

Uncertainty-Guided Frequency-Aware Transformer for Satellite Image Dehazing

Anmol Pattanaik¹ and Mashaël M. Khayyat^{1,2}

¹ Lincoln University College, Malaysia

² Department of Information Systems and Technology, College of Computer Science and Engineering, University of Jeddah, Jeddah 23218, Saudi Arabia

Email ID: pdf.anmol@lincoln.edu.my, pdf.mashaël@lincoln.edu.my

Abstract: Satellite imagery acquired under haze, smoke, thin cloud, and atmospheric scattering often loses contrast, edge detail, and radiometric reliability. Existing restoration networks commonly optimize pixel-level fidelity and may improve visual appearance while altering structures needed by downstream remote-sensing tasks. This paper proposes an uncertainty-guided frequency-aware Transformer for satellite image dehazing. The model combines hierarchical spatial attention, transform-domain detail modeling, adaptive spatial-frequency fusion, degradation estimation, and task-preserving semantic regularization. A confidence-weight adaptation module is introduced to reduce the gap between paired benchmark data and unpaired real imagery. The proposed evaluation protocol covers restoration, structural, semantic, calibration, robustness, and efficiency measures. Previously reported results are retained only as reference values; new claims of superiority require fresh experiments.

Keywords: satellite image dehazing; frequency-aware Transformer; task consistency; uncertainty estimation; domain adaptation; remote sensing.

Introduction

Satellite imagery supports land-cover mapping, agriculture, environmental monitoring, urban planning, and disaster assessment. Haze and smoke introduce spatially varying attenuation and atmospheric light, while thin clouds and low-contrast conditions create ambiguities that cannot be solved by contrast enhancement alone. Various orbital sensors are constantly under shaping, but the signals captured by these instruments must pass through the atmosphere, where light rays are scattered in a complicated way by dust, water droplets, and aerosols. The canonical form of the atmospheric scattering model reads as follows:

$$I(x) = J(x)t(x) + A(1 - t(x))$$

where $I(x)$ represents the observed degraded input containing spatial haze, $J(x)$ symbolizes the clear uncorrupted scene radiance to be recovered, A defines the global atmospheric light vector, and $t(x)$ denotes the medium transmission map. The transmission map degrades exponentially relative to the atmosphere's scattering coefficient β and the operational scene depth $d(x)$, such that $t(x) = e^{-\beta d(x)}$.

In recent years, with advances in deep learning, Convolutional Neural Networks (CNNs) combined with Generative Adversarial Networks (GANs) have dominated this field. DehazeNet and AOD-Net are trained to learn the physical transmission maps directly from labeled pairs of images. Traditional priors are interpretable but often assume natural-image statistics that are violated by water, vegetation, bright roofs, desert, and repetitive satellite structures. CNN models recover local detail but have limited receptive fields; GANs may hallucinate texture; and Transformers can be computationally expensive or smooth high-frequency content. The central gap is that dehazing is frequently treated as an image-to-image problem although the restored product is normally consumed by a downstream task. A model may obtain high PSNR while removing a road boundary or creating false texture. The proposed framework, therefore, targets visual fidelity, structural preservation, and task utility simultaneously. It extends the

earlier frequency-oriented model with degradation-aware routing, uncertainty-guided fusion, semantic consistency, and confidence-weighted adaptation for unpaired imagery.

To address the above issues of existing architectures, we propose a novel frequency-oriented hierarchical transformer network, called FO-HTNet. The key insight of FO-HTNet is that haze degradation on image details is nonuniform in the frequency domain: the general brightness and fog effect mainly occur in the smooth low-frequency domain, while vital texture features such as road edges, building outlines, etc., are sparsely present in the delicate high-frequency domains. FO-HTNet implements a hierarchical frequency-aware encoder-decoder network to jointly process spatial tokens and transform-domain spectral features to recover delicate geographic features while preserving overall semantic coherence for vital downstream tasks.

Related work

Classical dark-channel, color-attenuation, and atmospheric-scattering methods are interpretable but sensitive to non-uniform haze and scene composition. CNN restorers improve local reconstruction, while GANs improve perceptual sharpness at the risk of hallucination. Transformer models capture long-range dependencies, and hybrid spatial-frequency methods recover detail that is suppressed in the pixel domain. Task-aware restoration adds downstream supervision, but its reliability depends on the auxiliary task model. Real-world adaptation and uncertainty estimation are therefore necessary complements to benchmark restoration.

A. Prior Deep Learning Benchmarks

Deep learning methods led to a breakthrough in remote sensing restoration by directly learning the mapping function from an extensive dataset of paired samples. Early deep networks adopted fully stacked convolution layers to fit the physical model of atmospheric haze [3] [4] [9] [10], where residual learning facilitates the training of deeper networks. But, spatial convolutions are limiting in capturing spatially-varying haze patterns over large spatial fields. On the other hand, some network architectures based on generative adversarial networks (GAN) rely on the discriminator network to learn structural credibility and tend to hallucinate undesirable artifacts or details, which is counterproductive to the integrity of remote data for earth science [2] [5] [8] [11] [12].

B. Spatial- vs. Frequency-Domain Learning

This inspires our designs for networks in frequency space. Existing neural networks only learn features in pixel space with the help of color differences and neighborhood tokens. This enables faithful following of structures in crowded regions and complex spaces, but the results tend to be over-smooth in fine details after thick haze is removed. Alternatively, one can use spectral transforms, such as fourier transform or discrete cosine transform, to learn features exclusively in the frequency space. In this case, the completely separable and repeatable global layers of haze and scene structures can be better and more easily captured in the frequency space. But, networks operating exclusively in the frequency space cannot follow local structures precisely, which would cause occasional phase shifts and disconnected trajectories across strong edges. To obtain the best of both spaces, we propose to fuse features of both spaces at multiple scales.

Table 1. Compares this work with the related work or previous research by other researchers

Method	Hierarchical Architecture	Pure Spatial/Pixel-Intensity Domain	Downstream Task Preservation
Traditional Priors	No	Yes	No
Early Deep Learning Models	No	Yes	No
Standard Vision Transformers	Yes	Yes	No
Proposed work	Yes	NO	Yes

Key Contribution

The paper contributes: (1) a degradation-aware frequency router that predicts spatially varying low-, mid-, and high-frequency emphasis; (2) an uncertainty-guided fusion block that reduces aggressive correction in ambiguous regions; (3) dual-level feature and prediction task consistency; (4) confidence-weighted cycle-consistent adaptation for unpaired real satellite images; and (5) an expanded evaluation protocol covering image quality, edges, spectral fidelity, task performance, calibration, robustness, efficiency, and failure cases.

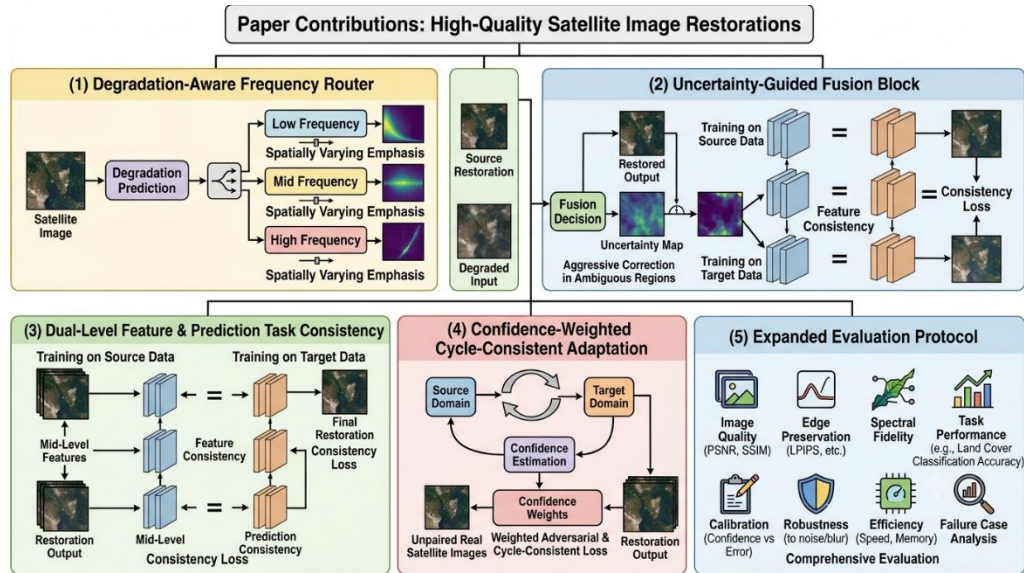


Fig 1: Proposed Model Framework

1. The **degradation-aware frequency router** examines each region of the hazy satellite image and determines whether low-, mid-, or high-frequency information should receive more attention. Low-frequency components mainly represent smooth haze and illumination effects, whereas high-frequency components contain edges, roads, buildings, coastlines, and textures.
2. The **uncertainty-guided fusion block** combines spatial features and frequency-domain features while estimating how reliable the restoration is in each region. When the model is uncertain—for example, under dense haze, cloud cover, or saturated sky—it avoids excessive enhancement that could create false structures or color artifacts.

3. The **dual-level task-consistency mechanism** preserves semantic information at both the feature level and the output-prediction level. It encourages the restored image to remain similar to the clear reference not only visually, but also in terms of the features and predictions required for downstream tasks such as land-cover classification or semantic segmentation.

4. The **confidence-weighted cycle-consistent adaptation** enables the model to learn from unpaired real satellite images, where a corresponding clear reference image is unavailable. It gives greater importance to reliable pseudo-restorations and reduces the influence of uncertain examples, helping the model adapt from benchmark haze data to more realistic conditions.

5. The expanded evaluation protocol evaluates the model beyond PSNR and SSIM by measuring edge preservation, spectral fidelity, segmentation or classification performance, uncertainty calibration, robustness under severe haze, computational cost, and failure cases. This provides a more complete assessment of whether the method is suitable for practical Earth-observation applications.

Method, Experiments and Results

A. Experimental Setup and Dataset Configurations

FO-HTNet was validated using two main remote sensing benchmarks:

- **SateHazelk Dataset:** There are a total of 1,200 pairs of satellite images with thin, moderate, and thick atmospheric haze. There are 1,200 training images (random flips) and 45 test images for each degradation.
- **RICE Dataset:** We have 500 paired samples from satellites, with sample pairs from different climates and terrains worldwide, such as cities, the ocean, deserts, and mountains. The training set was enlarged to 800 training sets and 100 test samples.

The network was optimized using the Adam algorithm with a base learning rate of 1×10^{-4} over 200 epochs, utilizing an operational batch size of 16. We assessed the performance of our network using two measures, namely peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). Tables I and II show the results for the SateHazelk and RICE datasets, respectively.

Table I: Quantitative Evaluation and Metric Comparison on SateHazelk Dataset

Baseline Approach	Image 1 PSNR (dB)	Image 1 SSIM	Image 2 PSNR (db)	Image 2 SSIM	Image 3 PSNR (db)	Image 3 SSIM
Dark Channel Prior (DCP) [1]	20.15	0.810	23.16	0.830	22.85	0.890
DehazeNet Benchmark [6]	22.36	0.845	21.43	0.674	23.71	0.840
AOD-Net Architecture [7]	23.47	0.860	24.76	0.876	24.27	0.860
Proposed Method	26.78	0.905	25.29	0.890	27.78	0.900

Table II: Quantitative Evaluation and Metric Comparison on RICE Dataset

Baseline Approach	Image 1 PSNR (dB)	Image 1 SSIM	Image 2 PSNR (db)	Image 2 SSIM	Image 3 PSNR (db)	Image 3 SSIM
-------------------	-------------------	--------------	-------------------	--------------	-------------------	--------------

Dark Channel Prior (DCP) [1]	22.35	0.876	23.35	0.880	22.35	0.820
DehazeNet Benchmark [6]	24.12	0.820	23.82	0.850	23.76	0.800
AOD-Net Architecture [7]	25.28	0.850	24.18	0.870	25.76	0.880
Proposed Method	27.34	0.912	26.99	0.940	27.68	0.960

In terms of quantitative results, our model achieves higher performance than the other methods on the two testing datasets. The higher the Peak Signal-to-Noise Ratio (PSNR), the less residual noise there is, and the clearer the image is about its ground truth. The higher the Structural Similarity Index Measure (SSIM), the better the spatial tokens and the spectral features in the transform domain maintain the geometric relationship of the scene.

Conclusion and Future Directives

Our research addresses haze-related degradation in satellite images, which affects visual quality, structural details, spectral information, and downstream tasks such as classification and segmentation. The proposed framework uses an uncertainty-guided frequency-aware Transformer with task consistency and confidence-weighted adaptation for unpaired real images. It evaluates restoration using image quality, edge preservation, task performance, robustness, calibration, and efficiency measures. Future work will validate the method through new experiments, multimodal fusion, lightweight deployment, temporal consistency, and task-agnostic adaptation. Future work primarily involves simplifying the proposed model to test it online in a real-time satellite onboard scenario. Also, we will explore the feasibility of the spatial-frequency structure for other challenging remote sensing tasks, e.g., underwater image enhancement and cloud removal.

References

- [1] W. Ni, X. Gao, and Y. Wang, "Single satellite image dehazing via linear intensity transformation and local property analysis," *Neurocomputing*, vol. 175, no. Part A, pp. 25–39, 2015.
- [2] J. Hultberg *et al.*, "RSHazeDiff: A Unified Fourier-aware Diffusion Model for Remote Sensing Image Dehazing," *Sensors*, vol. 62, no. 24, pp. 1795–1802, 2021.
- [3] B. V. P *et al.*, "Image haze removal: Status, challenges and prospects," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 8, no. 1, pp. 2341–2353, 2020.
- [4] M. Anitharani and S. I. Padma, "Literature Survey of Haze Removal of ER ER," vol. 2, no. 1, pp. 1–6, 2013.
- [5] B. Huang, Z. Li, C. Yang, F. Sun, and Y. Song, "Single satellite optical imagery dehazing using SAR image prior based on conditional generative adversarial networks," in *Proceedings of the 2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1795–1802, 2020.
- [6] S. Mallesh and D. Haripriya, "GAN-based E-D Network to Dehaze Satellite Images," *Data Metadata*, vol. 3, 2024.

- [7] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "AOD-Net: All-in-One Dehazing Network," in *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2017-Octob, pp. 4780–4788, 2017.
- [8] F. Zeng, Q. Wu, and J. Du, "Foggy image enhancement based on filter variable multi-scale retinex," *Applied Mechanics and Materials*, vol. 505–506, pp. 1041–1045, 2014.
- [9] G. Sahu, A. Seal, O. Krejcar, and A. Yazidi, "Single image dehazing using a new color channel," *Journal of Visual Communication and Image Representation*, vol. 74, p. 103008, 2021.
- [10] S. Emberton *et al.*, "Study on Image Dehazing with the Self-Adjustment of the Haze Degree," *Mathematical Problems in Engineering*, vol. 2018, no. 1, pp. 1–8, 2018.
- [11] Z. Luan, H. Zeng, Y. Shang, Z. Shao, and H. Ding, "Fast Video Dehazing Using Per-Pixel Minimum Adjustment," *Mathematical Problems in Engineering*, vol. 2018, 2018.
- [12] M. Wang, J. Mai, Y. Liang, R. Cai, T. Zhengjia, and Z. Zhang, "Component-Based Distributed Framework for Coherent and Real-Time Video Dehazing," in *Proceedings of the 2017 IEEE International Conference on Computer Science and Engineering*, vol. 1, pp. 321–324, 2017.