, noise, and low contrast are naturally present. Unlike the previous experiments with simulated hazy images, this evaluation demonstrates the model's robustness and adaptability to complex, real-world degradations.

MTIE-Net successfully restores essential details and preserves natural contrast without introducing halo artifacts. Also, using basic colorization methods on thermal images makes the

results easier to interpret, giving a clearer and more intuitive view of the scene. This improvement is crucial for tasks like autonomous driving and surveillance, where clarity and detail are vital.

4.2 Quantitative Comparison

To evaluate the effectiveness of MTIE-Net architecture, we utilize five image quality metrics:

(i) Peak Signal-to-Noise Ratio (PSNR), which measures the ratio between the maximum possible signal power and the power of noise that affects image quality, commonly used for assessing the reconstruction quality in lossy image compression codecs. Given a reference image I and a test image J, both of size $M \times N$, the PSNR between I and J is defined by:

$$PSNR(I,J) = 10 \log(\frac{255^2}{MSE(I,J)})$$

where MSE is defined in (1).

(ii) Structural Similarity Index Measure (SSIM) [20], which evaluates image degradation based on perceived structural information. SSIM considers factors such as luminance and contrast masking to assess the similarity between two images. SSIM is defined as:

$$SSIM(I,J) = l(I,J) \cdot c(I,J) \cdot s(I,J)$$

where $l(I,J)$ is the luminance comparison function that measures the closeness of the two images mean luminance, $c(I,J)$ is the contrast comparison function that measures the closeness of the contrast between the two images (where contrast is measured by the standard deviation), and $s(I,J)$ is the structure comparison function that measures the correlation coefficient between the two images [20, 43].

(iii) Measure of Enhancement (EME) [44], which assesses the entropy of block-wise image contrasts rather than individual pixel values. This metric is crucial for evaluating the quality of enhanced images, as it highlights contrast differences within blocks. The EME metric is defined as:

$$EME(I) = \frac{1}{n}\sum_{i=1}^{n} 20 \, ln(\frac{I_{max}^k}{I_{min}^k + c})$$

where I image is partitioned into n blocks, I_{max}^k, I_{min}^k are the maximum and minimum values of the k block, and c is a small constant to avoid division by zero.

(iv) Block-Based Information Measure (BDIM) [45], which evaluates the amount of information present in image blocks by considering both local contrast and structural details. This measure provides insight into the effectiveness of enhancement by focusing on small image regions, ensuring that fine details are preserved and enhanced across blocks.

(v) Global Contrast Measure of Enhancement (MDIMTE) [46], which combines features accounting for the human visual system, information theory, and distribution-based measures. This metric provides a holistic view of enhancement quality, considering global contrast improvements

that align with how humans perceive image quality and how efficiently information is distributed throughout the image.

High scores across these metrics indicate superior image enhancement and a more natural visual appearance. Table 2 presents the comparative analysis results, showcasing MTIE-Net performance against existing methods. Our method outperforms both traditional and CNN-based approaches, achieving the highest average scores in SSIM, PSNR. These results underscore the superior capability of MTIE-Net in enhancing thermal images while preserving essential details and maintaining a realistic look.

Table 2. Quantitative Comparison of MTIE-Net with Other State-of-the-Art Methods

Measure	AGCCPF [39]	BBCNN [41]	IE-CGAN [42]	WTHE [40]	MTI-Net
PSNR ↑	15.204	22.638	19.535	15.406	**24.303**
SSIM ↑	0.852	0.909	0.718	0.872	**0.948**
EME ↑	1.600	1.688	1.496	2.618	**3.001**
BDIM ↑	0.922	0.928	0.912	0.921	**0.939**
MDIMTE ↑	52.987	54.931	48.111	51.131	**57.584**

5. Ablation Study

We conducted a series of ablation experiments to assess the contribution of the SS2D and ED blocks to the thermal image enhancement task. Specifically, we trained MTIE-Net with and without these modules to understand their impact on performance. In one variant, only the ED block was retained while the SS2D block was removed, and in another variant, only the SS2D block was retained while the ED block was removed. As shown in Table 3, the highest PSNR, SSIM, EME, BDIM and MDIMTE values were achieved when both the ED and SS2D blocks were integrated, indicating that their combination significantly improves the model's ability to enhance overall image quality. This demonstrates that integrating ED and SS2D is critical for achieving superior performance in thermal image enhancement.

Table 3. Ablation Study on M3DF Dataset

Methods	MTIE-Net w/o SS2D	MTIE-Net w/o ED	MTIE-Net
PSNR ↑	22.915	23.143	**24.303**
SSIM ↑	0.911	0.932	**0.948**
EME ↑	2.837	2.877	**3.001**
BDIM ↑	0.934	0.934	**0.939**
MDIMTE ↑	56.821	57.193	**57.584**

In addition to the experiments conducted on the M3DF dataset, we further evaluated the data independence of MTIE-Net using real-world solar panel images provided by SOLARON LLC. These images offer a diverse set of real-world conditions, enabling us to test the model's ability to generalize beyond training data.

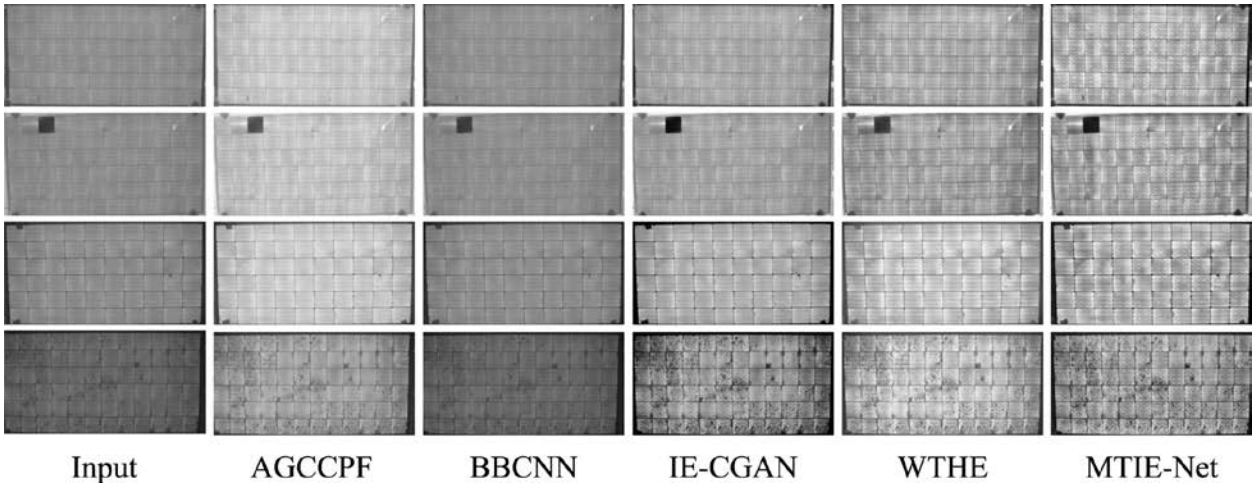

Fig. 9. Qualitative comparison of MTIE-Net on Solar Panel Images.

Fig. 9 presents a qualitative comparison of MTIE-Net against other enhancement techniques on the solar panel dataset. MTIE-Net consistently outperforms the other methods, making clearer and more detailed images while maintaining critical features, such as panel textures and fault regions. This demonstrates the robustness of MTIE-Net in handling different real-world datasets beyond the original thermal images used for training. To provide a more detailed visualization,

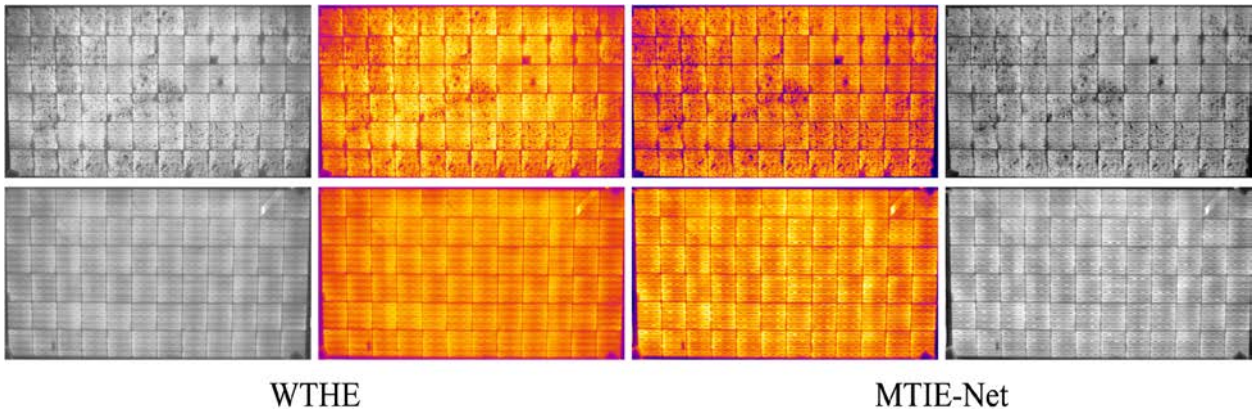

Fig. 10. Detailed Visualization of MTIE-Net (our) and WTHE on Solar Panel Images (the best one).

Fig. 10 shows two selected solar panel images enhanced by the best-performing model from the previous comparisons, with the simple colorization technique applied. These images highlight specific visual improvements, such as better contrast and detail preservation. The results underline the data-independent nature of MTIE-Net, confirming its ability to generalize effectively to new datasets.

6. Conclusion

In conclusion, this paper introduces the novel MTIE-Net architecture for thermal image dehazing, building upon the Mamba model to achieve superior performance. We enhanced feature extraction capabilities by integrating an advanced ED module with the SS2D model, significantly improving image quality and clarity. We also incorporated labeled thermal infrared (TIR) data for degradation enhancement, utilizing depth information from aligned visible-infrared (VIS-TIR) image pairs.

Our extensive quantitative and qualitative evaluations demonstrate that MTIE-Net outperforms state-of-the-art methods such as IE-CGAN, BBCNN, AGCCPF, and WTHE on the M3DF dataset. The results validate MTIE-Net's robustness and reliability across various imaging scenarios, showing consistently superior performance in both objective measures and visual assessments compared to other approaches. MTIE-Net shows significant potential for advancing thermal image-based applications. Future work will focus on developing a unified network capable of processing multi-scale and multi-spectral images, thereby broadening the versatility and applicability of our approach.

Additionally, we plan to integrate domain adaptation techniques to address the challenges of different imaging conditions and environments, ensuring more generalized performance. We will also explore real-time processing capabilities to make the network suitable for applications that require immediate analysis, such as autonomous navigation, surveillance, and disaster response. This ongoing development aims to enhance thermal imaging performance further, offering valuable improvements for diverse applications, including environmental monitoring, security, and beyond.

References

[1] B. Li et al., "Benchmarking single image dehazing and beyond", *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 492–505, 2019.

[2] S. Hovhannisyan et al., "AED-Net: A Single Image Dehazing", in *IEEE Access*, vol. 10, pp. 12465-12474, 2022.

[3] S. Hovhannisyan, H. Gasparyan and S. Agaian, "EOD-Net: enhancing object detection in challenging weather conditions using an innovative end-to-end dehazing network", *Twelfth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, Paris, France, pp. 1-6, 2023.

[4] D. Berman, T. Treibitz and S. Avidan, "Single image dehazing using haze-lines", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 3, pp. 720–734, 2020.

[5] Z. Chen, Z. He and Z. -M. Lu, "DEA-Net: single image dehazing based on detail-enhanced convolution and content-guided attention", *IEEE Transactions on Image Processing*, vol. 33, pp. 1002-1015, 2024.

[6] A. Oulefki, T. Trongtirakul, S. Agaian et al., "multi-view vr imaging for enhanced analysis of dust accumulation on solar panels", *Solar Energy*, vol. 279, pp. 112708, 2024.

[7] A. Parihar, Y. Gupta, Y. Singodia, V. Singh and K. Singh, "A comparative study of image dehazing algorithms", *5th International Conference on Communication and Electronics Systems (ICCES)*, IEEE, pp 766–771, 2020.

[8] A. Wilson et al., "Recent advances in thermal imaging and its applications using machine learning: A review", *IEEE Sens. J.*, vol. 23, pp. 3395–3407, 2023

[9] J. Baek, S. Hong, J. Kim and E. Kim, "Efficient Pedestrian detection at nighttime using a thermal camera" *Sensors*, vol. 17, no. 8, p. 1850, 2017.

[10] A. Goldberg, T. Fischer and Z. Derzko, "Application of dual-band infrared focal plane arrays to tactical and strategic military problems" *in Proc. SPIE*, vol. 4820, pp. 500–514, 2003.

[11] W. Wong, H. Lim, C. Loo and W. Lim, "Home alone faint detection surveillance system using thermal camera" *in Proc. 2nd Int. Conf. Comput. Res. Develop*., pp. 747–751, 2010.

[12] J. Berni, et al., "Remote sensing of vegetation from UAV platforms using lightweight multispectral and thermal imaging sensors", *Int. Arch. Photogramm. Remote Sens. Spat. Inform. Sci.* 38:6, 2008.

[13] T. Trongtirakul and S. Agaian "New retinex model-based infrared image enhancement", *Proc. SPIE 12526, Multimodal Image Exploitation and Learning*, 1252606, 2023.
[14] A. Gu and T. Dao, “Mamba: linear-time sequence modeling with selective state spaces”, *arXiv* preprint arXiv:2312.00752, 2023.
[15] E. Nguyen et al., “Modeling images and videos as multidimensional signals with state spaces”, *Advances in neural information processing systems*, vol. 35, pp. 2846–2861, 2022.
[16] J. Smith, A. Warrington and S. Linderman, “Simplified state space layers for sequence modeling, " *arXiv* preprint arXiv:2208.04933, 2023.
[17] B. Patro and V. Agneeswaran, “Mamba-360: Survey of state space models as transformer alternative for long sequence modelling: Methods, applications, and challenges”, *arXiv* preprint arXiv:2404.16112, 2024.
[18] Y. Cao and W. Zhang, “Mamba4KT: An Efficient and Effective Mamba-based Knowledge Tracing Model”, *arXiv* preprint arXiv:2405.16542, 2024.
[19] T. Dao and A. Gu, “Transformers are ssms: Generalized models and efficient algorithms through structured state space duality”, *arXiv* preprin arXiv:2405.21060, 2024.
[20] Z. Wang, A. Bovik, H. Sheikh and E. Simoncelli, “Image quality assessment: From error visibility to structural similarity”, *IEEE Trans. Image Process*., vol. 13, no. 4, pp. 600–612, Apr. 2004.
[21] L. Lu, et. al., Comparative study of histogram equalization algorithms for image enhancement. Mob. Multimed. Image Process. Secur. Appl. 2010, 7708, 337–347
[22] J. Wang, et al., “Range-restricted pixel difference global histogram equalization for infrared image contrast enhancement”, *Opt Rev 28*, 145–158, 2021.
[23] T. Mudavath and V. Niranjan, “Thermal image enhancement for adverse weather scenarios: a wavelet transform and histogram clipping approach”, *SIViP 18*, pp. 6547–6558, 2024.
[24] R. Soundrapandiyan, et al. “A comprehensive survey on image enhancement techniques with special emphasis on infrared images”, *Multimed Tools Appl 81*, 9045–9077, 2022.
[25] A. Grigoryan and S. Agaian, “Asymmetric and symmetric gradient operators with application in face recognition in Renaissance portrait art”, *Proc. of SPIE, Defense + Commercial Sensing, Mobile Multimedia/Image Processing, Security, and Applications*, vol. 10993, p. 12, Baltimore, Maryland, April 2019.
[26] R. Kalman, “A new approach to linear filtering and prediction problems”, 1960.
[27] M. Ahamed and Q. Cheng, “TSCMamba: Mamba meets multi-view learning for time series classification”, *arXiv* preprint arXiv:2406.04419, 2024.
[28] S. Hochreiter and J. Schmidhuber, “Long Short-Term memory”, *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[29] C. Yuan, D. Zhao and S. Agaian, “MUCM-Net: a mamba powered ucm-net for skin lesion segmentation”, *arXiv* preprint arXiv:2405.15925.
[30] H. Zhang et. al., “A survey on visual mamba”, *arXiv* preprint, arXiv:2404.15956v2, 2024.
[31] L. Zhu et al., “Vision mamba: Efficient visual representation learning with bidirectional state space model”, *arXiv* preprint arXiv:2401.09417, 2024.
[32] J. Liu et al., “Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection”, *IEEE/CVF International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.
[33] T. Trongtirakul and S. Agaian, “New Retinex model-based infrared image enhancement”, *Proc. SPIE 12526, Multimodal Image Exploitation and Learning*, 1252606, 2023.
[34] T. Trongtirakul, and S. Agaian, “Transmission map optimization for single image dehazing”, *Proc. SPIE 12100, Multimodal Image Expl. and Learning*, 121000C, May, 2022.

[35] H. Zhou, D. Greenwood and S. Taylor, “Self-supervised monocular depth estimation with internal feature fusion”, *British Machine Vision Conference (BMVC)*, 2021.
[36] F. Erlenbusch et al., “Thermal infrared single image dehazing and blind image quality assessment”, *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Vancouver, BC, Canada, 2023, pp. 459-469, 2023.
[37] A. Reza, “Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement”, *J. VLSI Sig. Proc. Syst. Signal Image Vid. Tech..*, vol. 38, no. 1, pp. 35-44, 2004.
[38] A. Buades, B. Coll and J. Morel, “Non-local means denoising," *Image Processing On Line*, 1, pp. 208–212, 2011.
[39] B. Gupta and M. Tiwari, “Minimum mean brightness error contrast enhancement of color images using adaptive gamma correction with color preserving framework”, *Optik* 127, no. 4, pp. 1671-1676, 2016.
[40] Q. Wang and and R. Ward, “Fast image/video contrast enhancement based on weighted thresholded histogram equalization”, *IEEE Trans. On Consumer Electronics* 53, no. 2, 757-764, 2007.
[41] K. Lee et al., “Brightness-based convolutional neural network for thermal image enhancement”, *IEEE Access*, vol. 5, pp. 26867-26879, 2017.
[42] K. Xiaodong et al., “Single infrared image enhancement using a deep convolutional neural network”, *Neurocomputing*, vol. 332, pp. 119-128, 2019.
[43] A. Horé and D. Ziou, “Image quality metrics: PSNR vs. SSIM”, *20th International Conference on Pattern Recognition*, pp. 2366-2369, Turkey, 2010.
[44] S. Agaian et al., “A new measure of image enhancement”, *IASTED Int. Conf. on Signal Proc. & Comm.*, Sept. 2000.
[45] W. Yang et al., “Blind image quality assessment with a multi-task CNN for enhanced measurement”, *Sig. Pr.: Im. Com.*, 105, 116672, 2022.
[46] S. Agaian, M. Roopaei and D. Akopian, “Thermal-image quality measurements”, *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014.

Մամբայի վրա հիմնված ջերմային պատկերի մառախուղի հեռացում

Սարգիս Ա. Հովհաննիսյան

Երևանի պետական համալսարան, Երևան, Հայաստան
e-mail: sargis.hovhannisyan@ysu.am

Ամփոփում

Մթնոլորտային երևույթները, ինչպիսիք են անձրևը, ձյունը, քաղաքային, անտառային հրդեհները և արհեստական աղետները, կարող են վատթարացնել պատկերի որակը տարբեր կիրառություններում, ներառյալ տրանսպորտը, վարորդների աջակցության

համակարգերը, հսկողությունը, ռազմական և հեռահաղորդակցական համակարգերը: Պատկերի մառախուղի հեռացման տեխնիկաները նպատակ ունեն նվազեցնել մառախուղի, փոշու և այլ մթնոլորտային աղավաղումների հետևանքները՝ բարելավելով պատկերի որակը՝ համակարգչային տեսողության առաջադրանքների ավելի լավ կատարման համար: Մառախուղը ոչ միայն թաքցնում է մանրամասները, այլև նվազեցնում է կոնտրաստը և գույնի հավաստիությունը՝ զգալիորեն ազդելով համակարգչային տեսողության (CV) մոդելների ճշգրտության վրա, որոնք օգտագործվում են օբյեկտների հայտնաբերման, պատկերների դասակարգման և սեգմենտավորման մեջ: Թեև ջերմային (TIR) պատկերները հաճախ նախընտրելի են երկար հեռավորության հսկողության և հեռահաղորդակցության համար՝ շնորհիվ դրանց դիմադրողականության մառախուղին, սակայն մթնոլորտային պայմանները կարող են վատթարացնել նրանց որակը, հատկապես եղանակային ծայրահեղ միջավայրերում: Այս հոդվածը ներկայացնում է MTIE-Net՝ նոր մամբայի վրա հիմնված ցանց՝ մթնոլորտային երևույթների շնորհիվ դեգրադացված ջերմային պատկերների բարելավման համար, ինչպիսիք են մառախուղը և ծուխը: MTIE-Net-ը օգտագործում է Enhancement and Denoising State Space Model (EDSSM), որը համատեղում է կոնվոլյուցիոն նեյրոնային ցանցերը State Space մոդելավորման հետ՝ արդյունավետ աղմուկի հեռացման և նկարի բարելավման համար: Մենք ստեղծում ենք սինթետիկ մառախուղապատ պատկերներ և կիրառում ոլորտին հատուկ փոխակերպումներ, որոնք հարմարեցված են ջերմային պատկերի հատկություններին՝ ցածր տեսանելիության պայմաններում ուսուցումը բարելավելու համար: Մեր հիմնական ներդրումներն են մամբա ճարտարապետության կիրառումը 2D ընտրողական սքանավորմամբ՝ ջերմային պատկերների բարելավման համար, աղմուկի հեռացման և պատկերի բարելավման մոդուլի մշակումը և ջերմային պատկերների համար սինթետիկ մառախուղապատ տվյալների հավաքածուի ստեղծումը: M3DF երկար հեռավորության ջերմային պատկերների տվյալների հավաքածուի վրա գնահատվելով՝ MTIE-Net-ը գերազանցում է ժամանակակից մեթոդները ինչպես քանակական չափումների (PSNR, SSIM), այնպես էլ տեսողական պարզության և եզրերի պահպանման որակական գնահատականներում: Այս առաջընթացը զգալիորեն բարելավում է հեռահաղորդակցման, հսկողության և ավտոմատ համակարգերի հուսալիությունն ու ճշգրտությունը՝ մթնոլորտային վատ միջավայրերում պատկերի որակի բարելավման շնորհիվ:

Բանալի բառեր՝ ջերմային պատկեր, պատկերի բարելավում, ջերմային պատկերի բարելավում, մամբա:

Удаление тумана с тепловизионных изображений на основе Мамба

Саргис А. Оганнисян

Ереванский государственный университет, Ереван, Армения
e-mail: sargis.hovhannisyan@ysu.am

Аннотация

Атмосферные явления, такие как дождь, снег, городские и лесные пожары, а также искусственные катастрофы, могут ухудшать качество изображения в различных областях, включая транспорт, системы помощи водителям, видеонаблюдение, военные и телекоммуникационные системы. Технологии удаления тумана с изображений направлены на снижение последствий тумана, пыли и других атмосферных искажений, улучшая качество изображения для более эффективного выполнения задач компьютерного зрения. Туман не только скрывает детали, но и снижает контраст и цветовую точность, что существенно влияет на точность моделей компьютерного зрения (CV), используемых для обнаружения объектов, классификации изображений и сегментации. Хотя тепловизионные изображения (TIR) часто предпочитаются для дальнего контроля и телекоммуникаций благодаря их стойкости к туману, атмосферные условия все же могут ухудшить их качество, особенно в экстремальных погодных условиях.

В этой статье представлен MTIE-Net — новая сеть на основе Mamba для улучшения тепловизионных изображений, искаженных атмосферными явлениями, такими как туман и дым. MTIE-Net использует Enhancement and Denoising State Space Model (EDSSM), которая сочетает в себе сверточные нейронные сети и моделирование состояний для эффективного удаления шума и улучшения изображения. Мы создаем синтетические изображения с туманом и применяем специфические для домена преобразования, адаптированные к характеристикам тепловизионных изображений, для улучшения обучения в условиях низкой видимости. Наши ключевые достижения включают использование архитектуры Mamba с 2D избирательным сканированием для улучшения тепловизионных изображений, разработку модуля для удаления шума и улучшения изображения, а также создание набора данных с синтетическим туманом для тепловизионных изображений. Оцененный на наборе данных M3DF для дальних тепловизионных изображений, MTIE-Net превосходит современные методы как по количественным показателям (PSNR, SSIM), так и по качественным оценкам визуальной четкости и сохранения краев. Это достижение значительно улучшает надежность и точность телекоммуникационных, систем наблюдения и автоматических систем за счет улучшения качества изображения в сложных атмосферных условиях

Ключевые слова: тепловизионное изображение; улучшение изображения; улучшение тепловизионного изображения; Мамба.