Explainable and Reliable Image Dehazing: A Survey of Deep Learning Approaches

Dr. Pulkit Dwivedi[1], Prof. (Dr.) Shashi Kant Gupta[1]
[1] Lincoln University College, 47301, Petaling Jaya, Selangor Darul Ehsan, Malaysia
Emails: pdwivedi1990@gmail.com , shashigupta@lincoln.edu.my

Abstract: Image dehazing has witnessed significant advancements with the adoption of deep learning techniques, particularly convolutional neural networks, generative adversarial networks, and transformer-based architectures. However, despite improvements in visual quality and quantitative performance, existing approaches often lack reliability and interpretability, which limits their deployment in real-world safety-critical applications. This survey presents a comprehensive review of deep learning-based image dehazing methods with a specific focus on explainability and reliability. We systematically categorize existing approaches into prior-based, CNN-based, GAN-based, and transformer-based methods, and analyze their strengths and limitations. Furthermore, we highlight the emerging need for confidence-aware modeling and explainable frameworks to enhance trustworthiness. Our analysis reveals that while GANs and transformers improve perceptual quality, they remain largely black box in nature. The survey also identifies key research gaps, including the lack of standardized evaluation for interpretability and limited real-world generalization. Potential applications include autonomous driving, surveillance, remote sensing, and medical imaging. This work aims to guide future research toward developing robust, interpretable, and deployment-ready dehazing solutions.

Keywords: Image Dehazing; Deep Learning; Explainable AI; Reliability; Generative Adversarial Networks

---

1. Introduction

Image dehazing is a fundamental problem in computer vision, aimed at restoring visibility in images degraded by atmospheric scattering and absorption. The presence of haze significantly reduces image contrast and distorts scene radiance, thereby affecting the performance of downstream vision tasks such as object detection, segmentation, and recognition. Traditional approaches rely on physical priors such as the Dark Channel Prior (DCP), which assume statistical properties of haze-free images. While effective in controlled environments, these methods struggle in complex real-world scenarios involving non-uniform haze and dynamic lighting conditions. Recent advances in deep learning have enabled data-driven solutions for image dehazing. Convolutional neural networks (CNNs), generative adversarial networks (GANs), and transformer-based models have demonstrated significant improvements in restoration quality. However, these models often function as black-box systems, lacking interpretability and reliability.

Motivated by these limitations, this survey focuses on analyzing deep learning-based dehazing approaches from the perspective of explainability and reliability, which are essential for real-world deployment. The problem statement aligns with the abstract by emphasizing the need for trustworthy dehazing systems.

2. Related Work

Image dehazing has evolved from prior-based formulations to deep learning frameworks, including CNNs, GANs, and transformers, which have significantly improved restoration quality. However, challenges related to reliability in real-world conditions and model interpretability remain largely unresolved. Early methods relied on physical models such as the Dark Channel Prior (DCP) [1], which estimates transmission using statistical assumptions. While interpretable, such approaches fail in bright regions and lack robustness. Subsequent feature-based methods [2] improved performance but remained dependent on handcrafted assumptions.

The shift to deep learning introduced data-driven approaches. DehazeNet [3] and AOD-Net [4] enabled learning-based and end-to-end dehazing, offering improved efficiency. Multi-task models such as DCPDN [5] further enhanced performance but remained black box in nature, limiting interpretability. GAN-based methods such as Cycle-Dehaze [6] and FFA-Net [7] improved perceptual quality and texture restoration, particularly in unpaired settings. However, they suffer from instability, artifacts, and lack of transparency, raising concerns about reliability. Transformer-based models like TransWeather [8] capture global dependencies and perform well under complex haze conditions, with further improvements through hybrid designs [9]. Despite their effectiveness, these models are computationally expensive and remain difficult to interpret.

A major challenge across all methods is the domain gap between synthetic datasets such as RESIDE [10] and real-world scenarios. Approaches such as domain adaptation and test-time training [11] aim to address this issue but remain limited. Additionally, while recent works explore interpretable components [12], there is no standardized framework for evaluating explainability or reliability. Most models continue to prioritize visual quality over trustworthy and transparent predictions.

Recent advancements in image dehazing focus on improving real-world generalization through hybrid, physics-guided, and data-efficient frameworks. Ali *et al.* [13] proposed Dehazing-DiffGAN, combining diffusion and GAN models to achieve high-quality restoration, while Liu *et al.* [14] introduced IHDCP, integrating physical priors for efficient and interpretable dehazing. To address domain gaps, Chen *et al.* [15] developed a self-supervised framework that enhances robustness without paired data. Hybrid models such as DAH-TrafficRSNet [16] and physics-based decomposition methods [17] further improve structural preservation and generalization. Additionally, CLIP-DQA V2 [18] advances blind quality assessment using multimodal learning, while recent surveys [19] emphasize real-time and deployment challenges. Overall, these works highlight a shift toward hybrid and data-driven approaches, though issues related to interpretability, reliability, and standardized evaluation remain open.

3. Key Contributions

1. This survey presents a comprehensive and structured taxonomy of image dehazing methods, categorizing them into prior-based, CNN-based, GAN-based, and transformer-based approaches.
2. This work emphasizes explainability by analyzing the interpretability aspects of dehazing models, including attention mechanisms and physics-guided representations.
3. This survey provides a reliability-oriented analysis by examining the robustness and generalization of existing methods under diverse and real-world haze conditions.
4. A detailed comparative evaluation is conducted using standard metrics such as PSNR, SSIM, NIQE, and BRISQUE, along with qualitative insights into model performance.
5. The study identifies key research gaps, including the lack of interpretability, absence of confidence-aware modeling, and limited real-world generalization.

4. Method, Experiments and Results

4.1 Evaluation Metrics

Image dehazing methods are evaluated using both full-reference and no-reference metrics. Full-reference metrics such as PSNR and SSIM measure reconstruction fidelity and structural similarity when ground truth images are available. However, since real-world haze-free references are often unavailable, no-reference metrics including NIQE, BRISQUE, and FADE are widely used to assess perceptual quality, naturalness, and haze density in practical scenarios.

4.2 Comparative Analysis

A comparative analysis shows that GAN-based methods achieve superior perceptual quality by generating visually realistic outputs, particularly in challenging haze conditions. CNN-based approaches, on the other hand, are computationally efficient and suitable for real-time applications but are limited in capturing global context. Transformer-based models address this limitation through self-attention mechanisms, enabling better handling of complex and non-uniform haze, although at the cost of increased computational complexity.

4.3 Observations

Overall, GAN-based methods improve visual realism but lack interpretability and may introduce artifacts, while CNN-based models are efficient but less effective in dense haze scenarios. Transformer-based approaches provide strong performance but are resource-intensive. Across all categories, explainability remains largely underexplored, highlighting the need for developing reliable and interpretable dehazing frameworks for real-world applications.

5. Discussion and Conclusions

Image dehazing has evolved from prior-based methods to advanced deep learning frameworks, significantly improving restoration quality; however, several challenges remain. A key observation is that performance gains are often driven by architectural complexity rather than a deeper understanding of haze formation, resulting in limited generalization under real-world conditions. The domain gap between synthetic datasets such as RESIDE and real-world scenarios continues to hinder robustness. Additionally, most deep learning models, including CNN, GAN, and transformer-based approaches, operate as black boxes, lacking interpretability and confidence estimation, which raises concerns about reliability, especially in safety-critical applications.

From a methodological perspective, GAN-based models achieve high perceptual quality but may introduce artifacts, CNN-based methods are efficient but limited in capturing global context, and transformer-based approaches provide strong performance at the cost of high computational complexity. Furthermore, the absence of standardized evaluation frameworks for explainability and reliability makes objective comparison challenging.

Overall, this survey highlights that despite significant advancements, current dehazing methods prioritize visual quality over trustworthiness and real-world applicability. Future research should focus on developing confidence-aware and explainable dehazing frameworks, improving real-world generalization, designing lightweight architectures for efficient deployment, and establishing standardized benchmarks. Additionally, task-driven dehazing approaches that enhance downstream vision performance represent a promising direction for practical applications.

References

[1] K. He, J. Sun and X. Tang, “Single image haze removal using dark channel prior,” Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2011.

[2] K. Tang, J. Yang and J. Wang, “Investigating haze-relevant features in a learning framework,” Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2014.

[3] B. Cai, X. Xu and K. Jia, “DehazeNet: An end-to-end system for single image haze removal,” IEEE Trans. Image Process., 2016.

[4] B. Li, X. Peng and Z. Wang, “AOD-Net: All-in-one dehazing network,” Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2017.

[5] H. Zhang and V. Patel, “Densely connected pyramid dehazing network,” Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018.

[6] D. Engin, A. Genç and H. K. Ekenel, “Cycle-Dehaze: Enhanced CycleGAN for image dehazing,” Proc. IEEE CVPR Workshops, 2018.

[7] X. Qin, Z. Wang and Y. Bai, “FFA-Net: Feature fusion attention network for image dehazing,” Proc. AAAI Conf. Artif. Intell., 2020.

[8] J. M. Valanarasu and V. M. Patel, “TransWeather: Transformer-based restoration under adverse weather conditions,” Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2022.

[9] Y. Liu et al., “Test-time training for multi-domain dehazing,” Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2022.

[10] B. Li et al., “RESIDE: A benchmark for single image dehazing,” Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018.

[11] Y. Zhang et al., “Density-aware image dehazing,” Proc. Asian Conf. Comput. Vis. (ACCV), 2020.

[12] Q. Xiao et al., “Interpretable image dehazing network,” arXiv preprint, 2022.

[13] A. M. Ali, B. Benjdira, W. Boulila and A. Koubaa, “Dehazing-DiffGAN: Sequential Fusion of Diffusion Models and GANs for High-Fidelity Remote Sensing Image Dehazing,” IEEE Trans. Geosci. Remote Sens., vol. 63, pp. 1–15, 2025.

[14] Y. Liu, T. Li, C. Tan, W. Ren, C. Ancuti and W. Lin, “IHDCP: Single Image Dehazing Using Inverted Haze Density Correction Prior,” IEEE Trans. Image Process., vol. 35, pp. 1448–1461, 2026.

[15] J. Chen, W. Ren, H. Zhao, Q. Xia and G. Yang, “You Only Need Clear Images: Self-Supervised Single Image Dehazing,” IEEE Trans. Multimedia, vol. 27, pp. 5800–5814, 2025.

[16] M. Liu, K. Chi and C. You, “DAH-TrafficRSNet: Dual-Branch Traffic Remote Sensing Image Dehazing Network,” IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens., vol. 18, pp. 25325–25337, 2025.

[17] N. Wang et al., “Weakly Supervised Image Dehazing via Physics-Based Decomposition,” IEEE Trans. Circuits Syst. Video Technol., vol. 36, no. 1, pp. 637–652, 2026.

[18] Y. Zeng, J. Fu, G. Yue, H. Liu and W. Zhou, “CLIP-DQA V2: Exploring CLIP for Dehazed Image Quality Assessment,” IEEE Signal Process. Lett., vol. 32, pp. 3829–3833, 2025.

[19] N. Raju, S. Kankanala, R. Thatikonda and B. M. Chintakindi, “A Review of Real-Time Single Image Dehazing Architectures for Embedded Vision,” IEEE Access, vol. 14, pp. 13264–13304, 2026.