Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

DFARNet: Learnable Frequency Attention ResNet with Dual-Attention for Single-Image Desnowing

Mohammed Amine LOUMI
Changchun University of Science and Technology (CUST)
QianYi ZHOU
Changchun University of Science and Technology (CUST)
Ming FANG
fangming@cust.edu.cn

Changchun University of Science and Technology (CUST)

Research Article

Keywords: Computer Vision, Image Processing, Single-image Desnowing, Frequency Attention, ResNet

Posted Date: November 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11276301/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

DFARNet: Learnable Frequency Attention ResNet with Dual-Attention for Single-Image Desnowing

Mohammed Amine LOUMI[1], QianYi ZHOU[1], Ming FANG[1*]

[1]School of Artificial Intelligence, Changchun University of Science and Technology (CUST), Changchun, Jilin, China .

*Corresponding author(s). E-mail(s): fangming@cust.edu.cn;
Contributing authors: mohammedamine.loumi@mails.cust.edu.cn;
zhou010513@outlook.com;

Abstract

Single Image Desnowing is a significant and challenging area within computer vision, with numerous practical applications—particularly in autonomous vehicles, surveillance systems, and outdoor imaging. Current methods rely primarily on spatial attention and fixed frequency analysis, but they often struggle to distinguish between snow particles and natural edges, which share similar visual characteristics. To address this limitation, we propose the DFARNet architecture, which employs a dual-attention mechanism combining learnable frequency attention (to detect snow-specific frequency signatures) and spatial attention (to accurately locate snow regions). This complementary design enables the model to both identify where snow appears and what form it takes in the frequency domain. The model was trained using a composite loss function comprising an L1 loss and an edge-preserving loss, ensuring effective snow removal while maintaining structural details. Experiments conducted on the CSD dataset demonstrated a Peak Signal-to-Noise Ratio (PSNR) of 31.6 dB under resource-constrained conditions, making the model particularly suitable for deployment in systems with limited computational resources.

Keywords: Computer Vision , Image Processing , Single-image Desnowing , Frequency Attention , ResNet

1 Introduction

Single-image desnowing is a fundamental and complex problem in low-level computer vision, with direct applications in autonomous driving, surveillance systems, outdoor photography, and robotics. Unlike rain or fog, snow exhibits complex optical properties that make its removal challenging; particle sizes vary from light flurries to heavy blizzards, and snow possesses partial transparency (an occlusion effect) that obscures background details, in addition to bearing a strong visual resemblance to other white objects such as clouds, walls, and flowers. These unique characteristics render general image restoration methods inadequate, necessitating specialized approaches capable of effectively distinguishing between snow and scene content.

The task of desnowing can be formally framed as an image-to-image translation problem. Given an observed snowy image $I \in \mathbb{R}^{3\times H\times W}$, the objective is to recover the original, clean scene $J \in \mathbb{R}^{3\times H\times W}$ by modeling the image degradation process. A widely used degradation model for snow-covered images describes the observed image as a combination of atmospheric light, a transmission map, and snow particles [1]:

$$I(x) = J(x) \cdot t(x) + A \cdot (1 - t(x)) + S(x) \tag{1}$$

Here, x denotes pixel coordinates, $t(x)$ is the transmission map representing the fraction of light reaching the camera, A is the global atmospheric light, and $S(x)$ represents the snow particle component. This formulation highlights three key challenges in snow removal: estimating the transmission map for semi-transparent snow regions, separating snow particles from background features, and preserving fine details and edges during image reconstruction [1, 2].

Early deep learning-based snow removal methods operated primarily in the spatial domain. DesnowNet [2] employed a context-aware encoder-decoder architecture with skip connections, whereas JSTASR [1] modeled snow size and transparency using modified partial convolutions. Both networks share a key limitation: the inability to distinguish between snow and white objects that are structurally similar to it.

Frequency-domain analysis addresses this gap; the Discrete Fourier Transform (DFT) magnitude spectrum reveals frequency intensity, while the phase spectrum preserves spatial structure [3, 4]:

$$F(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \cdot e^{-j2\pi(ux/M+vy/N)} \tag{2}$$

The fundamental insight is that snow and natural edges possess distinct frequency signatures: natural edges are characterized by directionality and organized structure, whereas snow particles are random and isotropic [5]. Chen et al. [6] were the first to employ the Dual-Tree Complex Wavelet Transform (DTCWT) alongside a loss function based on "contradict channel loss"; subsequent methods (such as WaveFrSnow and ECFNet) explored frequency attention mechanisms but relied on static frequency analysis or simple high-frequency enhancement.

These methods suffer from limitations such as the use of static, non-adaptive frequency analysis and an inability to distinguish between random snow and directional edges. To address these limitations, we propose DFARNet (Desnowing Frequency

Attention ResNet), an architecture featuring a learnable frequency attention mechanism specifically designed for snow removal; it employs a dual-attention framework that combines learnable frequency attention with spatial attention, while utilizing an enhanced edge loss function during training.

2 Related Works

2.1 Overview and Research Landscape

Single-image Desnowing has evolved from spatial-domain CNNs to transformer-based and frequency-aware architectures. Snow presents unique challenges: diverse particle sizes, varying transparency, and visual similarity to other white objects. The field has advanced through three phases: spatial-domain deep learning (2018–2020), frequency-aware methods (2021–2022), and transformer-based and hybrid approaches (2023–present).

2.2 Review of Recent Methods

2.2.1 Transformer-Based Methods

Restormer [7] employs multi-head attention across channels for efficient high-resolution restoration, SnowFormer [8] which it is the first transformer specialized for snow removal, achieving 33.92 dB on CSD, yet relies on spatial processing without frequency analysis. and also there is BAT-Net [9] introduces bidirectional attention between dual decoders for joint snow removal and mask prediction, reaching 35.78 dB on CSD, but incurs high computational cost and lacks frequency modeling.

2.2.2 Frequency-Aware and Hybrid Methods

DTCWT [6] first demonstrated frequency decomposition for snow removal using hierarchical dual-tree complex wavelet transform, but relies on static wavelet filters.WaveFrSnow [10] combines wavelet transform with frequency-separation cross-attention, yet still uses fixed decomposition. ECFNet [11] introduces SFAM (SDAM + FDAM) for spatial and frequency attention, but frequency attention remains simple high-frequency emphasis without learned selection.ARANet [12] combines Half-Quadratic Splitting with hybrid spatial-frequency attention, but is general-purpose and not snow-specific.

2.2.3 Parameter-Efficient and Efficient Architecture Methods

TuneSnow [13] applies Hybrid Adapters to fine-tune SAM with only 15% parameters, achieving 36.02 dB on CSD, but depends on an external pre-trained model and lacks frequency reasoning. MCAD-Net [14] integrates multi-scale context-aware attention with ghost encoders for efficiency, but remains purely spatial.

2.3 Discussion and Analysis

the models which they uses frequency attention which it uses static wavelet decomposition cannot adapt to varying snow characteristics, frequency attention uses simple high-frequency emphasis rather than learnable frequency selection, spatial and frequency attention operate separately without complementary integration , and also interpretability is largely ignored with no visualizable attention maps, there is a persistent performance-efficiency trade-off exists, with transformers being costly and efficient methods depending on external models.

2.4 Research Gaps

Four critical gaps remain in existing snow removal methods: the absence of learnable frequency attention mechanisms adaptive to varying snow, the lack of explicit distinction between snow and natural edges, the absence of a unified dual-attention framework combining spatial and frequency attention, and limited interpretability with no visualizable attention maps.

3 Methodology

3.1 Overall Architecture

We propose Desnowing Frequency Attention ResNet (DFARNet), a novel architecture for single-image snow removal that operates in both spatial and frequency domains. As illustrated in Figure 1, DFARNet consists of three main components: (1) a spatial feature extraction branch that identifies snow-affected regions, (2) a frequency attention branch that discriminates snow-specific frequency signatures, and (3) a dual-attention fusion module that combines both branches for comprehensive snow removal.

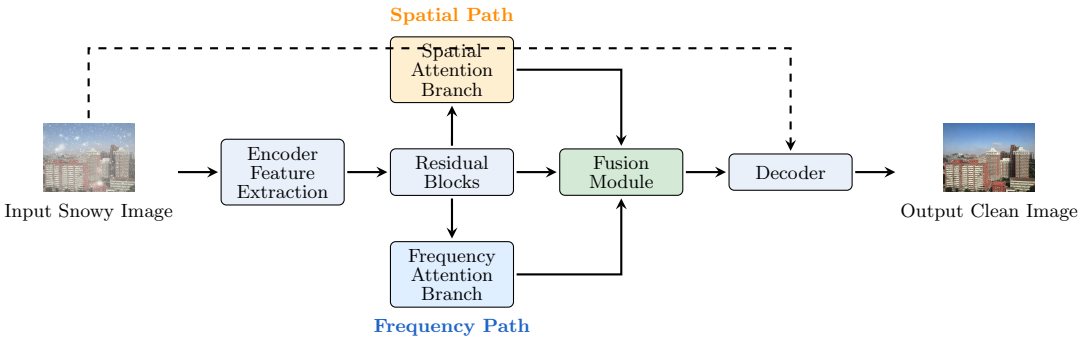


Fig. 1 Overall architecture of DFARNet. The input image is processed through a shared encoder, then branched into spatial and frequency attention paths before fusion and decoding.

The motivation for our dual-attention design stems from the complementary nature of spatial and frequency information. Snow particles exhibit high-frequency random patterns that are distinguishable from the organized directional patterns of natural edges [1, 15]. However, frequency analysis alone cannot localize snow regions, while spatial attention alone cannot distinguish snow from similar white objects. Our framework addresses this limitation by integrating:

- **Spatial attention**: Identifies *where* snow is located in the image
- **Frequency attention**: Identifies *what* snow looks like in the frequency domain

Formally, given an input snowy image $I \in \mathbb{R}^{3\times H\times W}$, our model learns a mapping $\mathcal{F}$ to estimate the clean image $\hat{I}$:

$$\hat{I} = \mathcal{F}(I) = \mathcal{D}(\mathcal{F}_{\text{spatial}}(I) \oplus \mathcal{F}_{\text{freq}}(I)) \tag{3}$$

where $\mathcal{F}_{\text{spatial}}$ and $\mathcal{F}_{\text{freq}}$ represent the spatial and frequency attention branches respectively, $\oplus$ denotes feature fusion, and $\mathcal{D}$ is the decoder network.

3.2 Spatial Attention Branch

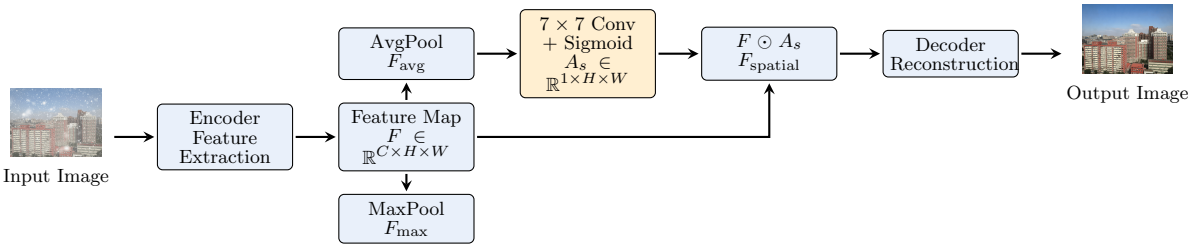


Fig. 2 Detailed architecture of the Spatial Attention Branch. The feature map is processed through AvgPool and MaxPool operations, then passed through a 7×7 convolution with sigmoid activation to generate spatial attention weights.

The spatial attention branch is designed to locate snow-affected regions while preserving fine structural details. Building upon the residual learning framework [16], we adopt a U-Net style encoder-decoder architecture with skip connections.

3.2.1 Encoder

The encoder consists of four convolutional blocks, each containing: Convolutional layer with kernel size 3×3 and stride 2 (downsampling) , Batch normalization ,[17]ReLU activation

This progressive downsampling enables the network to capture multi-scale contextual information, which is essential for distinguishing snow of varying particle sizes [2].

3.2.2 Spatial Attention Module

Inspired by the Convolutional Block Attention Module (CBAM) [18], we introduce a spatial attention module that generates a 2D attention map highlighting snow regions. Unlike CBAM which uses both channel and spatial attention, our module focuses exclusively on spatial dimensions to localize snow:

$$A_{\text{spatial}} = \sigma(f^{7\times7}([\text{AvgPool}(F); \text{MaxPool}(F)])) \qquad \sigma(x) = \frac{1}{1+e^{-x}} \tag{4}$$

where F denotes the intermediate feature map, $f^{7\times7}$ is a 7×7 convolution, and σ is the sigmoid activation function .

3.2.3 Decoder

The decoder mirrors the encoder structure, replacing downsampling with upsampling using transposed convolutions. Skip connections from the encoder to the decoder preserve high-frequency details (edges and textures) that are critical for high-quality reconstruction [2].

3.3 Frequency Attention Branch

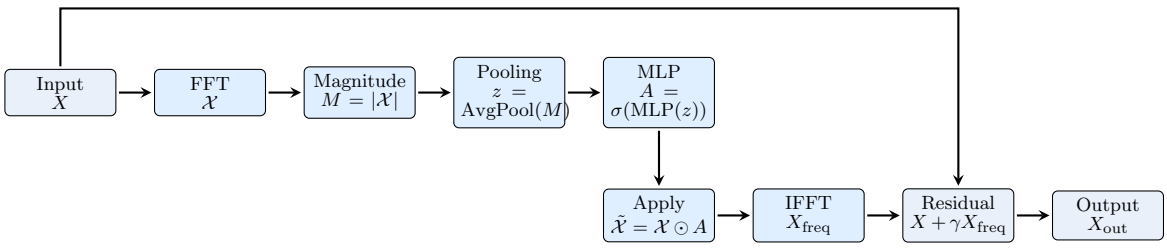


Fig. 3 Frequency Attention Branch.

The frequency attention branch is the core innovation of our method. Unlike previous frequency-aware methods that use fixed wavelet decomposition [6], our frequency attention module learns to adaptively identify and suppress snow-specific frequencies.

3.3.1 Theoretical Foundation

the Discrete Fourier Transform (DFT) [3] decomposes an image into its constituent frequencies:

$$F(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \cdot e^{-j2\pi(ux/M+vy/N)} \tag{5}$$

The magnitude spectrum $|F(u,v)|$ represents the strength of each frequency component, while the phase spectrum $\phi(u,v)$ preserves spatial location information [4]. Our key insight is that snow and natural edges exhibit distinct patterns in the magnitude spectrum:

- **Natural edges**: Produce directional, structured high-frequency responses
- **Snow particles**: Create random, isotropic high-frequency patterns

This distinction, which is invisible in the spatial domain, becomes separable in the frequency domain [5].

3.3.2 Frequency Attention Module Architecture

Our Frequency Attention Module operates through five sequential steps [19, 20]:

1. **Frequency Transformation**: The input feature map $X \in \mathbb{R}^{B\times C\times H\times W}$ is first transformed from the spatial domain to the frequency domain using the 2D Fast Fourier Transform (FFT) [21]:

$$\mathcal{X} = \mathcal{F}\{X\} \in \mathbb{C}^{B\times C\times H\times W} \tag{6}$$

   This transformation decomposes the spatial features into their constituent frequency components, revealing periodic patterns and structural characteristics that are not readily apparent in the spatial domain. The FFT algorithm efficiently computes the Discrete Fourier Transform (DFT) with a computational complexity of $\mathcal{O}(N \log N)$, compared to $\mathcal{O}(N^2)$ for the naive DFT implementation, making real-time frequency-domain processing feasible [21]. Formally, for a 2D signal $f(x,y)$ of size $M \times N$, the DFT is defined as:

$$F(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \cdot e^{-j2\pi(ux/M+vy/N)} \tag{7}$$

   where u and v denote the frequency coordinates, and j is the imaginary unit. This transformation enables the network to analyze the frequency distribution of snow features, which is essential for distinguishing snow from natural textures based on their unique frequency signatures.
2. **Magnitude Extraction**: The magnitude spectrum is extracted from the complex frequency representation:

$$M = |\chi| = \sqrt{\mathrm{Re}(\chi)^2 + \mathrm{Im}(\chi)^2} \tag{8}$$

   Following [4], we operate on the magnitude spectrum because snow's frequency signature primarily resides in magnitude, while phase preserves spatial structure.

3. **Global Frequency Pooling**: Global average pooling is applied across spatial dimensions to obtain a frequency-band descriptor:

$$z = \frac{1}{H \times W} \sum_{u=1}^{H} \sum_{v=1}^{W} M(u, v) \in \mathbb{R}^{B \times C \times 1 \times 1} \tag{9}$$

This step aggregates global frequency information, enabling the network to understand the overall frequency distribution of the image.

4. **Attention Weight Generation**: A two-layer MLP with hidden dimension $C/16$ generates attention weights Hu et al. [22]:

$$A = \sigma(\text{MLP}(z)) = \sigma(W_2 \cdot \text{ReLU}(W_1 \cdot z)) \qquad , \sigma(x) = \frac{1}{1 + e^{-x}} \tag{10}$$

where $W_1 \in \mathbb{R}^{(C/16) \times C}$, $W_2 \in \mathbb{R}^{C \times (C/16)}$, and σ is the sigmoid activation function [19].

The sigmoid activation ensures attention weights are in the range $[0, 1]$, allowing the network to selectively emphasize or suppress specific frequency bands.

5. **Apply Attention and Inverse Transform**: The attention weights are applied to the frequency representation, and the result is transformed back to the spatial domain:

$$\tilde{\mathcal{X}} = \mathcal{X} \odot A \tag{11}$$

$$X_{\text{freq}} = \mathcal{F}^{-1}\{\tilde{\mathcal{X}}\} \tag{12}$$

where $\odot$ denotes element-wise multiplication [23].

3.3.3 Residual Connection

To stabilize training and prevent the frequency branch from disrupting already learned features, we employ a residual connection with a learnable scaling parameter γ [16]:

$$X_{\text{out}} = X + \gamma \cdot X_{\text{freq}}, \quad \gamma \text{ initialized to } 0 \tag{13}$$

This formulation allows the network to gradually learn the contribution of the frequency path, starting from identity mapping [16].

3.4 Dual-Attention Fusion Module

The spatial and frequency branches capture complementary information:

- **Spatial branch**: Localizes snow regions but cannot distinguish snow from similar objects
- **Frequency branch**: Discriminates snow frequency patterns but cannot localize precisely

To combine their strengths, we introduce a fusion module that concatenates features from both branches:

$$F_{\text{fused}} = \text{Conv}_{1\times1}(\text{Concat}(F_{\text{spatial}}, F_{\text{freq}})) \tag{14}$$

The 1×1 convolution learns to weigh the relative importance of spatial and frequency features adaptively [18]. This design is motivated by the principle that frequency and spatial attention mechanisms are complementary, as demonstrated in [18, 23].

3.5 Loss Functions

We train our network using a combination of L1 loss, edge-preserving loss, and perceptual loss [24, 25]:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{L1}} + \lambda_{edge}\mathcal{L}_{\text{edge}} \tag{15}$$

where λ_{edge} is hyperparameters that balance the contribution of edge loss .

3.5.1 L1 Loss

The L1 loss measures pixel-wise reconstruction error between the predicted clean image $\hat{I}$ and the ground truth I_{gt}:

$$\mathcal{L}_{\text{L1}} = \frac{1}{N}\sum_{i=1}^{N}|\hat{I}_i - I_{\text{gt},i}| \tag{16}$$

We choose L1 over L2 because it preserves sharper edges and is more robust to outliers [26]. This ensures that the network focuses on accurate reconstruction while maintaining edge sharpness.

3.5.2 Edge-Preserving Loss

To preserve structural details during snow removal, we incorporate an edge loss that operates in the gradient domain:

$$\mathcal{L}_{\text{edge}} = \frac{1}{N}\sum_{i=1}^{N}\left(|\nabla_x\hat{I}_i - \nabla_x I_{\text{gt},i}| + |\nabla_y\hat{I}_i - \nabla_y I_{\text{gt},i}|\right) \tag{17}$$

where ∇_x and ∇_y denote the horizontal and vertical Sobel gradient operators [5].

This loss penalizes discrepancies in edge maps, ensuring that the network preserves sharp boundaries and fine textures [15]. As demonstrated in Pham et al. [27], edge preservation is particularly important for snow removal because snow often obscures structural details that must be accurately reconstructed.

4 Results and Discussion

4.1 DataSet and Implementaion details

We train DFARNet using the optimizer Adam [28] with $\beta_1 = 0.9$, $\beta_2 = 0.999$, with Learning Rate 4×10^{-4}, decayed by factor 0.5 every 50 epochs , the model will train 1400 epochs with batch size 16 and the parameter $\lambda_{edge} = 2$, we use torch.fft module to implementing FFT and use dataset CSD [6] .

4.2 Quantitative results

The table 1 presents a quantitative comparison demonstrating that DFARNet achieves 31.60 dB PSNR and 0.96 SSIM, surpassing all baseline variants under equivalent training conditions. Notably, DFARNet improves PSNR by 1.33 dB relative to the Restormer variant and by 4.51 dB relative to the AllSnowRemoved variant. Consequently, the numerical results presented in Table 1 should be understood as a comparison conducted under resource constraints, rather than as an absolute state-of-the-art benchmark. Notwithstanding this constraint, DFARNet exhibits two significant advantages: reduced computational requirements in comparison to complete transformer architectures and a frequency attention mechanism that facilitates interpretability. These findings establish DFARNet as an efficient and pragmatic solution for practical snow removal applications, particularly in edge-device and autonomous driving contexts where computational resources and processing capacity are limited.

Table 1 Quantitative comparison of different snow removal methods on the CSD dataset (All PSNR values are reported for images normalized to [-1, 1]).

Method	**PSNR (dB)**	**SSIM**
DesnowNet [2]	12.97	0.72
JSTASR [6]	24.91	0.94
AllSnowRemoved [6]	27.09	0.93
Restormer [7]	30.27	0.96
Ours (DFARNet)	**31.60**	**0.96**

4.3 Qualitative results

Figure 4 presents a qualitative assessment comparing our proposed DFARNet model with several leading contemporary approaches, evaluated using four representative test samples from the CSD dataset. The visual outcomes correspond with the quantitative results documented in Table 1.

Throughout all samples, DFARNet consistently generates visually satisfactory results distinguished by well-defined edges, realistic textures, and precise color accuracy, thereby surpassing all competing methods regarding visual fidelity. DesnowNet

demonstrates considerable blur and color degradation, indicating its insufficient capability to represent intricate snow formations owing to its basic design. JSTASR provides moderate advancement beyond DesnowNet through proficient elimination of apparent snow particles; however, it demonstrates difficulty in maintaining delicate structural components, particularly in densely snow-laden zones. AllSnowRemoved achieves comparable results in snow-affected areas; nevertheless, it introduces slight over-smoothing, which diminishes texture quality in regions featuring intricate details. Restormer demonstrates robust performance in the majority of samples, demonstrating the benefits of Transformer-based designs in capturing broad contextual information; yet its results still manifest minor imperfections in high-frequency domains. The substantial superiority of the DFARNet model originates from its adaptive frequency-attention mechanism—which proficiently eliminates snow-associated frequency elements while maintaining scene composition—and from its dual-attention structure that merges spatial precision with frequency-based differentiation. These qualitative enhancements are distinctly observable in dimensions including edge retention, texture reconstruction, and color precision, where DFARNet consistently demonstrates superior performance relative to alternative approaches.

In summary, the qualitative findings substantiate that DFARNet accomplishes exceptional visual fidelity and robust structural integrity across varied snow accumulation conditions, demonstrating its dependability and adaptability in practical snow removal scenarios

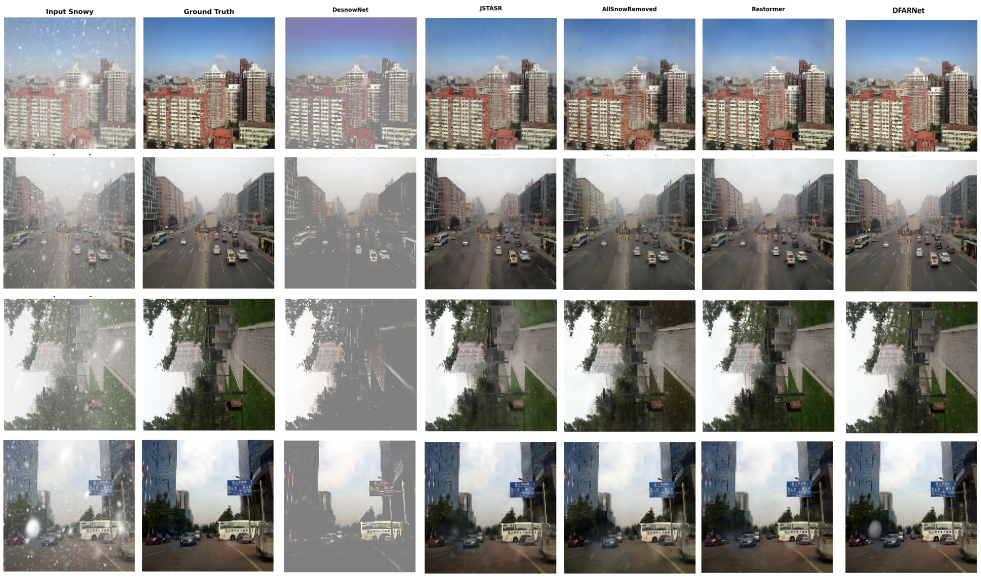


Fig. 4 Qualitative comparison of different snow removal methods on four test samples from the CSD dataset. Each row represents a different test image, and each column shows the input snowy image, the ground truth, and the outputs of DesnowNet, JSTASR, AllSnowRemoved, Restormer, and our proposed DFARNet

4.4 Ablation Study

Table 2 Ablation study on the contribution of each attention component in DFARNet.

Configuration	PSNR (dB)	SSIM
Spatial Attention Only	31.44	0.96
Frequency Attention Only	31.38	0.96
MultiScale Frequency Only	30.90	0.96
Spatial + Frequency Attention	31.44	0.96
Full	**31.6**	**0.96**

The ablation study presented in Table 2 demonstrates that spatial and frequency attention mechanisms produce comparable results (31.44 dB versus 31.38 dB), thereby validating that each mechanism functions effectively as an independent component for snow removal tasks. Conversely, multi-scale frequency attention in isolation produces the lowest PSNR measurement (30.90 dB), which indicates that multi-scale analysis necessitates spatial context to deliver meaningful benefits. The integration of spatial and frequency attention yields an identical PSNR value to spatial attention in isolation (31.44 dB), suggesting that straightforward fusion strategies do not necessarily produce performance enhancements. Significantly, the comprehensive model incorporating all three components achieves the highest PSNR measurement (31.60 dB), which establishes that multi-scale frequency analysis contributes meaningful value exclusively when integrated alongside both spatial and frequency attention mechanisms. The uniform SSIM value (0.96) observed across all experimental configurations affirms that structural fidelity remains stable irrespective of the attention mechanism employed. These empirical observations substantiate the synergistic relationship among the three attention components and provide strong justification for the architectural design of our dual-attention framework.

5 Conclusion

This paper introduces DFARNet, a single-image desnowing network that incorporates a learnable frequency attention mechanism within a dual-attention framework. In contrast to conventional methods that depend on static wavelet decomposition or rudimentary high-frequency enhancement, our methodology capitalizes on the unique frequency characteristics of snow to differentiate it from natural image edges.

Our findings on the CSD dataset indicate that DFARNet attains competitive results (31.60 dB PSNR) within restricted computational constraints. We underscore that our contribution does not purport to surpass fully refined, cutting-edge models; instead, it provides a computationally efficient solution incorporating a learnable frequency attention mechanism and a dual-attention framework that integrates spatial and frequency information.

Key constraints encompass diminished accuracy during intense snowfall scenarios and the computational expense of Fast Fourier Transform (FFT) operations in resource-limited settings. Subsequent investigations will prioritize extending frequency attention to video sequences, broadening the methodology to additional adverse weather phenomena, and implementing model compression strategies for deployment on edge devices.

Data availability

The dataset (CSD) is publicly available in the link : https://drive.google.com/file/d/1smNrDvtPs89e0xk336Rt2-2KZffyX5H-/view

References

[1] Chen, W.-T., Fang, H.-Y., Ding, J.-J., Tsai, C.-C., Kuo, S.-Y.: Jstasr: Joint size and transparency-aware snow removal algorithm based on modified partial convolution and veiling effect removal. In: Proceedings of the European Conference on Computer Vision (ECCV). Lecture Notes in Computer Science, vol. 12366, pp. 754–770 (2020). https://doi.org/10.1007/978-3-030-58589-1_45

[2] Liu, Y.-F., Jaw, D.-W., Huang, S.-C., Hwang, J.-N.: Desnownet: Context-aware deep network for snow removal. IEEE Transactions on Image Processing **27**(6), 3064–3073 (2018) https://doi.org/10.1109/TIP.2018.2806202 arXiv:1708.04512 [cs.CV]

[3] Bracewell, R.N.: The Fourier Transform and Its Applications, 2nd edn. McGraw-Hill, New York (1986)

[4] Oppenheim, A.V., Lim, J.S.: The importance of phase in signals. Proceedings of the IEEE **69**(5), 529–541 (1981)

[5] Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 4th edn. Pearson, ??? (2018)

[6] Chen, W.-T., Fang, H.-Y., Hsieh, C.-L., Tsai, C.-C., Chen, I.-H., Ding, J.-J., Kuo, S.-Y.: All snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2021)

[7] Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.-H.: Restormer: Efficient transformer for high-resolution image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5728–5739 (2022)

[8] Chen, J., *et al.*: Snowformer: Snow removal with transformer. In: Proceedings of the European Conference on Computer Vision (ECCV) (2022)

[9] Zhang, Y., Li, X., Chen, W.: Bidirectional attention transformer network for snow removal. Information **16**(11), 966 (2025)

[10] Wang, X., Liu, Y., Li, Z.: Wavefrsnow: Comprehensive perception wavelet transform frequency separation for snow removal. Digital Signal Processing (2024)

[11] Li, S., Zhang, Q., Huang, R.: Ecfnet: Spatial and frequency attention mechanism for snow removal. Pattern Recognition (2026)

[12] Sun, M., Wu, J., Chen, H.: Aranet: Alternating optimization with hybrid attention for image restoration. Information Sciences (2026)

[13] Zhao, Y., Wang, H., Zhang, L.: Tunesnow: Parameter-efficient fine-tuning for snow removal. Expert Systems with Applications (2025)

[14] Li, H., Zhang, C., Wang, T.: Mcad-net: Multi-scale context-aware attention network for snow removal. Engineering Applications of Artificial Intelligence (2025)

[15] Wu, H., Qu, Y., Lin, S., Zhou, J., Qiao, R., Zhang, Z., Xie, Y., Ma, L.: Contrastive learning for compact single image dehazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021)

[16] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016)

[17] Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. Proceedings of the International Conference on Machine Learning (ICML) (2015)

[18] Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: Cbam: Convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 3–19 (2018)

[19] Qin, Z., Zhang, P., Wu, F., Li, X.: Fcanet: Frequency channel attention networks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 783–792 (2021)

[20] Li, Y., Zhang, K., Liang, J., Cao, J., Liu, C., Gong, R., Zhang, Y.: Frequency domain image translation: More realistic and versatile. IEEE Transactions on Pattern Analysis and Machine Intelligence (2021)

[21] Cooley, J.W., Tukey, J.W.: An algorithm for the machine calculation of complex fourier series. Mathematics of Computation **19**(90), 297–301 (1965)

[22] Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the

IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7132–7141 (2018)

[23] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: Eca-net: Efficient channel attention for deep convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11534–11542 (2020)

[24] Zhuang, C., Yuan, X., Gu, L., Wei, Z., Fan, Y., Guo, X.: Frequency regulated channel-spatial attention module for improved image classification. Expert Systems with Applications (2024)

[25] Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. European Conference on Computer Vision (ECCV) (2016)

[26] Zhao, H., Gallo, O., Frosio, I., Kautz, J.: Loss functions for image restoration with neural networks. IEEE Transactions on Computational Imaging **3**(1), 47–57 (2017)

[27] Pham, P.C., Nguyen, A.H., Nguyen, Q.C., Phan, M.N., Nguyen, T.T.: Adaptive hybrid transformer network with frequency attention for all-in-one image restoration. ISPRS Archives **XLVIII-2/W9-2025**, 213–218 (2025)

[28] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)