ISSN (ONLINE): 2456-5717
ISSN (PRINT): 2456-5717

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol. 12, Issue 10, October 2026

Utilizing a Multi-Component Neural Network Architecture for Single Image Dehazing Based on Deep Learning

Kanchan Jha[1], Poonam Sharma[2]
Department of Computer Science and Engineering, Ramdeobaba University Nagpur, India [1]
kanchanvnitphd@gmail.com
Department of Computer Science and Engineering, VNIT, Nagpur, Nagpur, India [2]
poonamsharma@cse.vnit.ac.in

***Abstract*— Image dehazing is an important preprocessing step used in almost every computer vision application. The quality of an image deteriorates due to the presence of haze. Good quality images are needed mainly in areas such as object detection, autonomous driving, and remote sensing. Due to the presence of haze in an image, these areas are affected. In this paper, a novel deep learning approach based on fusion is presented. The hazy image is divided among three specialised networks: Tθ for the transmission map, Kθ for the atmospheric light, and Fθ for fusion. The first branch Tθ uses UNet and DCP architectures to estimate the amount of haze in an image. The second branch Kθ is used to estimate the atmospheric light. The final fusion Fθ network merges the input hazy image with the outputs of Tθ branch and Kθ branch to produce the final clear dehazed image. The experiment is carried out on the RESIDE outdoor dataset. The quantitative PSNR value obtained is 27.76 dB and the SSIM value is 0.9623. These values are competitive with existing methods under the adopted evaluation protocol. In the best case, the PSNR value achieved is 36.43 dB.**

***Index Terms*—CNN, Deep Learning, Atmospheric light, PSNR, SSIM, Image Dehazing, Transmission Map, Attention Mechanism.**

I. Introduction

The images taken in foggy morning weather are usually blurry and washed out. It is not a trivial problem. This type of image is a serious issue in the case of autonomous cars [1], surveillance cameras [2] and drones [3]. When there is a presence of haze in an image, the objects cannot be detected properly [4], [5], colours get faded, and real objects get hidden behind the haze. Physically, it can be understood as:- the haze occurs due to Atmospheric particles which scatter the light before reaching the camera. This phenomenon can be described by the atmospheric scattering model [6, 7].

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

Here, I(x) is the hazy image captured by the camera, J(x) is the clear haze-free image which has to be recovered, and t(x) is the transmission map, which tells us the amount of light that actually reaches the camera. A is the atmospheric light. Basically, the goal is to obtain the true image J(x) by reversing equation (1). Traditional methods like the Dark Channel Prior (DCP) [8] have worked well, but they have a few limitations, such as failure in sky regions, slow processing, etc. After the arrival of the deep learning approach, end-to-end learning became possible, and the results improved significantly [9]. This approach combines the best of both: physics-based understanding plus deep learning power.

Recent progress in deep learning has shown astonishing improvement in low-level vision work, which includes image super-resolution [10], image denoising [11], and image restoration [12]. With this success, many deep learning-based dehazing methods have been proposed. Most existing methods either estimate intermediate physical parameters or directly predict the clear image in an end-to-end manner. The proposed method bridges this gap by combining physically guided estimation with learned fusion.

First, the methods previously used in dehazing will be described, and then their performance will be compared with the proposed method.

II. Related Work

A. Traditional Methods

The Dark Channel Prior (DCP) [8] was the first work to observe that haze-free images have some pixels that are very dark. Based on this observation, the transmission is estimated. The PSNR value on the RESIDE dataset is about 16 to 19 dB. This low PSNR value is due to the presence of sky regions, where the Dark Channel Prior is violated. Due to the presence of soft matting, the processing speed is also very slow.

The Colour Attenuation Prior (CAP) [13] observed that, due to the presence of haze, the brightness of an image increases while the saturation decreases. The depth of haze is estimated using a simple linear model. This method was fast, but the accuracy was very low. A PSNR value of around 19 to 21 dB is achieved. It works well with simple scenes but struggles with complex scenes. In Non-local Dehazing [14], it was assumed that in haze-free images the colours form tight clusters. It was a clever approach, but computationally it was expensive. In the case of dense haze, the accuracy drops. The PSNR value of the

ISSN (ONLINE): 2456-5717
ISSN (PRINT): 2456-5717

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol. 12, Issue 10, October 2026

output image is around 17 to 20 dB. Fattal [15] proposed a method based on the assumption of local statistical independence between surface shading and transmission. Tarel and Hautière [16] proposed a fast visibility restoration approach using median filtering for transmission estimation.

B. Deep Learning Based Methods

DehazeNet [17] – It was the first CNN-based method that directly estimated the transmission map. The BRELU activation function was used, which bounded the range of the transmission. The PSNR value found on the RESIDE dataset was around 21 to 22 dB. It was better than the traditional methods, but it still had limited network capacity.

MSCNN [18] – The Multi-Scale CNN estimates the transmission in a coarse-to-fine manner. First, it estimates the coarse transmission, and then it estimates the refined transmission. The PSNR value achieved is around 19 to 21 dB. The multi-scale approach was helpful, but end-to-end training was missing, which limited its performance.

AOD-Net [19] – This was the first end-to-end method that directly predicted the clear image from a hazy image without separately estimating the transmission map and the atmospheric light A. The reformulated atmospheric model was $J(x) = K(x)\cdot I(x) - K(x) + b$. It was a lightweight network, with a PSNR value of around 20 dB on the outdoor benchmark. It was fast, but sometimes there was a chance of colour artefacts.

GFN (Gated Fusion Network) [20] – This method generates multiple dehazed results and then fuses them using confidence maps. It was a smart approach, but it had a complex architecture. It had better visual quality, with a PSNR value of around 22 to 24 dB. The main drawback was that it was memory-intensive.

GridDehazeNet [21] – It was an attention-based grid network that combined multi-scale features effectively. It had quite good results, with a PSNR value of around 31 dB on the outdoor benchmark. There was no need for pre-processing and post-processing steps.FFA-Net [22] – It was a feature fusion attention network that used both channel attention and pixel attention. It achieved one of the better state-of-the-art results. The PSNR value was around 33.6 dB on the outdoor benchmark, but it had more than four million parameters, which may cause issues in deployment.AECR-Net [23] – It was a contrastive learning-based approach with an autoencoder. It was a novel idea, but the training process was complex. It achieved around 37 dB on the indoor benchmark.

In addition, many recent methods have made great achievements in the field of image dehazing. MAXIM [24] proposed a multi-axis MLP architecture for image restoration, including dehazing. DehazeFormer [25] introduced a vision transformer-based approach, which can capture long-range dependencies for better haze removal. PMNet [26] introduced a progressive scale network with patch-based training for efficient dehazing. Zheng et al. [27] proposed a curriculum learning approach for dehazing that trains the network from easy to hard samples. RIDCP [28] addressed real-world image dehazing using contrastive prior learning.

III. The Proposed Method

The proposed work is described in detail in this section. The core formula is:

$$J(x) = F\theta\,(I(x), T\theta\,(I(x)), K\theta\,(I(x))) \quad (2)$$

This equation indicates that the input hazy image is processed by three networks: Tθ calculates the transmission map, Kθ computes the atmospheric light, and Fθ combines these to produce the final clear image.

A. Tθ – Transmission Map Estimation Network

This network is the largest component, with nearly 10.5 million parameters. First, the dark channel prior of the input is computed by a DCP module and concatenated with the R, G, and B channels of the hazy image, forming a 4-channel input that provides a physics-based initial cue of the haze density. A U-Net-style encoder of four stages (64, 128, 256, and 512 channels) then extracts features, where each stage uses residual blocks followed by stride-2 downsampling. At the bottleneck (resolution $H/16 \times W/16$), atrous spatial pyramid pooling with dilation rates 1, 2, 4, and 8 captures multi-scale context, and a channel attention module highlights the important channels. The decoder upsamples the features in four stages (512, 256, 128, and 64 channels) using skip connections from the encoder together with residual blocks. Finally, a sigmoid activation followed by guided filter refinement produces a single-channel transmission map tO (x) in the range (0, 1], where values close to 0 indicate dense haze and values close to 1 indicate a clear region.

B. Kθ Atmospheric Light Estimation Network

This network estimates the global atmospheric light together with a local scattering representation. A shared ResNet-18-style backbone, consisting of a 7×7 convolution followed by four residual stages (64, 64, 128, and 256 channels) with batch normalisation and ReLU, extracts features from the hazy image. Two heads are attached to this backbone. The first head estimates the global light: global average pooling is followed by two fully connected layers ($256 \rightarrow 128 \rightarrow 3$) and a sigmoid activation, producing the global atmospheric light $\hat{A}$ as an RGB triplet. As atmospheric light is a global, scene-level property, global pooling is an important step for aggregating the overall context. The second head models the local scattering: a feature pyramid combines the backbone features into 64 channels, the estimated $\hat{A}$ is tiled and injected through concatenation and

ISSN (ONLINE): 2456-5717
ISSN (PRINT): 2456-5717

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol. 12, Issue 10, October 2026

convolution, and the result is upsampled 4× to full resolution to produce a 32-channel scattering feature map $\hat{S}(x)$. In addition, a physics consistency check based on the atmospheric scattering model produces a coarse dehazed estimate, which is passed to the fusion network described in the next subsection.

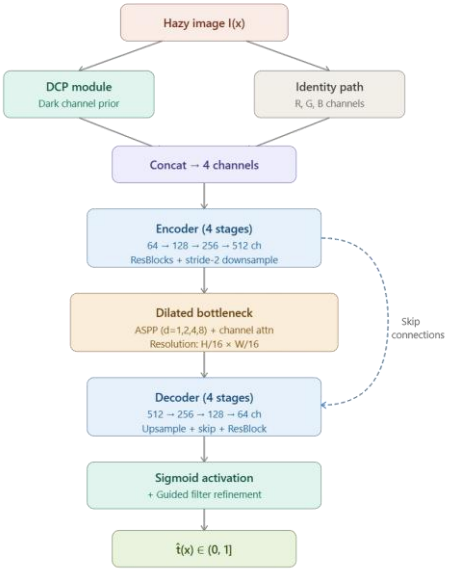


Fig. 1. The Tθ network architecture.

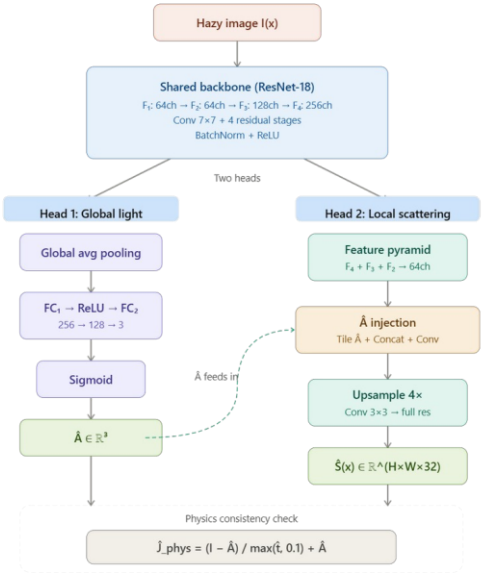


Fig. 2. The Kθ network architecture.

C. *Fθ: Fusion Network*

The fusion network combines the physical estimates with learned image features to produce the final clear image. It has two encoder branches with an identical structure of four levels (48, 96, 192, and 384 channels) built from residual blocks with downsampling. The image branch encodes the hazy image I(x), while the physics branch encodes the concatenation of the transmission map tO (x), the scattering features $\hat{S}(x)$, and a coarse physics estimate $\hat{J}$phys (36 channels in total). The coarse estimate is obtained by inverting the atmospheric scattering model:

$$\hat{J}\text{phys} = (I - \hat{A}) / \max(tO , 0.1) + \hat{A} \qquad (3)$$

The two branches are merged by cross-attention fusion, in which the queries are taken from the physics branch and the keys and values from the image branch, i.e., $F = Q + \gamma \cdot \text{Softmax}(QKT/\sqrt{d}) \cdot V$. A transmission-guided spatial attention (TGSA) module then uses the inverted transmission $W = 1 - tO(x)$ as a spatial prior, refines it with a convolution and a sigmoid activation, and modulates the fused features so that dense-haze regions receive a stronger correction while clear regions are only lightly modified. A four-stage decoder (384, 192, 96, 48, and 24 channels) with skip connections and residual blocks restores the full resolution, and a final convolution predicts a residual R(x). The dehazed image is obtained as $\hat{J}(x) = I(x) +$

ISSN (ONLINE): 2456-5717
ISSN (PRINT): 2456-5717

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol. 12, Issue 10, October 2026

R(x), clamped to [0, 1].

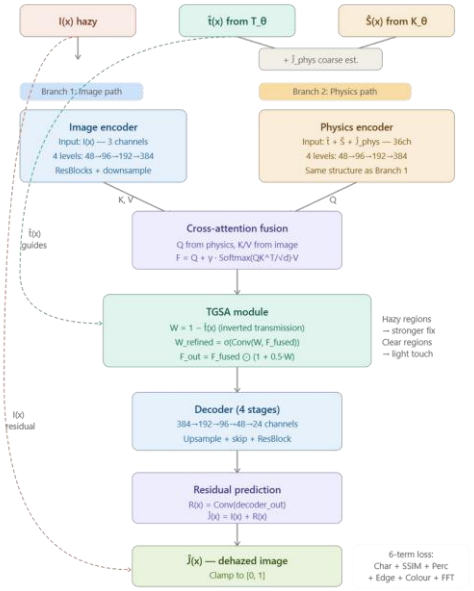


Fig. 3. The Fθ network architecture.

IV. Experiments and Results

A. Dataset

The RESIDE-Outdoor dataset [29] has been used, which is a standard benchmark for dehazing research. There are 500 hazy-clear image pairs with different outdoor scenes. The different haze densities and lighting conditions present in the images make the model robust. We have split the dataset in an 80:20 ratio, where 400 images have been used for training and 100 images for validation. We have resized the images to 256 × 256. It is noted that these 500 pairs constitute the SOTS-outdoor evaluation split of RESIDE; in this work, they were repartitioned into 400 training and 100 held-out images. Consequently, the reported values are not directly comparable to methods trained on the full OTS training set, and the comparison in Table 2 is indicative only. The RESIDE dataset has been widely used as a benchmark in recent dehazing studies.

B. Training Setup

TABLE 1: TRAINING CONFIGURATION

Setting	Value
Epochs	50
Batch Size	8
Optimizer	Adam (1e-4), cosine annealing
Loss	0.84×SSIM + 0.16×L1
GPU	Tesla T4
Image Size	256 × 256
Training Images	400
Validation Images	100

Table 1 depicts the training configuration. A total of 400 images are trained with 50 epochs and a batch size of 8. The remaining 100 images are used for validation purposes. 50 epochs were applied to train the dataset on a Tesla T4 GPU (Google Colab). The Adam optimizer with a learning rate of 1e-4 has been used. The batch size is kept at 8 due to memory constraints. Each epoch took nearly 1.5 minutes to execute.

C. Quantitative Results

All the methods are compared in Table 2. This comparison is based on the RESIDE-Outdoor benchmark (published values under the standard protocol).

D. Comparison Analysis

From the table, it can be observed that the proposed method achieves 27.76 dB PSNR and 0.9623 SSIM under the protocol

ISSN (ONLINE): 2456-5717
ISSN (PRINT): 2456-5717

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol. 12, Issue 10, October 2026

described in Section IV-A. Since the compared values were obtained under the standard OTS training protocol, they are indicative rather than directly comparable. A few observations can be made:

TABLE 2: COMPARISON WITH STATE-OF-THE-ART METHODS

Method	Year	PSNR (dB)	SSIM	Key Feature
DCP [8]	2011	19.13	0.8148	Dark Channel Prior – Traditional
CAP [13]	2015	19.05†	0.8364†	Color Attenuation Prior
Non-Local [14]	2016	17.29†	0.7489†	Haze-line clustering
DehazeNet [17]	2016	22.46	0.8514	First CNN for dehazing
MSCNN [18]	2016	19.84†	0.8327†	Multi-scale CNN
AOD-Net [19]	2017	20.29	0.8765	End-to-end, lightweight
GFN [20]	2018	22.30†	0.8800†	Gated fusion of results
GridDehazeNet [21]	2019	30.86	0.9819	Grid attention network
FFA-Net [22]	2020	33.57	0.9840	Feature Fusion Attention
AECR-Net [23]	2021	37.17†	0.9901†	Contrastive learning
(Proposed)	2026	27.76	0.9623	Tθ + Kθ + Fθ Modular

Table 2 shows the comparison of the proposed method with other state-of-the-art methods.

Traditional vs. Deep Learning Gap: The PSNR value of the traditional DCP method is 19.13 dB, whereas modern deep learning methods such as FFA-Net reach 33.57 dB — an improvement of more than fourteen dB in the PSNR value. The use of deep learning has transformed this field.

Impact of Attention Mechanisms: The FFA-Net method and suggested approach both use attention networks and give better results. Channel plus spatial attention really helped in achieving a better PSNR value. The effectiveness of attention mechanisms has been demonstrated across various vision tasks.

End-to-End Training: Before the AOD-Net method, the training was done separately. With the help of end-to-end training, joint optimisation became possible, which gives better results.

Physics-Guided vs. Pure Learning: the proposed approach is physics-guided (T and K are explicitly modelled), whereas FFA-Net is pure learning. Both approaches work, but physics-guided methods are more interpretable.

Best Case Performance: The best PSNR is 36.43 dB, which indicates that, on some images, highly accurate reconstruction has occurred.

E. Strengths of the Proposed Method

Modular Architecture: The suggested Tθ, Kθ, Fθ approach is modular. This signifies that, in future, if a better transmission estimator comes along, then only that component can be swapped. This flexibility is not present in monolithic architectures like FFA-Net.

Interpretability: We can explicitly visualise the transmission map and atmospheric light. It is helpful in debugging, and a physically meaningful output is obtained. This is not found in pure black-box methods.

Attention at the Right Places: Channel attention is used in the transmission network, which is critical for transmission estimation, while transmission-guided spatial attention is employed in the fusion network. In the atmospheric light network, global pooling is used, which captures the global context. This is perfect for atmospheric light estimation.

Combined Loss Function: A combination of SSIM and L1 loss has been used (0.84 × SSIM + 0.16 × L1). SSIM preserves the structural similarity, whereas L1 loss ensures pixel accuracy. Many older methods only used MSE, which gives blurry results.

The following results are obtained by testing on 100 validation images.

TABLE 3. EPOCH-WISE AVERAGE PSNR AND SSIM VALUES.

ISSN (ONLINE): 2456-5717
ISSN (PRINT): 2456-5717

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol. 12, Issue 10, October 2026

Epoch	Train PSNR	Val PSNR	Val SSIM
1	19.22 dB	20.87 dB	0.8466
10	25.23 dB	26.21 dB	0.9474
25	25.95 dB	26.21 dB	0.9529
40	26.89 dB	27.47 dB	0.9613
50 (Final)	26.98 dB	27.76 dB	0.9623

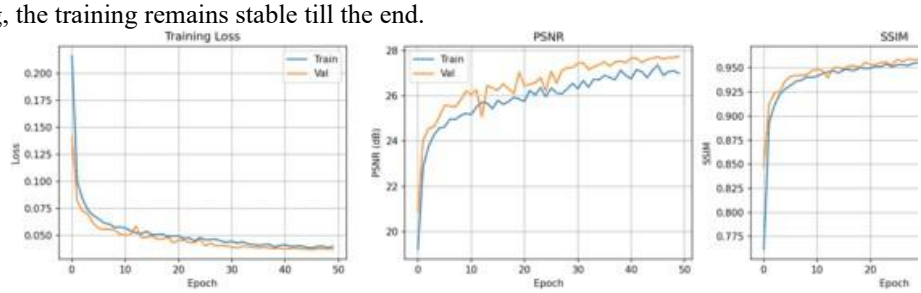
TABLE 4: FINAL RESULTS

Metric	Mean ± Std	Best
PSNR	27.76 ± 3.57 dB	36.43 dB
SSIM	0.9623 ± 0.0251	0.9889
Test Images	100	–

Table 4 gives the best values of PSNR and SSIM. The average PSNR value is 27.76 dB, which is an excellent result. The standard deviation is 3.57 dB, which indicates that some images show better results (easy scenes), while some images show lower values, which indicate dense haze. The best-case PSNR value of 36.43 dB reflects highly accurate reconstruction on favourable scenes. The SSIM value is also very high, so the details of the image are well preserved.

F. Training Progress

The training dataset converges smoothly without overfitting. In epoch 1, the validation PSNR value was 20.87 dB, which increased smoothly to 27.76 dB. By the 50th epoch, the SSIM value also improved from 0.8466 to 0.9623. Due to the presence of cosine annealing, the training remains stable till the end.



(a) Training Loss (b) PSNR curve (c) SSIM curve

Fig. 4. Training progress across epochs: (a) training loss; (b) PSNR (dB); (c) SSIM.

G. Qualitative Results (Visual Comparison)

The PSNR value is excellent, but the real test is through the visual images. The results in Fig.5 shows that the input hazy images are blurred and have low contrast. In the dehazed output, the colours are vibrant, the details are sharp, and there is much improvement in the visibility. The transmission map is also physically meaningful, as distant objects have more haze and darker pixel values, whereas nearby objects have less haze and lighter pixel values.

Hazy Input	Dehazed	Ground Truth

ISSN (ONLINE): 2456-5717
ISSN (PRINT): 2456-5717

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol. 12, Issue 10, October 2026

Hazy Input	Dehazed	Ground Truth

Hazy Input	Dehazed	Ground Truth

Hazy Input	Dehazed	Ground Truth

Fig. 5. Visual comparison of the dehazing results, showing the hazy input, the dehazed image produced by the proposed method, and the corresponding ground truth.

V. Discussion

The suggested multi-component architecture reveals many advantages over the existing methods. The modular design of Tθ, Kθ and Fθ allows separate optimisation of each component, which leads to better overall performance. The combined estimation of the transmission map and atmospheric light provides physical interpretability, which was lacking in black-box approaches [19, 22]. The combined loss function (0.84 × SSIM + 0.16 × L1) is vital for preserving structural information and maintaining pixel-level accuracy. The attention mechanism employed in the network plays a crucial role in performance. The channel attention in Tθ gives the network the capacity to focus on the informative feature channels for transmission estimation. The transmission-guided spatial attention in the fusion network provides support in identifying regions with different haze densities [30, 31]. The global pooling in Kθ effectively captures the global context, which is important for estimating atmospheric light, as it is a scene-level property.

A limitation of our current work is the relatively small training dataset (400 images). Training on larger datasets, such as the full RESIDE-β [29] with thousands of image pairs, could further improve generalisation. Additionally, the current model has been evaluated only on synthetic haze; evaluation on real-world hazy images would provide stronger evidence of practical applicability.

VI. Conclusion and Future Work

In this paper, an effective deep learning-based image dehazing method has been presented. The key contributions are as

follows: the novel Tθ, Kθ, Fθ architecture has physical meaning and is modular. Under the adopted evaluation protocol, the PSNR value is 27.76 dB and the SSIM value is 0.9623. These values are obtained on the RESIDE-Outdoor dataset [29]. In the best case, the PSNR value is 36.43 dB. The PSNR value indicates that the suggested method has excellent reconstruction capability. The proposed work also shows the effective use of a combination of channel attention, transmission-guided spatial attention, global context aggregation, and the combined loss function.

In the future, more work can be done on video dehazing and real-time optimization. Testing can be done on more diverse datasets. The integration of vision transformers and diffusion-based approaches for image dehazing presents promising research directions. Exploring domain adaptation techniques for transferring knowledge from synthetic to real-world hazy images is another important future direction. Overall, the method is well suited for practical applications.

ACKNOWLEDGMENT

We would like to thank the creators of the RESIDE dataset [29] for making the benchmark publicly available for research purposes.

REFERENCES

[1] J. Janai, F. Güney, A. Behl, and A. Geiger, "Computer Vision for Autonomous Vehicles: Problems, Datasets and State of the Art," Found. Trends Comput. Graph. Vis., vol. 12, no. 1–3, pp. 1–308, 2020.
[2] Y. Tian, P. Luo, X. Wang, and X. Tang, "Deep Learning Strong Parts for Pedestrian Detection," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), pp. 1904–1912, 2015.
[3] P. Zhu, L. Wen, D. Du, X. Bian, H. Fan, Q. Hu, and H. Ling, "Detection and Tracking Meet Drones Challenge," IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 11, pp. 7380–7399, 2022.
[4] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, 2017.
[5] G. Cheng and J. Han, "A Survey on Object Detection in Optical Remote Sensing Images," ISPRS J. Photogrammetry Remote Sens., vol. 117, pp. 11–28, 2016.
[6] S. G. Narasimhan and S. K. Nayar, "Vision and the Atmosphere," Int. J. Comput. Vis., vol. 48, no. 3, pp. 233–254, 2002.
[7] S. G. Narasimhan and S. K. Nayar, "Contrast Restoration of Weather Degraded Images," IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 6, pp. 713–724, 2003.
[8] K. He, J. Sun, and X. Tang, "Single Image Haze Removal Using Dark Channel Prior," IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI), vol. 33, no. 12, pp. 2341–2353, 2011.
[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.
[10] C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 2, pp. 295–307, 2016.
[11] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising," IEEE Trans. Image Process., vol. 26, no. 7, pp. 3142–3155, 2017.
[12] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Multi-Stage Progressive Image Restoration," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 14821–14831, 2021.
[13] Q. Zhu, J. Mai, and L. Shao, "A Fast Single Image Haze Removal Algorithm Using Color Attenuation Prior," IEEE Trans. Image Process. (TIP), vol. 24, no. 11, pp. 3522–3533, 2015.
[14] D. Berman, T. Treibitz, and S. Avidan, "Non-local Image Dehazing," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 1674–1682, 2016.
[15] R. Fattal, "Single Image Dehazing," ACM Trans. Graph., vol. 27, no. 3, pp. 1–9, 2008.
[16] J.-P. Tarel and N. Hautiere, "Fast Visibility Restoration from a Single Color or Gray Level Image," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), pp. 2201–2208, 2009.
[17] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An End-to-End System for Single Image Haze Removal," IEEE Trans. Image Process. (TIP), vol. 25, no. 11, pp. 5187–5198, 2016.
[18] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single Image Dehazing via Multi-Scale Convolutional Neural Networks," in Proc. European Conf. Comput. Vis. (ECCV), pp. 154–169, 2016.
[19] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "AOD-Net: All-in-One Dehazing Network," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), pp. 4770–4778, 2017.
[20] W. Ren, L. Ma, J. Zhang, J. Pan, X. Cao, W. Liu, and M.-H. Yang, "Gated Fusion Network for Single Image Dehazing," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 3253–3261, 2018.
[21] X. Liu, Y. Ma, Z. Shi, and J. Chen, "GridDehazeNet: Attention-Based Multi-Scale Network for Image Dehazing," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), pp. 7314–7323, 2019.
[22] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "FFA-Net: Feature Fusion Attention Network for Single Image Dehazing," in Proc. AAAI Conf. Artif. Intell., vol. 34, no. 07, pp. 11908–11915, 2020.
[23] H. Wu, Y. Qu, S. Lin, J. Zhou, R. Qiao, Z. Zhang, Y. Xie, and L. Ma, "Contrastive Learning for Compact Single Image Dehazing," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 10551–10560, 2021.
[24] Z. Tu, H. Talebi, H. Zhang, F. Yang, P. Milanfar, A. Bovik, and Y. Li, "MAXIM: Multi-Axis MLP for Image Processing," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 5769–5780, 2022.
[25] Y. Song, Z. He, H. Qian, and X. Du, "Vision Transformers for Single Image Dehazing," IEEE Trans. Image Process., vol. 32, pp. 1927–1941, 2023.
[26] Y. Ye, C. Yu, Y. Chang, L. Zhu, X. Zhao, L. Yan, and Y. Tian, "Perceiving and Modeling Density for Image Dehazing," in Proc. European Conf. Comput. Vis. (ECCV), pp. 130–145, 2022.
[27] Y. Zheng, J. Zhan, S. He, J. Dong, and Y. Du, "Curricular Contrastive Regularization for Physics-Aware Single Image Dehazing," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2023.
[28] R.-Q. Wu, Z.-P. Duan, C.-L. Guo, Z. Chai, and C. Li, "RIDCP: Revitalizing Real Image Dehazing via High-Quality Codebook Priors," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2023.
[29] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking Single-Image Dehazing and Beyond," IEEE Trans. Image Process. (TIP), vol. 28, no. 1, pp. 492–505, 2019.

[30] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 7132–7141, 2018.
[31] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in Proc. European Conf. Comput. Vis. (ECCV), pp. 3–19, 2018.