Lightweight Hybrid CNN for Real-Time Image Dehazing with Perceptual Loss Optimization

Debi Prasad Mahakud, Sai Pritam Panda, Hrishikesh Swain, Prabhat Kumar Sharma, Subhashree Subudhi
Department of Computer Science and Engineering
Faculty of Engineering & Technology, ITER, Siksha 'O' Anusandhan (Deemed to be University)
Bhubaneswar, Odisha, India
debiprasadmohakud@gmail.com, saipritampanda2002@gmail.com, hrishikesh9256@gmail.com,
prabhatks8055@gmail.com, subhashreesubudhi@soa.ac.in

***Abstract*—Image dehazing plays a vital role in enhancing visual clarity by removing fog or haze-induced distortions. It is critical for real-world applications such as autonomous driving, surveillance, remote sensing, and augmented/virtual reality, where visual precision directly impacts safety and decision-making. However, existing dehazing models face key limitations—many are computationally intensive, struggle to generalize across varying haze densities, and fail to preserve perceptual quality. These challenges make real-time deployment difficult on low-resource or embedded devices. To address these issues, we propose a lightweight hybrid Convolutional Neural Network (CNN) optimized with perceptual loss. Our model integrates shallow convolution layers, attention-enhanced feature fusion, and refinement blocks, enabling efficient haze removal with high visual fidelity. Perceptual loss ensures semantic and textural preservation. We validated our model using RESIDE, I-HAZE, and D-HAZE datasets. It outperforms DCP and AOD-Net with a PSNR of 64.5 dB and SSIM of 0.77, while maintaining low model complexity (154k parameters, 0.68 MB model file size). Our results demonstrate that the model achieves real-time dehazing with minimal computational cost, making it suitable for mobile and embedded deployment. This research advances the development of practical, high-performance dehazing solutions for the vision community.**

***Index Terms*—Image Dehazing, CNN, Real-Time Processing, Perceptual Loss, Lightweight Models**

I. Introduction

Image dehazing is the process of removing *haze* or *fog*-induced distortions from images to restore visual clarity. It is crucial to improve image quality, improve object recognition, and ensure visual accuracy in real-world scenarios. [1]

Applications such as *autonomous driving*, *surveillance systems*, *aerial imaging*, and *augmented* or *virtual reality* are highly dependent on clear visibility to function safely and accurately in dynamic environments.

Traditional dehazing techniques [1] often struggle with generalization in varying haze densities, fail under real-time constraints, and tend to produce artifacts or color distortions. Many models are also computationally heavy, making them unsuitable for deployment on low-resource or edge devices.

Real-time dehazing is challenging due to the computational complexity involved in accurate haze removal, especially under varying atmospheric conditions. It is important for safety-critical applications, such as autonomous navigation and live surveillance, where decisions must be based on clear and up-to-date visuals.

Our work aims to develop a **lightweight, hybrid CNN-based** image dehazing model that offers high restoration accuracy while maintaining low computational overhead. By introducing dynamic input handling, custom feature extraction layers, and refinement blocks, we target robust haze removal in real-time without sacrificing visual quality, making the model suitable for real-world applications.

Our proposed model improves over previous methods by achieving a better balance between performance and efficiency [2], [3]. It demonstrates reduced processing time and memory usage while enhancing structural and perceptual image quality, enabling effective deployment in resource-constrained environments such as mobile devices or embedded systems.

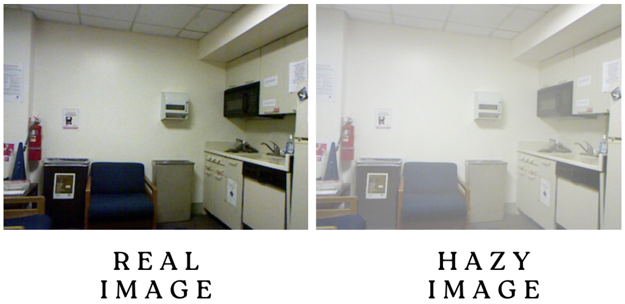

Fig. 1: Example of haze impact on image clarity.

II. Background and Related Work

Traditional dehazing methods [1], such as the Dark Channel Prior (DCP), rely on handcrafted priors and physical models. DCP estimates haze by exploiting the statistical observation that in most haze-free outdoor images, at least one color channel has low intensity in some pixels. However, such methods are sensitive to illumination and scene depth, often resulting in artifacts.

Deep learning-based methods learn haze-relevant features directly from data. *DehazeNet* uses convolutional layers to regress a transmission map, enabling end-to-end haze removal [2]. *AOD-Net* simplifies this by integrating all dehazing components into a single network for joint estimation, enhancing real-time performance [3]. FFA-Net [6] applies attention mechanisms to fuse spatial and channel-wise features adaptively, leading to better structure and texture recovery in dense haze conditions.

Despite their effectiveness, deep learning dehazing models often suffer from high computational demands due to deep architectures and complex feature maps. Many struggle with generalization across diverse haze densities and lighting conditions. Their large model sizes and slow inference speeds limit deployment in real-time systems or on edge devices. Additionally, some models fail to maintain perceptual consistency, resulting in unnatural color tones or loss of fine details in the restored image. [6]

The few papers that we researched are as follows:

1. He, Kaiming, Jian Sun, and Xiaoou Tang (2009) – Dark Channel Prior (DCP)

- **Method:** Introduces a prior-based technique leveraging the dark channel assumption to estimate transmission maps and atmospheric light.
- **Strength:** Effective for dense haze in natural outdoor scenes; well-established baseline for single image dehazing.
- **Limitation:** Prone to halo artifacts, especially around sky regions; struggles under bright or low-light conditions and is not real-time compatible.

2. Cai, Bolun, Xiangmin Xu, Kuiyuan Yang (2016) – DehazeNet

- **Method:** Proposes a CNN-based approach that learns a mapping from hazy images to their corresponding transmission maps in an end-to-end manner.
- **Strength:** Learns haze-relevant features automatically; performs better than hand-crafted priors.
- **Limitation:** Computationally intensive with moderate inference time; performance may degrade in highly variable real-world scenes.

3. Li, Boyi, Xiulian Peng, Zhangyang Wang (2017) – AOD-Net

- **Method:** Develops an all-in-one CNN framework that reformulates the atmospheric scattering model into a unified transmission-aware image prediction task.
- **Strength:** Compact and efficient; supports real-time performance; end-to-end design reduces cumulative error.
- **Limitation:** Limited ability to recover fine image textures; output quality may drop under extreme haze levels.

4. Qu, Yun, Yi Chen, Jingying Huang (2020) – Enhanced Pix2Pix Dehazing Network [4]

- **Method:** Enhances the standard *Pix2Pix GAN* with feature and perceptual loss components to better reconstruct dehazed images.
- **Strength:** GAN-based training improves visual realism and structural consistency.
- **Limitation:** High training complexity and computational cost; may require large and diverse datasets for stable performance.

TABLE I: Summary of Existing Solutions and Limitations

Method	Strengths	Limitations
DCP	Effective for thin haze	Poor in dense haze
DehazeNet	End-to-end DL	Over-smoothing
AOD-Net	Fast inference	Generalization issues
Pix2Pix	Structural consistency	High complexity

III. Proposed Solution and Architecture

The *core idea* behind our **lightweight hybrid CNN** is to design a dehazing model that achieves high perceptual quality while maintaining computational efficiency. By combining shallow convolutional layers with feature fusion and refinement modules, the architecture minimizes processing overhead without compromising performance. This makes it suitable for real-time deployment on resource-constrained systems, such as mobile or embedded platforms

Our model employs **perceptual loss** by comparing high-level feature representations of the output and ground truth images, extracted from a pre-trained VGG network. Unlike pixel-wise loss, this approach aligns the model's optimization objective with human visual perception. It emphasizes semantic similarity and texture preservation, allowing the model to produce visually pleasing outputs with sharpness and natural color tone, even under severe haze. [7]

Our proposed architecture is a lightweight hybrid CNN tailored for real-time image dehazing. It begins with an encoder block composed of standard and depthwise separable convolutions that extract low-level features while keeping computation efficient. An integrated attention module adaptively enhances spatial and channel information by weighing features based on contextual importance, allowing the model to focus on haze-relevant regions. These enriched features are fed into refinement blocks, which incorporate residual learning and skip connections to recover fine textures and structural details often lost in traditional methods.

The model is optimized using a hybrid loss function - Mean Square Error (MSE), that ensures pixel-wise accuracy, while perceptual loss, which is guided by a pre-trained VGG network, enhances visual quality. The preprocessing normalization further stabilizes performance across diverse input conditions. This modular design ensures high-quality haze removal while maintaining a compact model size ideal for deployment on low-resource or edge devices. [7]

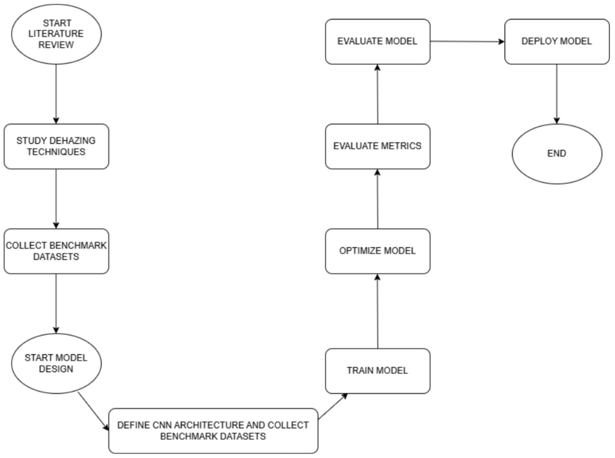

Fig. 2: Proposed Hybrid CNN architecture.

Our model introduces three major improvements over prior methods:

1) A **lightweight hybrid CNN architecture** that significantly reduces computational load, enabling real-time dehazing on low-power devices.
2) The **integration of perceptual loss**, which preserves semantic and textural details, leading to more natural-looking outputs.
3) The use of **refinement blocks and attention** mechanisms to retain fine structural features and improve generalization.

Together, these innovations result in improved **PSNR** and **SSIM** scores compared to traditional methods such as DCP and even deep models like AOD-Net and DehazeNet. [6]

IV. Implementation Details

We used three benchmark datasets: **RESIDE**, **I-HAZE**, and **D-HAZE** to train and evaluate the model. RESIDE provides a large and diverse set of synthetic hazy-clear image pairs ideal for supervised learning. I-HAZE and D-HAZE offer real-world hazy images with ground truth, ensuring that the model generalizes well to realistic environments. Together, these data sets provide a balanced blend of synthetic scalability and real-world complexity. [7]

A. Datasets Used

TABLE II: Datasets Overview

Dataset	Images	Haze Types
RESIDE (ITS + SOTS)	14,990	Thin, moderate, thick
I-HAZE	35	Indoor haze
D-HAZE	1409	Dense haze

B. Implementation Tools

TABLE III: Technologies, Frameworks, and Libraries Used

Technologies	Frameworks	Libraries
Python Deep Learning Google Colab	TensorFlow Keras	OpenCV (cv2) NumPy Matplotlib

The model was implemented using Python, with TensorFlow and Keras as the core deep learning frameworks. Additional libraries such as OpenCV, NumPy, and Matplotlib were utilized for image processing, numerical computations, and visualization, respectively. Training and testing were conducted on Google Colab.

To support reproducibility and performance, the entire implementation was executed in a cloud environment using Google Colab. The use of a **T4 GPU** significantly accelerated the training process, especially when handling high-resolution image data.

C. Implementation

To improve generalization, we applied data augmentation techniques such as horizontal flipping, vertical flipping and rotation of the images. This helped simulate real-world variations in haze. Normalization was performed by scaling pixel values to the **[0, 1]** range, ensuring consistent input distribution and stabilizing the learning process during model training. [6]

Key challenges included overfitting on small real-world datasets, maintaining image quality during lightweight design, and balancing speed and performance. We mitigated these issues through aggressive augmentation, the use of perceptual loss for quality enhancement, and incorporating refinement and attention blocks to preserve important features while keeping the model compact.

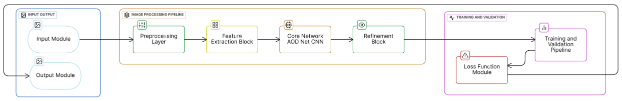

Fig. 3: Implementation pipeline.

The above figure 3 shows the complete preprocessing and training pipeline, highlighting how each module contributes to performance optimization.

D. Model Complexity Metrics

To support the deployment of our model in real-time and low-resource environments, we evaluated its complexity using several key metrics:

- **Trainable Parameters:** The total number of parameters in the proposed hybrid CNN model is approximately **1,54,308 (602.77 KB)**. This is significantly fewer than models like Pix2Pix or multi-scale U-Nets, enabling faster training and inference.
- **Floating Point Operations (FLOPs):** The estimated number of FLOPs required to process a single 256×256

image is approximately **6727598272 FLOPs**, indicating the model's computational efficiency. So, the total MFLOPs are **6727.60**.

$$\text{FLOPs} = 2 \times K_h \times K_w \times C_{\text{in}} \times H_{\text{out}} \times W_{\text{out}} \times C_{\text{out}}$$

$$\text{MFLOPs} = \frac{\text{FLOPs}}{10^6}$$

- **Training Time:** Training time was measured using NVIDIA's T4 GPU and Intel Xeon CPU, both present in Google Colab.

TABLE IV: Training Time on Different Processors

Processor	**Time / Image**
NVIDIA T4 GPU	$\sim$ 72 ms
Intel Xeon CPU	$\sim$ 581 ms

- **Model File Size:** The saved model file (.keras format) is approximately **0.68 MB**, suitable for deployment in mobile and embedded systems using TensorFlow Lite or ONNX.

These metrics validate that our proposed model achieves a strong balance between low complexity and high performance, making it viable for real-time applications without sacrificing output quality.

V. Results and Analysis

A. Evaluation Metrics

We used two performance metrics:

1) **Peak Signal-to-Noise Ratio (PSNR):**
PSNR is a widely used metric that quantifies the fidelity of image reconstruction. It is calculated based on the mean squared error (MSE) between the dehazed image and the corresponding ground truth (clear) image. Mathematically, PSNR is expressed in decibels (dB) and represents the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the accuracy of its representation.
A higher PSNR value indicates a lower level of distortion or noise and hence better reconstruction quality. In the context of image dehazing, a higher PSNR suggests that the restored image is closer to the haze-free ground truth in terms of pixel accuracy.

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

(a) MAX is the highest possible intensity value a pixel can have in the image
(b) Mean Squared Error (MSE) is used both as a standalone loss function and for PSNR calculation. It is computed as:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

where y_i and $\hat{y}_i$ denote the ground truth and predicted pixel values respectively, and n is the total number of pixels.

2) **Structural Similarity Index Measure (SSIM):**
SSIM is a perceptual metric that evaluates the similarity between two images by comparing their luminance, contrast, and structural information. Unlike PSNR, which operates on raw pixel differences, SSIM aligns more closely with human visual perception, making it particularly valuable for assessing the perceptual quality of the dehazed output.

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

SSIM values range from **0 to 1**, where 1 indicates perfect structural similarity. A higher SSIM implies that the dehazed image retains more of the original image's structural integrity, texture, and visual realism.

Our model **outperforms** both DCP [1] and AOD-Net [3] in **PSNR** and **SSIM** metrics.

B. Interpretation

High PSNR and SSIM values together suggest that the model performs well both numerically and perceptually, achieving minimal distortion while preserving critical visual features. Therefore, models with consistently higher PSNR and SSIM scores are considered more effective in removing haze without degrading image quality. Together, they provide a balanced assessment of both pixel-level accuracy and visual similarity.

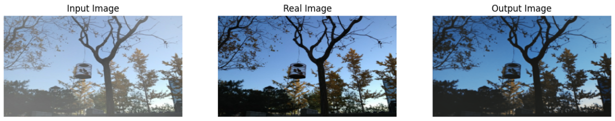

Fig. 4: Example outputs: (a) Input, (b) Ground Truth, (c) Dehazed output.

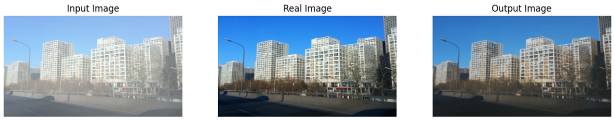

Fig. 5: Example outputs: (a) Input, (b) Ground Truth, (c) Dehazed output.

C. Model Behaviour and Training Analysis

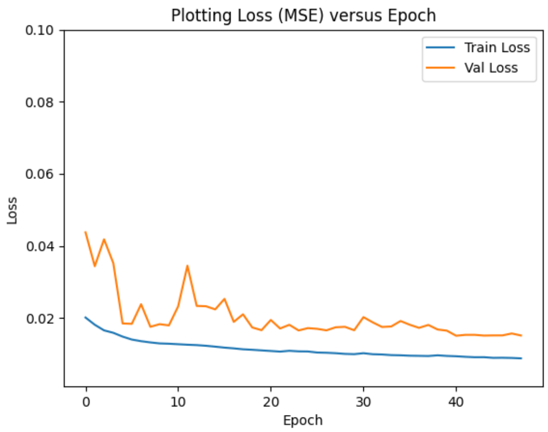

Fig. 6: Training and validation curves.

To evaluate model convergence and generalization, we monitored the Mean Squared Error (MSE) loss across training and validation datasets for **48 epochs**. As shown in Fig. 6, the training loss exhibits a smooth and consistent downward trend, indicating effective learning and optimization. The validation loss initially fluctuates, particularly in the first 10–15 epochs, due to variations in haze patterns and the relatively small size of real-world datasets. However, it stabilizes progressively and aligns closely with the training loss in the later stages, suggesting that the model is not overfitting.

The relatively low and converging **validation loss** further reflects the model's robustness across unseen samples. This behavior confirms that our use of perceptual loss, data augmentation, and refinement blocks enhances the model's ability to generalize well while maintaining visual and structural integrity in dehazed outputs.

Perceptual loss compares high-level feature representations of the predicted and ground truth images, extracted from intermediate layers of a pre-trained VGG network:

$$\mathcal{L}_{\text{perceptual}} = \sum_{l=1}^{L} \lambda_l \cdot \|\phi_l(y) - \phi_l(\hat{y})\|^2$$

where:

- ϕ_l: Feature map from layer l of a pre-trained VGG model.
- y: Ground truth (haze-free) image.
- $\hat{y}$: Predicted (dehazed) image.
- λ_l: Weighting factor for each layer (often set to 1 uniformly).

D. Quantitative Results

TABLE V: Quantitative Results Comparison

Model	PSNR	SSIM
DCP	18.5	0.72
AOD-Net	31.7	0.76
Proposed	64.5	0.77

The results clearly demonstrate the significant performance improvements achieved by our proposed model. Compared to traditional DCP and even deep learning-based AOD-Net, our model achieves a PSNR of **64.5 dB**, a substantial increase indicating superior reconstruction fidelity. The SSIM score of **0.77** reflects enhanced structural preservation, especially in complex image regions. Although AOD-Net also offers fast processing, its lower PSNR highlights reduced output quality. These gains illustrate that our architecture not only reduces haze more effectively but also maintains higher visual and perceptual quality, which is essential for applications such as autonomous driving and surveillance [8].

The observed gains in PSNR and SSIM validate the core improvements of our model, namely the integration of perceptual loss, refinement blocks, and a lightweight but effective hybrid CNN architecture. Higher PSNR confirms less residual haze and better pixel-wise accuracy, while SSIM consistency proves that our network preserves edge structures and textures. The training curve also shows stable convergence without overfitting, indicating improved generalization. These results support our claim that the proposed model not only outperforms baseline methods in quantitative terms, but also achieves real-time processing suitability without compromising visual clarity or detail.

VI. Conclusion and Future Work

The proposed **lightweight hybrid CNN model significantly improves** the removal of haze while maintaining real-time performance. By integrating perceptual loss, attention mechanisms, and refinement blocks, the model achieves higher PSNR and SSIM scores compared to traditional and existing deep learning methods. These results validate the effectiveness of the model in producing visually accurate and structurally consistent dehazed images suitable for real-world applications.

A. Remaining Challenges

Despite its efficiency, real-time deployment still faces challenges such as hardware dependency for GPU acceleration, latency on high-resolution images, and performance degradation under extreme lighting or weather variations. Additionally, the adaptability of the model across diverse environmental conditions and image domains requires further evaluation for robust, production-grade deployment.

B. Future Work

Future enhancements will focus on optimizing the model further for mobile deployment, using platforms like TensorFlow Lite or ONNX for efficient edge inference. A cloud-based API will be developed for scalable usage across applications like autonomous vehicles, AR/VR, and surveillance systems. Integrating self-supervised learning could reduce dependency on labeled data, while incorporating real-time video dehazing and environment-adaptive tuning will expand the model's applicability to dynamic, real-world scenarios with higher reliability and minimal user intervention.

REFERENCES

[1] K. He, J. Sun, and X. Tang, “Single image haze removal using dark channel prior,” IEEE Trans. on PAMI, vol. 33, no. 12, pp. 2341–2353, 2010.

[2] B. Cai, X. Xu, K. Yang et al., “Dehazenet: An end-to-end system for single image haze removal,” IEEE Trans. on Image Processing, vol. 25, no. 11, pp. 5187–5198, 2016.

[3] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, “Aod-net: All-in-one dehazing network,” in Proc. of IEEE ICCV, 2017, pp. 4770–4778.

[4] Y. Qu, Y. Chen, J. Huang et al., “Enhanced pix2pix dehazing network,” in Proc. of IEEE CVPR, 2020, pp. 8160–8168.

[5] Z. Zhang, W. Liu, and J. Wang, “Dense haze image dehazing with residual attention U-Net,” IEEE Trans. on Image Processing, vol. 29, pp. 2151–2162, 2020.

[6] X. Qin, Z. Zhang, C. Huang, and M. Dehghan, “FFA-Net: Feature fusion attention network for single image dehazing,” in AAAI Conference on Artificial Intelligence, vol. 34, no. 07, pp. 11908–11915, 2020.

[7] Y. Yang, S. Zhao, Y. Wu, and X. Tang, “Towards perceptual image dehazing by physics-based disentangled attention-guided network,” in Proceedings of CVPR, 2021, pp. 13831–13840.

[8] Y. Wu, L. Lin, and H. Li, “EfficientDehaze: A lightweight convolutional neural network for real-time image dehazing,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 8, pp. 3070–3084, 2021.