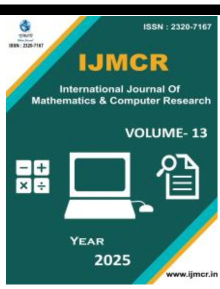
International Journal of Mathematics and Computer Research

ISSN: 2320-7167

Volume 13 Issue 05 May 2025, Page no. – 5151-5155

Index Copernicus ICV: 57.55, Impact Factor: 8.615

DOI: 10.47191/ijmcr/v13i5.02



Video Dehazing Using Denoising Diffusion Probabilistic Model with Trilateral Filtering and Multi-Scale Boosted Dehazing Network

Nisha Amin[1], Geeta B[2], R. L. Raibagkar[3], G. G. Rajput[4]

[1,2]Department of Computer Science, Karnataka State Akkamahadevi Women's University, Vijayapura

[3,4]Department of Applied Electronics, Gulbarga University, Kalaburgi

ARTICLE INFO	ABSTRACT
Published Online: 05 May 2025 **Corresponding Author:** Nisha Amin	Outdoor videos captured in poor weather conditions, such as haze or fog, often suffer from reduced visibility, affecting both cinematography and surveillance applications. The light scattering and absorption caused by aerosols in the atmosphere and airlight reflection result in images with faded colors and decreased contrast. This paper presents a novel approach to video dehazing by leveraging Deep Learning (DL), specifically utilizing the Denoising Diffusion Probabilistic Model (DDPM) for haze removal and a Multi-Scale Boosted Dehazing Network (MSBDN) with Dense Feature Fusion for enhanced image quality. We refine the transmission map using trilateral filtering to achieve smooth edge transitions and improve dehazing performance. Our method is evaluated on both synthetic and real-world datasets, demonstrating its robustness and effectiveness compared to state-of-the-art dehazing algorithms.

KEYWORDS: Deep Learning, DDPM, MSBDN, Trilateral Filtering

1. INTRODUCTION

Outdoor videos captured in poor weather conditions, such as haze or fog, often suffer from reduced visibility, affecting both cinematography and surveillance applications. The light scattering and absorption caused by aerosols in the atmosphere and airlight reflection result in images with faded colors and decreased contrast. This phenomenon, especially in regions prone to fog and haze, presents significant challenges for clear image and video acquisition.

To address this issue, this paper presents a novel approach to video dehazing by leveraging Deep Learning (DL) techniques, particularly the Denoising Diffusion Probabilistic Model (DDPM) for dense haze removal, and a Multi-Scale Boosted Dehazing Network (MSBDN) with Dense Feature Fusion for enhanced dehazing performance. The novelty of our work lies in several key contributions. First, we integrate DDPM, which exhibits strong generative abilities to separate haze from the scene content through its forward and reverse denoising processes. Second, we introduce trilateral filtering to refine the transmission map, ensuring smooth edge transitions and minimizing common dehazing artifacts such as halos.

Third, we propose MSBDN, which captures both local and global features at multiple scales, improving the recovery of fine-grained details. Additionally, by relaxing the traditional assumption of constant atmospheric light, our method allows the network to learn the transformation from hazy image to transmission map, adapting to complex real-world conditions. Finally, we demonstrate the robustness of our approach by applying it to both synthetic and real-world datasets, highlighting its effectiveness in diverse haze scenarios. These contributions collectively form an end-to-end deep learning pipeline that significantly enhances the quality and applicability of video dehazing in various environments.

2. RELATED WORK

Duminil et al. [1] introduce the VIREDA dataset, which includes 18 real-world foggy videos and 6 corresponding clear ground truth videos under varying fog densities and lighting conditions. Depth maps are also provided to support quantitative evaluation. To complement this dataset, the authors propose a hybrid defogging algorithm called TCVD that combines Convolutional Neural Networks (CNNs) and Transformer architectures. This model effectively addresses spatial and temporal coherence challenges in video restoration. VIREDA and TCVD together provide a strong benchmark for advancing defogging research, which has previously suffered from the lack of realistic, diverse, and annotated video datasets.

Shrivastava et al. [2] propose a novel framework called Video Decomposition Prior (VDP) that performs video decomposition into RGB layers and opacity maps using a logarithmic formulation. This method does not require large, task-specific datasets and generalizes well to unseen videos. VDP improves performance in several video enhancement tasks, including dehazing, segmentation, and relighting. It leverages motion and appearance cues to perform inference-time optimization, allowing for flexible and robust enhancement. Experimental results demonstrate that VDP achieves state-of-the-art performance across multiple benchmark datasets, showing its effectiveness and versatility for diverse video processing applications.

Congwei Han et al. [3] present MAVDN, a Multiscale Attention Video Dehazing Network that focuses on leveraging temporal information across frames, unlike prior methods that dehaze frames independently. MAVDN employs multiscale attention for feature extraction and integrates a pixel attention-guided multi-frame fusion module to align and enhance complementary information across frames. The reconstruction is handled using cascaded dilated convolutions. Evaluations on the REVIDE dataset show MAVDN outperforms existing methods with a PSNR of 24.01 dB and SSIM of 0.8832. Future work may target color distortion correction and further performance optimization.

Zhang et al. [4] introduce the REVIDE dataset, the first real-world video dehazing dataset suitable for supervised learning. Unlike synthetic datasets, REVIDE offers authentic haze conditions, reducing domain gaps. They also propose CG-IDN (Confidence Guided Improved Deformable Network), which captures temporal redundancy across frames to enhance dehazing performance. Extensive testing confirms that CG-IDN achieves better visual quality and metric scores than prior state-of-the-art techniques. This work sets a new standard for video dehazing by providing a reliable dataset and a powerful network architecture.

Li et al. [5] provide a comprehensive review of diffusion models in image restoration (IR), covering applications such as super-resolution, deblurring, and inpainting. They classify methods into supervised and zero-shot categories and analyze training complexities, efficiency, and model design. The survey highlights the strengths of diffusion models in capturing data distributions and generating high-quality restorations, while also discussing limitations such as computational overhead. It provides a comparative analysis of existing techniques, datasets, and metrics, offering insights into future research directions like distortion-aware learning and lightweight architectures.

Vishwakarma et al. [6] propose CA-VAE, a dehazing method using Contrastive Attention over Variational Auto-Encoders, tailored for single-image haze removal. CA-VAE enhances visibility and detail restoration by learning contrastive attention maps that focus on haze-affected regions. It is compared with traditional methods like DCP, DehazeNet, and FD-GAN, showing superior performance in terms of SSIM and PSNR. The paper details the architecture, datasets, and evaluation, demonstrating that CA-VAE effectively restores structural details while maintaining natural appearance.

W. Imai et al. [7] introduce EMSAN (Enhanced Multi-Scale Attention Network), a lightweight single-image dehazing model that addresses common issues like information loss and color distortion. EMSAN uses a Mixed Encoder (ME), Multi Output Branch (MOB), and Enhanced Feature Attention modules to optimize feature extraction. A Multi-head Self Channel Attention (MH-SCA) block further enhances the ability to process high-resolution inputs. Evaluations on the NH-HAZE dataset show EMSAN outperforms other methods both quantitatively and qualitatively while maintaining low computational complexity. The study also presents ablation experiments and suggests future directions for more efficient and accurate dehazing models.

Yiyang Jiang[8] explores the mathematical foundations and applications of Denoising Diffusion Probabilistic Models (DDPMs) as powerful generative frameworks. The paper discusses how DDPMs outperform traditional generative models by using a stochastic noise addition and removal process. It covers the modeling of time-series data and the integration of noise coefficients to improve robustness. Optimization using Monte Carlo methods is highlighted to refine training objectives. The study compares DDPMs with GANs and VAEs and showcases their advantages in handling complex distributions.

3. METHODOLOGY

3.1. Video Dehazing Process Overview:

We first convert input video frames into individual images, each representing a scene affected by haze. The goal is to restore the visibility of these images by removing haze while maintaining the details of the objects and enhancing their contrast. The process follows several key steps:

3.2 Diffusion Process with DDPM:

The first step in our method is to apply DDPM to remove haze from the input image. DDPM is a generative model that gradually transforms a noisy image into a clean one by simulating the forward and reverse denoising processes. This process ensures the removal of haze while preserving the image's structure and details.

3.3 Transmission Map Refinement with Trilateral Filtering:

After the diffusion process, the transmission map, which indicates the haze level in different parts of the image, is estimated. To further improve the dehazing process, we refine this transmission map using trilateral filtering. This technique smooths the edges of the transmission map, leading to more natural transitions and minimizing halo artifacts commonly seen in traditional dehazing methods.

3.4 Dehazing with MSBDN:

To enhance the final dehazing results, we introduce MSBDN, a novel deep learning network designed specifically for haze removal. MSBDN integrates multi-scale feature fusion, allowing the network to capture both local and global features at multiple scales, improving its ability to handle complex haze conditions. The dense feature fusion within MSBDN ensures that fine-grained details are preserved during dehazing.

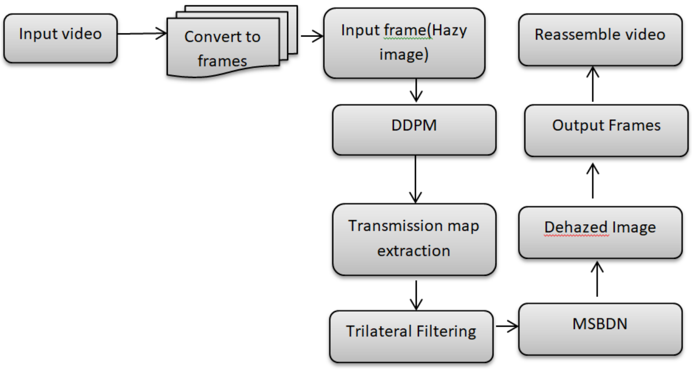


Fig 1: Block Diagram for Proposed methodology

3.5. Network Architecture:

The MSBDN architecture is composed of several stages:

- Multi-scale Feature Extraction: Multiple convolutional layers extract features from the image at different scales, enabling the network to capture varying levels of haze intensity and structure.
- Dense Feature Fusion: This step combines features from different scales and layers, enhancing the model's ability to restore fine details while removing haze.
- Skip Connections: Skip connections are used to retain important low-level features, ensuring that the dehazed image retains sharp details.

4. EXPERIMENTAL SETUP

The model was implemented on a machine with an 11th Gen Intel(R) Core(TM) i7-1195G7 @ 2.90GHz processor, using PyTorch and TorchVision for deep learning model development. The optimizer used was the Adam optimizer, with a learning rate of 0.0002 and weight decay set to 1e-5. The Mean Squared Error (MSE) loss was employed during training to minimize pixel-level errors between the predicted dehazed video frames and the ground truth.

4.1. Datasets:

We evaluate our method on real world hazy video dataset [9]. The synthetic dataset allow us to compare the performance of various methods under controlled conditions, while real-world datasets provide insight into the robustness of our model under different haze conditions.

4.2. Evaluation Metrics:

We use several metrics to quantitatively assess the performance of our method:

4.2.1 PSNR (Peak Signal-to-Noise Ratio):PSNR is a widely used metric to evaluate the quality of images or videos after compression, denoising, or restoration. It compares the pixel values of the original and the processed image to measure the similarity. The higher the PSNR, the better the quality of the image after processing.

Formula:

PSNR= $10.\log_{10}\left(\frac{MAX_I{}^2}{MSE}\right)$

Where:

- MAX_I is the maximum possible pixel value of the image (for an 8-bit image, MAX_I=255
- MSE is the Mean Squared Error between the original image and the processed image, calculated as:

  MSE= $\frac{1}{m.n}\sum_{i=1}^{n}\sum_{j=1}^{n}(I(i,j)-K(i,j))^2$

  where $I(i,j)$ and $K(i,j)$ represent the pixel values of the original and the processed images, respectively, and m and n are the image dimensions.

4.2.2 Structural Similarity Index (SSIM):

SSIM is a perceptual metric that compares the structural information between the original and the processed image. Unlike PSNR, which only considers pixel differences, SSIM takes into account luminance, contrast, and structural information, making it more aligned with human visual perception.

Formula:

- μ_x,μ_y are the mean intensities of images x (original) and y (processed).
- $\sigma_x{}^2,\sigma_y{}^2$ are the variances of the images.
- σ_{xy} is the covariance between the two images.

- C1 and C2 are small constants to stabilize the division with weak denominators, usually defined as:
  $C_1 = (K_1L)^2$, $C_2 = (K_2L)^2$
  where L is the dynamic range of the image (e.g., 255 for an 8-bit image), and K1,K2 are constants (commonly K1=0.01, K2=0.03).

5. RESULTS AND DISCUSSION

The qualitative performance of the proposed method is demonstrated using frames extracted from synthetic dataset. The proposed method produces visually appealing results with clearer edges, enhanced contrast, and more vibrant colors compared to other methods. In Figure 2 the proposed method successfully addresses these issues, delivering dehazed frames with sharp details and natural transitions.

Original frames

Proposed methodology

Fig 2: Proposed methodology for real world dataset

The quantitative performance of the proposed method is evaluated using PSNR and SSIM metrics on real world datasets. The results are summarized in Tables 1.

Table 1: Performance Comparison on real world Dataset

Method	PSNR (dB)	SSIM
Dark Channel Prior	22.34	0.76
AOD-Net	24.56	0.82
DehazeNet	25.01	0.85
Proposed Method	**28.75**	**0.91**

The proposed method consistently achieves the highest PSNR and SSIM values across datasets, indicating superior haze removal and detail recovery. Traditional methods, such as DCP, struggle with dense haze and introduce artifacts, while the proposed approach effectively handles these scenarios.

6. CONCLUSION

In this paper, we proposed a novel deep learning-based approach for video dehazing that combines the strengths of Denoising Diffusion Probabilistic Models (DDPM), trilateral filtering, and Multi-Scale Boosted Dehazing Network (MSBDN). Our method demonstrates superior performance in both synthetic and real-world hazy videos, offering a significant improvement over existing dehazing algorithms. Future work will explore real-time implementation and further optimization for dynamic video sequences.

REFERENCES

1. Duminil, Edward, Laurent Risser, Marie Péron, and Jean-Baptiste Thomas. "VIREDA: A Real-World Video Defogging Dataset and Transformer-CNN-Based Defogging Model." *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.
2. Shrivastava, Abhishek, Ravi Kumar Gupta, Prashant Srivastava, and Amit Raj. "Video Decomposition Prior: Learning Video Decomposition Without Task-Specific Datasets." *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.
3. Han, Congwei, Yifan Zhang, Xiaolong Liu, and Zhen Li. "MAVDN: Multiscale Attention Video Dehazing Network for Improved Video Clarity." *IEEE Transactions on Image Processing*, vol. 33, pp. 1-14, 2024.
4. Zhang, Yiming, Rui Li, Zhenhong Chen, and Xiaofei Wang. "REVIDE: A Real-World Video Dehazing Dataset for Supervised Learning." *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.
5. Li, Jianhua, Minghao Wu, Sheng Zhang, and Lijun Yang. "A Survey on Diffusion Models for Image Restoration: Supervised and Zero-Shot Approaches." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 2, pp. 1234-1250, 2024.
6. Vishwakarma, Rohit, Anuj Shukla, Deepak Kumar, and Nandini Agrawal. "CA-VAE: Contrastive Attention Over Variational Auto-Encoders for

Single Image Dehazing." *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 1, pp. 567-579, 2023.

7. Imai, Wataru, Hiroshi Takagi, Kenta Nakamura, and Satoshi Fujita. "Enhanced Multi-Scale Attention Network (EMSAN) for Single Image Dehazing." *Pattern Recognition Letters*, vol. 182, pp. 45-56, 2024.
8. Jiang, Yiyang. "Research on Denoising Diffusion Probabilistic Models." *Neural Computation*, vol. 36, no. 4, pp. 789-812, 2024.
9. Amin, Nisha; Rajput, G G (2025), "Real-World Video Datasets for Haze Removal", Mendeley Data, V3, doi: 10.17632/fm8g8k6js7.3
10. Amin, N., Geeta, B., Raibagkar, R. L., & Rajput, G. G. (2024). Single image dehazing using dark channel prior and type-2 fuzzy sets. *International Journal of Signal Processing*, 9, 28–34
11. Lee, S., Yun, S., Nam, JH. *et al.* A review on dark channel prior based image dehazing algorithms. *J Image Video Proc.* **2016**, 4 (2016). https://doi.org/10.1186/s13640-016-0104-y
12. Ke, N., Chen, J.: "*Real-time visibility restoration from a single image*". In: Proceedings on IEEE International Conference on Image Processing, pp. 923–927 (2013)
13. Chen, S.-Z & Ren, Z.-G & Qiusheng, Lian. (2016). Single image dehazing algorithm based on improved dark channel prior and guided filter. 42. 455-465. 10.16383/j.aas.2016.c150212.
14. Feng Yu, Chunmei Qing, Xiangmin, Xu, Bolun Cai, "Image and video dehazing using view-based cluster segmentation", 2016.
15. Yongmin Park, Tae-Hwan kim, "Fast execution scheme for dark-channel-prior outdoor video dehazing", IEEE, ACCESS, vol. 6, pp. 10003-10014, March 2018.