

ECA LDNet A Dual Attention Network with Atmospheric Scattering Guidance for Single Image Dehazing

Manpreet Singh

Boston University

Rohith Reddy Bellibatl

Florida International University

Sai Deekshith Lekkala

Kakatiya University

Rahul Joshi

`rahul.j@sitpune.edu.in`

Symbiosis International University

Method Article

Keywords: Atmospheric scattering model, compact neural network, depthwise-separable convolution, efficient channel attention, model-guided learning, pixel attention, image restoration, single-image dehazing

Posted Date: August 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10626300/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

ECA LDNet A Dual Attention Network with Atmospheric Scattering Guidance for Single Image Dehazing

Manpreet Singh ¹, Rohith Reddy Bellibatlu²,
Sai Deekshith Lekkala³, Rahul Joshi ^{4*}

¹Boston University, Boston, MA, 02215, USA.

²Florida International University, Miami, FL, 33199, USA.

³Kakatiya Institute of Technology and Science, Warangal, India.

^{4*}Symbiosis Institute of Technology, Pune, Symbiosis International
(Deemed University), Pune, India.

*Corresponding author(s). E-mail(s): rahulj@sitpune.edu.in;
Contributing authors: manni@bu.edu; rohithreddybc@gmail.com;
b23ai005@kitsw.ac.in;

Abstract

Single-image dehazing requires balancing restoration fidelity against model compactness: high-capacity attention and Transformer models achieve strong accuracy, whereas very small convolutional models often sacrifice fidelity. We present ECA-LDNet, a 1.482-million-parameter depthwise-separable U-Net that combines reduction-free Efficient Channel Attention (ECA), pixel attention, and a softly blended atmospheric-scattering-model branch. The auxiliary branch predicts latent transmission-like and atmospheric-light-like variables and contributes 8% of the output; these are guidance signals, not physically calibrated estimates. ECA-LDNet achieves the 32.53 dB peak signal-to-noise ratio (PSNR) and 0.9717 structural similarity (SSIM) on SOTS-Indoor, 31.94 dB / 0.9769 on SOTS-Outdoor, and 30.20 dB / 0.9641 on RESIDE-6K. Because training uses a mixed RESIDE-6K and ITS corpus, published cross-method results are treated as protocol-unmatched references rather than controlled rankings. Controlled single-seed ablations show that ECA gives the most consistent PSNR gain per added parameter, whereas pixel attention and the atmospheric branch trade PSNR against structural similarity. At 256×256 the model needs 2.171 GMACs by a layer-wise counter released and runs at about 12–14 frames/s on an NVIDIA Tesla P100. Under one identical evaluation pipeline, a 22-image comparison of five

released pretrained models gives ECA-LDNet the highest overall PSNR, though the set is too small for population-level ranking, and the models differ in training data. ECA-LDNet is therefore a compact design with explicitly characterized component trade-offs rather than a state-of-the-art result. Code, weights, and evaluation scripts are publicly released.

Keywords: Atmospheric scattering model, compact neural network, depthwise-separable convolution, efficient channel attention, model-guided learning, pixel attention, image restoration, single-image dehazing

1 Introduction

1.1 Haze, Visual Degradation, and the Need for Dehazing

Atmospheric haze, fog, and smoke degrade the quality of images captured in outdoor and adverse-weather environments. As light traverses a scattering medium, suspended particles attenuate the radiance reflected from a scene and add an additive airlight component, producing observations with reduced contrast, washed-out colors, and obscured fine detail. This process is described by the Atmospheric Scattering Model (ASM), formalized for computer vision by Narasimhan and Nayar [1], which expresses a hazy observation as $I(x) = J(x)t(x) + A(1 - t(x))$, where J is the haze-free radiance, A is the global atmospheric light, and $t(x)$ is the medium transmission.

The practical consequences of haze extend well beyond aesthetics. Autonomous driving, video surveillance, remote sensing, and outdoor robotics all rely on clear visual input, and downstream tasks such as object detection, semantic segmentation, and tracking degrade sharply when the input is hazy. Restoring a clean image J from a single hazy observation, the single-image dehazing problem, is therefore a safety-relevant enabling technology for vision-based perception systems. Because a single image provides only one equation per pixel for the unknown radiance, transmission, and atmospheric light, the inversion is severely ill-posed, motivating decades of work on priors and, more recently, learned models.

1.2 From Hand-Crafted Priors to Learned Restoration

Classical dehazing addresses the ill-posed inversion with hand-crafted priors. Fattal decorrelated surface shading from transmission [9]; He et al.’s Dark Channel Prior (DCP) observed that haze-free outdoor patches contain near-zero pixels in some color channel, enabling direct estimation of transmission and atmospheric light [2]. Subsequent priors include the color attenuation prior [10], the non-local haze-line prior [11], and boundary-constrained contextual regularization [12]. These methods are interpretable but degrade on sky regions, bright objects, and whenever the atmospheric light is misestimated, and they lack perceptual fidelity. In the SOTS-Indoor reference comparison used in this paper, DCP is listed at 16.62 dB PSNR.

The first wave of learning-based methods replaced these priors with CNNs that regress ASM-related variables. DehazeNet [13] and MSCNN [14] learn the transmission map and then recover the clean image by inverting the model, a two-stage pipeline in which small transmission errors are amplified by the division and the heuristically estimated atmospheric light compounds the error. AOD-Net algebraically reformulated the ASM into a single learned variable, yielding an extremely compact network (0.002M parameters) that is reported at 20.51 dB on SOTS-Indoor [15]. GFN sidestepped explicit parameter regression with a gated fusion of derived inputs [16]. These methods illustrate a recurring trade-off: physically interpretable formulations are compact, but their output quality can be limited when real haze deviates from the idealized homogeneous model or when division-based recovery propagates estimation error.

Attention-based networks subsequently relaxed strict ASM constraints and moved toward direct end-to-end restoration. Building on Squeeze-and-Excitation channel recalibration [17] and the channel-then-spatial design of CBAM [18], FFA-Net stacks channel attention and pixel attention and is reported at 36.39 dB on SOTS-Indoor [6], while GridDehazeNet embeds attention in a multi-scale grid and is reported at 32.16 dB [19]. These results motivate attention-based restoration, although the reported methods differ in parameter count, training protocol, and evaluation setting.

1.3 The Accuracy-versus-Efficiency Gap

The current high-PSNR frontier is defined by Transformer architectures. DehazeFormer adapts the Swin Transformer to dehazing and scales to a large variant, DehazeFormer-L, reported at 25.44M parameters and 40.05 dB on SOTS-Indoor [3], while general-purpose restoration Transformers such as Restormer [20], Uformer [21], and SwinIR [22] rely on multi-head self-attention whose cost grows quadratically with spatial resolution, yielding tens of millions of parameters and hundreds of GMACs. Recent efficient designs explicitly attack this overhead, for example MB-TaylorFormer through a Taylor-approximated linear attention [23] and DPTE-Net by replacing self-attention with pooling [24], evidencing an active push toward lightweight models.

This produces a broad spectrum between high-capacity restoration models and very small physics-regression CNNs. Rather than claiming that the intermediate regime is unoccupied, we investigate one specific design point: a depthwise-separable U-Net that combines reduction-free channel attention, pixel attention, and a weak atmospheric-model branch. The research question is whether these components provide measurable benefits under a compact parameter budget.

1.4 Gap Statement and Contributions

We study whether channel attention, pixel attention, and atmospheric-model guidance can be integrated into a compact convolutional dehazing network while retaining useful restoration fidelity. ECA-LDNet is a 1.482M-parameter depthwise-separable U-Net designed for this operating regime.

The contribution is not a new attention operator or physical imaging model. Instead, it is the integration of reduction-free channel attention, pixel attention, and

softly blended atmospheric-model guidance within a compact depthwise-separable network, together with a controlled decomposition of their parameter costs and metric-specific effects.

The main contributions are framed at the level supported by the experiments:

- **An integrated compact dehazing architecture.** We compose reduction-free Efficient Channel Attention [5] (a single 1D convolution over global descriptors) with a compact pixel-attention head for spatial gating, applying the two in sequence within depthwise-separable residual blocks [4], and add a soft atmospheric-model auxiliary branch. This recombines FFA-Net’s channel-plus-pixel philosophy [6] and the classical scattering model [1], [2] in a compact, fully convolutional form. We present this composition as an architectural design; controlled knock-out results show that the complete architecture achieves the highest PSNR across all evaluated datasets, although pixel attention presents a trade-off between PSNR and SSIM.
- **A softly blended atmospheric-model auxiliary branch.** Rather than imposing the ASM as a hard output constraint, we add a compact module that predicts atmospheric-light-like and transmission-like latent variables at the bottleneck and computes an ASM-derived auxiliary estimate, blended at only 8% of the output. This retains an explicit link to the scattering model of classical priors [2], [1] while letting the learned CNN dominate. The latent variables are not claimed to be physically calibrated, and the optimal blend weight is not sensitivity-tested here.
- **Controlled component and cost analysis.** We evaluate variants without ECA, pixel attention, atmospheric-model guidance, and both attention mechanisms under an identical single-seed epoch-28 protocol. ECA produces the strongest PSNR contribution while adding only 27 parameters. Atmospheric-model guidance produces a smaller PSNR benefit with a structural-metric trade-off. Pixel attention also improves PSNR but reduces SSIM, indicating a fidelity trade-off. Parameter analysis shows that pixel attention and atmospheric-model guidance add 109,433 and 51,588 parameters, respectively.
- **Explicitly bounded empirical positioning.** ECA-LDNet contains 1.482M parameters and runs at approximately 12–14 FPS at 256×256 on an NVIDIA Tesla P100. It attains 30.20 dB/0.9641 SSIM on RESIDE-6K but remains below the listed DehazeFormer variants in fidelity. Published benchmark values are reported as contextual references because training corpora and preprocessing differ. A separate 22-image common-image comparison, in which all five released models are evaluated under one identical pipeline with no model-specific output rescaling, gives ECA-LDNet the highest overall PSNR and competitive SSIM; the set is small and the models differ in training data, so it is reported as case-level rather than population-level evidence.

Our main benchmark comparisons are reported on the RESIDE/SOTS benchmark [25]; the common-image comparison in Section 4 additionally samples Haze-1K, I-HAZE, and O-HAZE images.

1.5 Paper Organization

The remainder of this paper is organized as follows. Section 2 reviews related work on classical prior-based dehazing, early CNN and physics-regression methods, attention mechanisms, and Transformer-based restoration. Section 3 details the ECA-LDNet architecture: the depthwise-separable backbone, the dual-attention block, the atmospheric-model auxiliary branch, the ASM-derived estimate, and the composite training objective. Section 4 presents the experimental setup, datasets, and quantitative and qualitative comparisons against representative baselines, together with the multi-stage training-strategy study. Section 5 discusses limitations, including single-seed evaluation, synthetic-data dependence, and unresolved blend-weight sensitivity. Section 6 concludes the paper.

2 Related Work

Single-image dehazing has developed through several overlapping paradigms: physics-driven prior-based inversion of the Atmospheric Scattering Model (ASM), data-driven CNN parameter regression, attention-augmented restoration, and, more recently, Transformer-based and compact architectures. We review each line and position ECA-LDNet at the intersection of compactness, attention, and atmospheric-model guidance.

2.1 Prior-Based Dehazing and the Atmospheric Scattering Model

The physical foundation for most dehazing work is the Atmospheric Scattering Model (ASM), as formalized for computer vision by Narasimhan and Nayar [1], which expresses a hazy observation as

$$I(x) = J(x)t(x) + A(1 - t(x)), \quad t(x) = e^{-\beta d(x)}, \quad (1)$$

where J is the haze-free radiance, A the global atmospheric light, $t(x) \in [0, 1]$ the medium transmission, β the scattering coefficient, and $d(x)$ scene depth. Recovering J from a single observation is ill-posed, since both t and A are unknown, and the classical literature confronts this by introducing hand-crafted priors. Fattal [9] decorrelated surface shading from transmission under a local statistical independence assumption. He et al. [2] proposed the Dark Channel Prior (DCP), observing that haze-free outdoor patches contain near-zero intensities in at least one color channel, which enables direct estimation of transmission and atmospheric light; DCP remains the canonical baseline and attains 16.62 dB on SOTS-Indoor in our comparison. Zhu et al. [10] modeled scene depth as a linear function of brightness minus saturation through a supervised Color Attenuation Prior, while Berman et al. [11] replaced patch-based assumptions with a global non-local prior in which haze-free colors form clusters that degrade into haze-lines under scattering. Meng et al. [12] imposed an explicit boundary constraint on transmission combined with contextual regularization.

These priors are interpretable and require no training, but they degrade in sky regions, on bright objects, and whenever the atmospheric light is misestimated, and

they offer no mechanism for perceptual fidelity. ECA-LDNet retains an explicit link to the scattering model through an atmospheric-model auxiliary branch that predicts atmospheric-light-like and transmission-like latent variables and applies the scattering inversion

$$J_{\text{asm}} = \frac{I - A}{t + \epsilon} + A, \quad \epsilon = 0.1, \quad (2)$$

as a correction head. Critically, this ASM-derived estimate is blended at only 8% of the output,

$$\text{out} = 0.92 \text{ raw} + 0.08 J_{\text{asm}}, \quad (3)$$

so the learned network dominates while the scattering-model estimate acts as a weak auxiliary signal. The resulting 32.53 dB / 0.9717 SSIM on SOTS-Indoor at 1.482M parameters is substantially higher than the published DCP reference value, although the comparison is not controlled because the methods use different protocols.

2.2 CNN-Based Physics-Regression and End-to-End Methods

The first wave of learning-based dehazing replaced hand-crafted priors with convolutional networks that regress the parameters of the ASM. DehazeNet [13] and MSCNN [14] both learn the transmission map t and then recover J by inverting the ASM with a separately estimated atmospheric light. This two-stage design is fragile: small transmission errors are amplified by the division-based recovery, and A is estimated heuristically rather than jointly, so errors compound. AOD-Net [15] addressed this coupling by algebraically reformulating the ASM into a single learned variable and regressing it end-to-end; the resulting network is extremely compact (0.002M parameters) but plateaus in fidelity at 20.51 dB on SOTS-Indoor, because a shallow model constrained to the ASM form has limited representational capacity. Subsequent methods loosened the physical constraint: GFN [16] fused white-balance, contrast-enhanced, and gamma-corrected derivations of the input rather than regressing physical parameters; DCPDN [26] embedded the ASM within a densely connected pyramid network to jointly learn t , A , and J under an edge-preserving objective; and EPDN [27] abandoned the physical model entirely, casting dehazing as a pix2pix image-translation problem with an adversarial enhancer.

The recurring lesson across this family is that physics-only or tightly physics-constrained CNNs can be limited in restoration quality, with DehazeNet (19.82 dB), MSCNN (19.84 dB), AOD-Net (20.51 dB), and GFN (22.30 dB) on SOTS-Indoor as reported in the respective papers and the RESIDE benchmark [25]. This limitation is consistent with the difficulty of modeling real haze using an idealized homogeneous model, the ill-posedness of transmission estimation, and the error amplification introduced by division-based recovery. ECA-LDNet is a direct end-to-end restoration network in which the CNN carries 92% of the output weight, while an atmospheric-model auxiliary branch contributes the remaining 8%. This design encourages atmospheric-model consistency without imposing the ASM as a hard output constraint.

2.3 Attention-Based Dehazing and Channel/Spatial Attention

Because haze is spatially non-uniform and channel responses carry uneven information, attention mechanisms have become central to modern dehazing. The lineage we build on begins with Squeeze-and-Excitation (SE) networks [17], which recalibrate channels through global average pooling followed by a dimensionality-reducing fully-connected bottleneck, and CBAM [18], which appends spatial attention sequentially after channel attention. FFA-Net [6] adapts this CBAM-style design specifically to dehazing: its Feature Attention block stacks channel attention with pixel attention, attaining 36.39 dB on SOTS-Indoor, the strongest published number in our comparison. GridDehazeNet [19] embeds channel attention within a multi-scale grid architecture, reaching 32.16 dB / 0.9836 at 0.956M parameters. AECR-Net [33] pushes the compact regime further, using contrastive-learning regularization to reach 37.17 dB on SOTS-Indoor at roughly 2.6M parameters; it is directly relevant to our efficiency-oriented goal, though it remains larger and more accurate than ECA-LDNet and does not target the compact, roughly 2-GMAC compute budget we emphasize.

The key compactness insight we adopt comes from ECA-Net [5], which replaces the fully connected dimensionality reduction of SE-style attention with a parameter-frugal local cross-channel interaction implemented as a single 1D convolution. ECA-LDNet preserves the channel-plus-pixel philosophy of FFA-Net and CBAM but uses ECA for the channel branch and a compact pixel-attention head for spatial gating. At 1.482M parameters, ECA-LDNet reaches 32.53 dB / 0.9717 on SOTS-Indoor. GridDehazeNet has slightly lower published PSNR but higher SSIM with fewer parameters, while FFA-Net and the listed DehazeFormer variants achieve substantially higher fidelity. These values are contextual because their training protocols differ.

2.4 Transformer-Based and Compact Methods

Transformer architectures now define the high-PSNR frontier of single-image dehazing, but at substantial parameter and compute cost. DehazeFormer [3] adapts the Swin Transformer with modified normalization, a rescaled spatial-aggregation scheme, and a revised activation, scaling across several variants; in our comparison these attain 36.82 dB (S, 1.283M), 37.84 dB (B, 2.514M), and 40.05 dB (L, 25.44M) on SOTS-Indoor. General-purpose restoration Transformers manage the cost of self-attention through different approximations: Restormer [20] applies channel-transposed (MDTA) attention with a gated feed-forward network, Uformer [21] employs locally-enhanced windowed (LeWin) attention in a U-shaped backbone, and SwinIR [22] stacks residual Swin Transformer blocks. These designs yield strong fidelity but still incur large parameter and MAC budgets, since unrestricted global self-attention scales quadratically with spatial resolution. Hybrid CNN-Transformer designs such as Dehamer [34] inject transmission-aware 3D positional embeddings to couple convolutional locality with global attention. Recent work explicitly targets the attention overhead: MB-TaylorFormer [23] Taylor-approximates softmax attention to achieve linear complexity, and DPTE-Net [24] replaces self-attention with pooling in a distilled encoder, evidencing an active push toward lightweight Transformers.

A parallel and highly relevant line of work pushes the accuracy frontier with efficient *convolutional* designs, which is the regime ECA-LDNet targets. DEA-Net [28] combines detail-enhanced convolution with content-guided attention; ConvIR [31] shows that carefully designed convolutions can rival Transformer restoration models; and gUNet [29] argues that much of the gain attributed to attention can arise from the U-Net design itself. On the physics-aware side, C2PNet [30] enforces atmospheric-scattering consistency through curricular contrastive regularization. These methods define a substantially higher fidelity envelope. ECA-LDNet does not compete with that envelope and instead examines a compact, approximately 1.5M-parameter design.

Compact restoration has long used depthwise-separable convolutions introduced in MobileNets [4], which ECA-LDNet adopts throughout its U-Net. ECA-LDNet combines this backbone with an atmospheric-model auxiliary branch and dual attention, reaching 32.53 dB / 0.9717 at 1.482M parameters. The composite training objective combines VGG perceptual loss [8], SSIM [7], Charbonnier, Sobel-edge, and FFT-magnitude terms, and benchmarking uses the RESIDE/SOTS protocol [25].

ECA-LDNet therefore sits deliberately on the CNN side of the compactness-versus-quality trade-off. It is a 1.482M-parameter depthwise-separable U-Net that delivers 32.53 dB / 0.9717 on SOTS-Indoor and runs at approximately 12–14 FPS on the measured P100 setup. On RESIDE-6K it obtains 30.20 dB / 0.9641, below the listed DehazeFormer variants. The relevant contribution is the integration of reduction-free channel attention, pixel attention, depthwise-separable convolution, and soft atmospheric-model guidance, together with a controlled analysis of their individual effects.

3 Proposed Method

3.1 Overview

ECA-LDNet is a compact single-image dehazing convolutional network. It accepts a hazy RGB image normalized to $[0,1]$ and produces a restored image of the same spatial dimensions. Training and evaluation in this paper use $256 \times 256 \times 3$ inputs; because the network is fully convolutional, other resolutions are structurally possible but not evaluated here. The complete model contains 1,481,871 parameters (1,466,985 trainable plus 14,886 non-trainable batch-normalization statistics; ≈ 1.482 M total, ≈ 5.65 MB), requires 2.171 GMACs (4.343 GFLOPs) at 256×256 by an analytic layer-wise count (Section 4.9), and runs at approximately 72–82 ms per image (≈ 12 –14 FPS) on the measured NVIDIA Tesla P100 setup. The overall architecture is shown in Fig. 1.

Architecturally, ECA-LDNet is a U-Net encoder-decoder with skip connections, in which every convolutional unit is realized as a depthwise-separable block to drive parameter economy [4]. The design combines two historically distinct threads in the dehazing literature: atmospheric-model inversion [1], [2] and data-driven attention-based restoration [6], [5]. Rather than constraining the output to the idealized ASM form, ECA-LDNet lets a learned CNN dominate while retaining a compact atmospheric-model auxiliary branch blended at only 8%. This encourages atmospheric-model consistency without claiming physical calibration. Channel attention follows

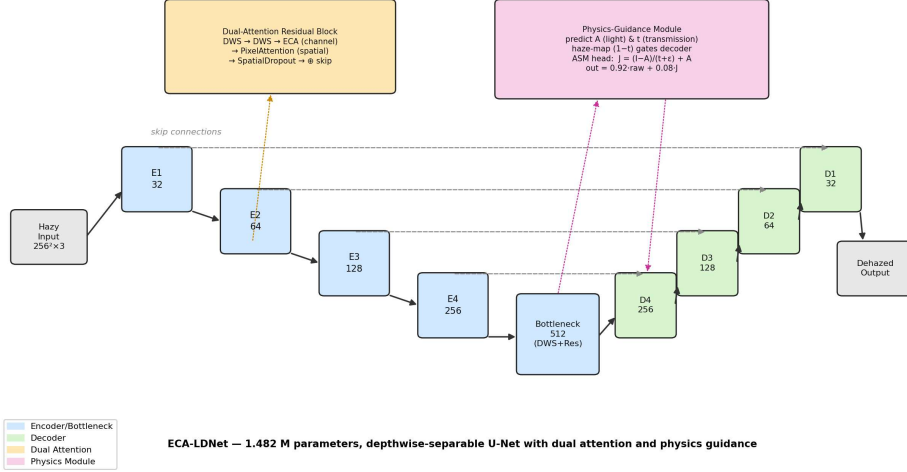


Fig. 1 ECA-LDNet architecture: a depthwise-separable U-Net with dual-attention residual blocks (ECA channel attention followed by pixel attention) and an atmospheric-model auxiliary branch that predicts atmospheric-light-like and transmission-like latent variables and forms an ASM-derived auxiliary estimate. *Takeaway:* the learned CNN path carries 92% of the output and the ASM-derived branch only 8%, so the network dominates while the scattering model acts as a weak prior.

the reduction-free Efficient Channel Attention formulation [5], and spatial attention follows the pixel-attention idea popularized for dehazing by FFA-Net [6].

We characterize the contribution of the multi-stage training schedule through a training-strategy study (Stages S1–S5, reported in Section 4), which varies the optimization schedule rather than the architecture.

3.2 Atmospheric Scattering Model Preliminaries

The Atmospheric Scattering Model (ASM), formalized for computer vision by Narasimhan and Nayar [1], expresses a hazy observation as a weighted sum of the scene radiance and the global atmospheric light:

$$I(x) = J(x)t(x) + A(1 - t(x)) \quad (4)$$

where I is the observed hazy image, J is the haze-free scene radiance, A is the global atmospheric light, and $t(x) \in [0, 1]$ is the medium transmission, with $t(x) = \exp(-\beta d(x))$ for scattering coefficient β and scene depth $d(x)$. Inverting this relation recovers the clean radiance:

$$J = \frac{I - A}{t} \quad (5)$$

This inversion is ill-posed for a single image and is numerically unstable when t is small, since transmission errors are amplified by the division [2], [13]. ECA-LDNet therefore uses the ASM not as a hard output constraint but as a soft, blended guide: it predicts A and t explicitly and applies the inversion only as a minority correction term.

3.3 Encoder-Decoder with Depthwise-Separable Blocks

The backbone is a symmetric U-Net with skip connections. The encoder uses filter widths [32, 64, 128, 256, 512] across its successive stages, and the decoder mirrors this with [256, 128, 64, 32], concatenating encoder features through skip connections at matching scales.

The core computational unit throughout the network is the depthwise-separable convolution block, which is the principal efficiency driver [4]. Each block applies the sequence:

$$\text{DepthwiseConv2D} \rightarrow \text{BN} \rightarrow \text{ReLU6} \rightarrow 1 \times 1 \text{ Conv} \rightarrow \text{BN} \rightarrow \text{ReLU6} \quad (6)$$

The depthwise stage performs per-channel spatial filtering and the subsequent 1×1 (pointwise) convolution mixes channel information, factorizing a standard convolution into a substantially cheaper two-step operation. Batch normalization and the bounded ReLU6 activation follow each convolution.

3.4 Dual-Attention Residual Block

The representational workhorse of the network is the residual block of dual-attention. Each residual block comprises two depthwise-separable blocks followed by a dual-attention module, a SpatialDropout2D layer with rate 0.05, and an additive skip connection (with a 1×1 projection applied on the skip path when the channel count changes).

The dual-attention module applies channel attention followed by spatial attention, in the order ECA-then-PixelAttention, combining FFA-Net’s dual channel+pixel philosophy [6] with ECA’s reduction-free channel branch [5].

3.4.1 ECA Channel Attention

Channel attention follows Efficient Channel Attention [5], which replaces the dimensionality-reducing fully-connected bottleneck of SE-style modules [17] with a parameter-frugal local cross-channel interaction. Given an input feature map, a global average pooling (GAP) produces a per-channel descriptor, over which a single 1D convolution with kernel size $k = 3$ (no bias) models local cross-channel dependence, followed by a sigmoid that yields the channel weights:

$$\mathbf{w} = \sigma(\text{Conv1D}_{k=3}(\text{GAP}(\mathbf{X}))), \quad \mathbf{X}' = \mathbf{w} \odot \mathbf{X} \quad (7)$$

where σ is the sigmoid function and $\odot$ denotes channel-wise reweighting. With kernel size $k = 3$ and no bias, this branch adds only 3 parameters per ECA module, avoiding the SE-style reduction that has been shown to affect channel-attention learning [5].

3.4.2 Pixel Attention

Spatial attention follows the pixel-attention design popularized for dehazing by FFA-Net [6], learning *where* haze concentrates. It is implemented as a 1×1 convolution to $C/4$ channels with ReLU (where C is the input channel count), followed by a 1×1

convolution to a single channel with sigmoid, producing a spatial mask that multiplies the feature map:

$$\mathbf{m} = \sigma\left(\text{Conv}_{1 \times 1}^1\left(\text{ReLU}\left(\text{Conv}_{1 \times 1}^{C/4}(\mathbf{X}')\right)\right)\right), \quad \mathbf{X}'' = \mathbf{m} \odot \mathbf{X}' \quad (8)$$

where $\mathbf{m}$ is a single-channel spatial mask broadcast across channels. This dual ECA-then-PA design is a compact analog of FFA-Net’s CA+PA pairing and the channel-then-spatial ordering of CBAM [18].

3.5 Bottleneck

At the deepest stage, the bottleneck applies a depthwise-separable block expanding to 512 channels, followed by one dual-attention residual block and a Dropout layer with rate 0.10. This bottleneck representation also serves as the input to the atmospheric-model auxiliary branch described next.

3.6 Atmospheric-Model Auxiliary Branch

A compact auxiliary branch uses the bottleneck features to predict atmospheric-light-like and transmission-like latent variables used in the ASM-derived estimate.

The global atmospheric light A is predicted as a 3-vector (one value per RGB channel) through a global average pooling followed by a dense head of widths 64, 32, and 3 with a final sigmoid:

$$A = \sigma(\text{Dense}_3(\text{Dense}_{32}(\text{Dense}_{64}(\text{GAP}(\mathbf{F}_{\text{bottleneck}})))) \quad (9)$$

The transmission map t is estimated by a 1×1 convolution to 32 channels followed by BN and ReLU, then a 1×1 convolution to a single channel with sigmoid, computed at 16×16 resolution and bilinearly upsampled to the decoder feature resolutions:

$$t = \sigma\left(\text{Conv}_{1 \times 1}^1\left(\text{ReLU}\left(\text{BN}\left(\text{Conv}_{1 \times 1}^{32}(\mathbf{F}_{\text{bottleneck}})\right)\right)\right)\right) \quad (10)$$

The estimated transmission is converted into a haze map $(1 - t)$, which gates the decoder features at two scales. This guidance makes the decoder spatially responsive to the branch’s haze-density proxy, while leaving the final radiance estimate primarily to the learned decoder.

3.7 ASM-Derived Auxiliary Estimate

The final stage applies an explicit ASM inversion as a minority correction to the CNN output. Using the predicted A and t , the ASM inversion is computed with a numerical stabilizer $\epsilon = 0.1$ in the denominator, yielding an auxiliary estimate J_{asm} :

$$J_{\text{asm}} = \frac{I - A}{t + \epsilon} + A \quad (11)$$

The final output blends the raw CNN prediction with this ASM-derived auxiliary estimate at a fixed 92%/8% ratio:

$$\hat{J} = 0.92 \cdot \hat{J}_{\text{raw}} + 0.08 \cdot J_{\text{asm}} \quad (12)$$

This 8% blend links the output to the same scattering model used by classical priors [1], [2], while the CNN carries 92% of the output weight. The stabilizer $\epsilon = 0.1$ damps the inversion. The branch is intended to provide an ASM-inspired auxiliary estimate; it is not evidence that the latent variables are physically calibrated.

3.8 Composite Loss Function

ECA-LDNet is trained with a composite objective that combines a robust pixel-fidelity term with structural, edge, frequency, and perceptual constraints. The total loss is

$$\mathcal{L} = 0.35 \mathcal{L}_{\text{Char}} + 0.20 (1 - \text{SSIM}) + 0.10 \mathcal{L}_{\text{Edge}} + 0.10 \mathcal{L}_{\text{FFT}} + 0.25 \mathcal{L}_{\text{VGG}} \quad (13)$$

with the following terms:

- **Charbonnier loss** ($\mathcal{L}_{\text{Char}}$, weight 0.35): a robust ℓ_1 -like pixel fidelity term, $\mathcal{L}_{\text{Char}} = \frac{1}{N} \sum_i \sqrt{(J_i - \hat{J}_i)^2 + \epsilon_c^2}$ with $\epsilon_c = 10^{-3}$.
- **SSIM loss** ($1 - \text{SSIM}$, weight 0.20): enforces structural similarity between prediction and ground truth [7].
- **Edge loss** ($\mathcal{L}_{\text{Edge}}$, weight 0.10): a Sobel-gradient consistency term that preserves edges.
- **FFT loss** ($\mathcal{L}_{\text{FFT}}$, weight 0.10): an FFT-magnitude term that aligns the frequency spectra of prediction and target.
- **Perceptual loss** ($\mathcal{L}_{\text{VGG}}$, weight 0.25): a VGG16 perceptual loss [8] computed over features from layers block1_conv2, block2_conv2, and block3_conv3 of a frozen ImageNet-pretrained network.

In the final polish stage, the term weights are reweighted to 0.40 (Charbonnier), 0.20 (SSIM), 0.10 (Edge), 0.05 (FFT), and 0.25 (perceptual), shifting emphasis toward pixel fidelity by reducing the frequency term while holding the perceptual weight fixed at 0.25.

4 Experiments and Results

4.1 Datasets

We train and evaluate ECA-LDNet on the RESIDE benchmark suite [25], a standard benchmark family for single-image dehazing. The training corpus combines the RESIDE-6K training partition with the Indoor Training Set (ITS) of RESIDE, which provides paired hazy/clean indoor images synthesized under the atmospheric scattering model [1]. The combined pool is divided into training and validation subsets using a 90/10 split with a fixed random seed (42) to make the split reproducible. Because the split is random over the combined pool rather than scene-stratified, possible overlap of clean-scene identities across splits has not been fully excluded; we therefore treat validation values as internal model-selection signals rather than as evidence of benchmark generalization.

Evaluation is conducted on three held-out benchmarks: two drawn from the Synthetic Objective Testing Set (SOTS) of the RESIDE benchmark [25], and the mixed RESIDE-6K split introduced by DehazeFormer [3]:

- **SOTS-Indoor:** the canonical indoor test set on which the bulk of the published dehazing literature reports results, enabling comparison against prior work.
- **SOTS-Outdoor:** the outdoor counterpart, which probes behavior on scenes with sky regions and large depth variation.
- **RESIDE-6K test:** the mixed indoor/outdoor test partition of the RESIDE-6K split [3].

All hazy inputs are pre-processed with a gamma correction of 0.9 followed by per-channel min-max normalization, and all images are processed at a resolution of $256 \times 256 \times 3$ in the $[0,1]$ range. This preprocessing is part of the ECA-LDNet pipeline and is not shared by all literature baselines.

Protocol disclosure. A point of methodological importance must be stated up front. ECA-LDNet is trained on a *mixed* corpus that merges the RESIDE-6K training partition with ITS. This differs from the single-set protocols under which most baselines are reported: the SOTS-Indoor leaderboard figures in the literature are produced by models trained on ITS only, whereas the RESIDE-6K figures are produced by models trained on the RESIDE-6K training partition only. Consequently, the cross-method SOTS-Indoor comparison (Table 1) is *indicative rather than controlled*, because our training distribution does not match the ITS-only protocol of the baselines. The RESIDE-6K comparison in Table 2 is the closest available reference comparison because RESIDE-6K is a major component of our training corpus, but it is still not strictly matched because ECA-LDNet additionally uses ITS data and a different preprocessing pipeline. We are explicit about this distinction throughout the analysis below.

4.2 Evaluation Metrics

Following standard practice in the dehazing literature [3], [25], we report Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index (SSIM) [7]. PSNR measures pixel-level fidelity and is reported in decibels; SSIM measures structural agreement. The controlled ablations additionally report multi-scale SSIM (MS-SSIM). Higher values are better for all three metrics.

Common-output protocol. Outputs from AOD-Net [15], GCANet [35], FFA-Net [6], DehazeFormer-S [3], and ECA-LDNet were evaluated with one script on 22 paired images: six Haze-1K, two I-HAZE, ten O-HAZE, two SOTS-Indoor, and two SOTS-Outdoor images. Inputs, predictions, and references were matched by filename, resized to 256×256 using Lanczos interpolation, scaled to $[0,1]$, and evaluated in RGB using PSNR and SSIM. Every model, including ECA-LDNet, is run with the same plain clip-to- $[0,1]$ output handling under one identical evaluation pipeline; no model-specific output post-processing or contrast rescaling is applied to any method. Each model nonetheless retains its own pretrained weights and the input normalization required by its public implementation, and the models are trained on different corpora. This comparison therefore evaluates released pretrained models on a common test set

under one evaluation protocol; it is not a controlled same-training comparison, and the 22-image set is small and heterogeneous.

4.3 Implementation Details

ECA-LDNet is implemented in TensorFlow 2.16.1 / Keras 3.9.0 (Python 3.10, CUDA 11.8) and optimized with AdamW under a warmup-cosine-decay learning-rate schedule. All experiments use a fixed seed (42) with deterministic operations to improve reproducibility. Training was performed on an NVIDIA Tesla P100 16 GB GPU.

Training proceeds through a multi-stage curriculum that progressively anneals the learning rate and refines the model. Stage S1 trains for 60 epochs at a peak learning rate of $1e-3$ with 5 warmup epochs and a batch size of 16. Stage S2 continues for 30 epochs at $1e-4$, and Stage S3 for 20 epochs at $5e-5$ with the batch size reduced to 8. Stages S4 and S5 perform transfer-learning refinement initialized from S3, using an AdamW cosine schedule decaying from $1e-5$ to $1e-7$, an exponential moving average (EMA) of the weights with decay 0.999, a perceptual-heavy loss reweighting (the polish configuration), and early stopping with patience 8.

The composite training objective combines a Charbonnier data term ($\text{eps}=1e-3$, weight 0.35), a structural term $0.20 \cdot (1 - \text{SSIM})$, a Sobel-based edge term (0.10), an FFT-magnitude term (0.10), and a VGG16 perceptual term (0.25) computed on frozen ImageNet features from layers `block1_conv2`, `block2_conv2`, and `block3_conv3` [7], [8]. In the polish stages the weights are rebalanced to 0.40/0.20/0.10/0.05/0.25, raising the Charbonnier fidelity term and lowering the FFT term while keeping the perceptual weight fixed. Data augmentation comprises horizontal and vertical flips, 90-degree rotations, brightness jitter of ± 0.08 , contrast jitter of ± 0.10 , and additive Gaussian noise ($\sigma = 0.01$) applied to the hazy input only.

The final model reported in all tables is the EMA model evaluated on the test benchmarks.

4.4 Quantitative Comparison on SOTS-Indoor

Table 1 compares ECA-LDNet against representative classical, CNN-based, attention-based, and Transformer-based methods on SOTS-Indoor. Baseline PSNR/SSIM/parameter figures are quoted from the cited literature [2], [3], [6], [13], [14], [15], [16], [19]. The baselines generally use ITS-only training, whereas ECA-LDNet uses a mixed RESIDE-6K-train + ITS corpus and its own preprocessing. The table is therefore indicative rather than controlled.

On SOTS-Indoor, ECA-LDNet attains 32.53 dB PSNR and 0.9717 SSIM. It does not lead this benchmark: every listed DehazeFormer variant [3] and FFA-Net [6] exceeds it in both metrics. Relative to GridDehazeNet [19], ECA-LDNet is nominally higher in PSNR but lower in SSIM and uses more parameters. These differences cannot be attributed to architecture because the training and preprocessing protocols differ. The older DCP, DehazeNet, MSCNN, AOD-Net, and GFN entries are retained to show historical benchmark progression, not to define controlled wins.

Table 1 Indicative quantitative comparison on SOTS-Indoor. Baseline PSNR/SSIM/parameter figures are quoted from the cited literature; baselines are trained on ITS only while ECA-LDNet is trained on a mixed RESIDE-6K-train + ITS corpus, so this comparison is indicative, not controlled. *Takeaway:* ECA-LDNet exceeds the classical priors and early CNNs by a wide margin (10+ dB) but trails FFA-Net and the DehazeFormer family, which use larger capacity and a different training protocol.

Method	Year	Params	PSNR (dB) $\uparrow$	SSIM $\uparrow$
DCP [2]	2009	-	16.62	0.8546
DehazeNet [13]	2016	0.009M	19.82	0.8210
MSCNN [14]	2016	0.008M	19.84	0.8330
AOD-Net [15]	2017	0.002M	20.51	0.8160
GFN [16]	2018	0.499M	22.30	0.8800
GridDehazeNet [19]	2019	0.956M	32.16	0.9836
FFA-Net [6]	2020	4.456M	36.39	0.9886
DehazeFormer-T [3]	2023	0.686M	35.15	0.9890
DehazeFormer-S [3]	2023	1.283M	36.82	0.9920
DehazeFormer-B [3]	2023	2.514M	37.84	0.9940
DehazeFormer-L [3]	2023	25.44M	40.05	0.9960
ECA-LDNet (Ours)	2026	1.482M	32.53	0.9717

4.5 Closest Available Protocol Comparison on RESIDE-6K

Because RESIDE-6K forms the dominant component of ECA-LDNet’s training corpus, its test set provides the closest available comparison with DehazeFormer. This comparison is not strictly controlled because ECA-LDNet also uses ITS during training. Table 2 should therefore be interpreted as a reference comparison rather than an isolated architecture comparison [3].

Table 2 Closest available reference comparison on RESIDE-6K (mixed indoor and outdoor). ECA-LDNet is trained on RESIDE-6K plus ITS, whereas DehazeFormer models use the RESIDE-6K protocol. Therefore, this table provides a reference comparison rather than a strictly controlled same-training-data comparison. DehazeFormer results are taken from [3]; the variant set (T/S/B/M) follows the variants reported for the RESIDE-6K split in [3] and therefore differs from the SOTS-Indoor variant set in Table 1. *Takeaway:* on the protocol closest to its own training data, ECA-LDNet lands within roughly 0.4 dB of the comparable-scale DehazeFormer-T and -S but does not exceed any listed variant.

Method	Params	PSNR (dB) $\uparrow$	SSIM $\uparrow$
DehazeFormer-T [3]	0.686M	30.36	0.9730
DehazeFormer-S [3]	1.283M	30.62	0.9760
DehazeFormer-B [3]	2.514M	31.45	0.9800
DehazeFormer-M [3]	4.634M	30.89	0.9770
ECA-LDNet (Ours)	1.482M	30.20	0.9641

ECA-LDNet reaches 30.20 dB/0.9641 SSIM on RESIDE-6K, below every listed DehazeFormer variant. The PSNR differences to DehazeFormer-T and DehazeFormer-S are 0.16 and 0.42 dB, respectively, while the SSIM gap is larger. This is the closest dataset-level reference comparison because RESIDE-6K is included in ECA-LDNet’s training corpus, but it is not controlled: ECA-LDNet additionally uses ITS and different preprocessing. The table therefore provides scale and fidelity context rather than evidence of architectural equivalence.

4.6 Results Across All Benchmarks

Table 3 reports ECA-LDNet on all three evaluation benchmarks. Performance remains stable across the indoor, outdoor, and mixed protocols, spanning a 30.20–32.53 dB PSNR range. With the exception of the RESIDE-6K comparison in Table 2, no cross-method comparison was run on SOTS-Outdoor, so the outdoor figure characterizes our own model’s behavior across distributions rather than its standing relative to prior work on that set. Moreover, none of our training data is drawn from the RESIDE Outdoor Training Set (OTS); the SOTS-Outdoor result therefore reflects the cross-distribution generalization of an indoor/mixed-trained model and is reported descriptively, not as a benchmarked comparison.

Table 3 ECA-LDNet results across all evaluation benchmarks (final EMA model). *Takeaway:* performance is stable across the indoor, outdoor, and mixed protocols, spanning a narrow 30.20–32.53 dB PSNR range.

Benchmark	PSNR (dB) $\uparrow$	SSIM $\uparrow$
SOTS-Indoor	32.53	0.9717
SOTS-Outdoor	31.94	0.9769
RESIDE-6K (test)	30.20	0.9641

The model attains its highest SSIM on SOTS-Outdoor (0.9769), a descriptive single-model observation suggesting strong structural preservation on outdoor scenes despite the additional challenge of sky regions and large depth ranges that historically destabilize prior-based methods [2], [11]. The mixed RESIDE-6K test partition yields the most conservative numbers (30.20 dB / 0.9641), as expected for a heterogeneous indoor/outdoor distribution.

4.7 Controlled Component Ablation

Table 4 provides the principal controlled architectural evidence. Every variant uses the same training pool, split, preprocessing, seed, optimizer configuration, epoch-28 checkpoint rule, and evaluation code. These checkpoints are separate from the final multi-stage EMA model; only differences among ablation variants are interpreted.

Removing ECA lowers PSNR by 0.848 dB on SOTS-Indoor, 1.684 dB on SOTS-Outdoor, and 0.687 dB on RESIDE-6K while removing only 27 parameters. MS-SSIM also falls on all three datasets. The ordinary SSIM change on SOTS-Indoor is +0.0008,

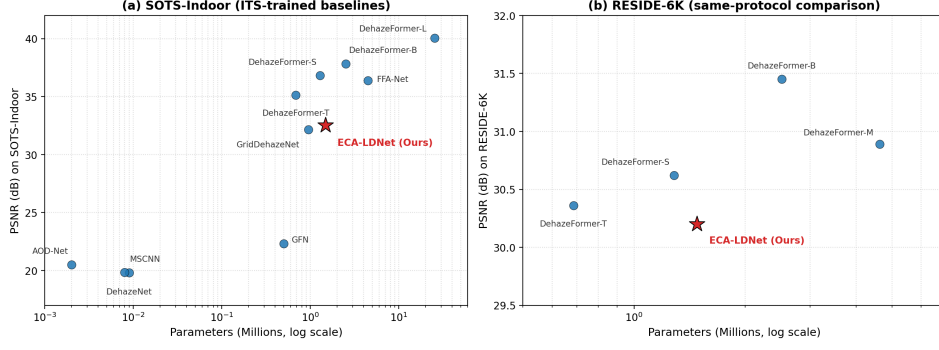


Fig. 2 Published accuracy versus parameter count (log scale) on (a) SOTS-Indoor and (b) RESIDE-6K. ECA-LDNet is marked with a star. *Takeaway:* ECA-LDNet occupies a compact, low-parameter region of the plot while remaining within a few dB of much larger models. The plot visualizes reported operating points only; it should not be read as a Pareto frontier or a controlled comparison because training data, preprocessing, and evaluation implementations differ.

Table 4 Controlled single-seed ablation results. Each cell reports PSNR/SSIM/MS-SSIM. Variants use a shortened single-seed epoch-28 protocol; the “Full model” row (29.65 dB on SOTS-Indoor) is therefore not the final multi-stage EMA model reported in Tables 1 and 3 (32.53 dB), and only differences among ablation variants are interpreted. *Takeaway:* removing ECA causes the largest PSNR drop on two of three datasets at only 27 fewer parameters, while removing pixel attention or atmospheric-model guidance lowers PSNR but slightly raises SSIM/MS-SSIM.

Variant	Params	SOTS-Indoor	SOTS-Outdoor	RESIDE-6K
Full model	1,481,871	29.653/0.9467/0.9759	29.118/0.9622/0.9839	28.953/0.9531/0.9797
Without ECA	1,481,844	28.805/0.9475/0.9746	27.434/0.9585/0.9825	28.266/0.9520/0.9785
Without pixel attention	1,372,438	29.088/0.9564/0.9786	28.662/0.9646/0.9860	28.930/0.9590/0.9820
Without atmospheric-model guidance	1,430,283	28.763/0.9492/0.9761	28.609/0.9637/0.9855	28.774/0.9551/0.9806
Without all attention	1,372,411	28.780/0.9472/0.9751	27.794/0.9597/0.9827	28.369/0.9520/0.9787

but SSIM falls on the other datasets. Within this run, ECA therefore provides the largest and most consistent PSNR contribution per added parameter.

Removing atmospheric-model guidance lowers PSNR by 0.890, 0.509, and 0.179 dB on SOTS-Indoor, SOTS-Outdoor, and RESIDE-6K, respectively. In contrast, SSIM and MS-SSIM increase slightly after removal. The branch therefore favors pixel-level fidelity as measured by PSNR while introducing a small structural-metric trade-off. This establishes that the branch affects restoration behavior, but not that its latent variables are physically calibrated.

Pixel attention produces the clearest metric trade-off. Removing it lowers PSNR by 0.565 dB on SOTS-Indoor and 0.456 dB on SOTS-Outdoor, while RESIDE-6K changes by only 0.023 dB. SSIM and MS-SSIM increase on every dataset after removal. Pixel attention is therefore retained because it favors PSNR in this run, not because it uniformly improves structural fidelity. All ablations are single-seed point estimates and do not establish statistical significance.

Table 5 Component contribution and parameter-cost summary. Metric deltas are computed as ablated variant minus full model. *Takeaway*: ECA is effectively free (27 parameters) yet yields the largest PSNR-per-parameter contribution, whereas pixel attention and atmospheric-model guidance cost more parameters and trade PSNR against the structural metrics.

Removed component	Parameter reduction	Share	Indoor Δ PSNR/ Δ SSIM/ Δ MS-SSIM	Outdoor Δ PSNR/ Δ SSIM/ Δ MS-SSIM	RESIDE-6K Δ PSNR/ Δ SSIM/ Δ MS-SSIM
ECA	27	0.002%	-0.848/ + 0.0008/ - 0.0013	-1.684/ - 0.0037/ - 0.0014	-0.687/ - 0.0011/ - 0.0011
Pixel attention	109,433	7.385%	-0.565/ + 0.0098/ + 0.0027	-0.456/ + 0.0023/ + 0.0021	-0.023/ + 0.0059/ + 0.0023
Atmospheric-model guidance	51,588	3.481%	-0.890/ + 0.0025/ + 0.0002	-0.509/ + 0.0015/ + 0.0017	-0.179/ + 0.0020/ + 0.0009
All attention	109,460	7.387%	-0.873/ + 0.0005/ - 0.0008	-1.325/ - 0.0025/ - 0.0012	-0.584/ - 0.0011/ - 0.0010

4.8 Common-Image Comparison of Pretrained Models

To complement the published reference tables, we evaluated the outputs of AOD-Net, GCANet, FFA-Net, DehazeFormer-S, and ECA-LDNet on the same 22 paired images under one identical evaluation pipeline. This controls the evaluated files, resolution, output handling, and metric implementation: all five models, including ECA-LDNet, use the same plain clip-to-[0,1] output with no model-specific rescaling. It does not control training data or the input normalization each public implementation requires, and the 22-image set is small and heterogeneous. Table 6 and Fig. 3 therefore report a fair common-test-set comparison of released pretrained models rather than a controlled same-training architecture ranking.

Table 6 Common-image evaluation at 256×256 on 22 paired images. All models, including ECA-LDNet, are evaluated under one identical pipeline (Lanczos resize, scale to [0,1], RGB PSNR/SSIM) with no model-specific output rescaling. Each cell reports PSNR/SSIM; the best value per column is in bold. The set is small and heterogeneous and the models differ in training data and required input normalization, so values are case-level evidence rather than population-level rankings. *Takeaway*: under this shared evaluation ECA-LDNet attains the highest overall PSNR (marginally above GCANet) and competitive SSIM, leading both metrics on the outdoor subsets.

Model	Haze-1K (6)	I-HAZE (2)	O-HAZE (10)	SOTS-In (2)	SOTS-Out (2)	Overall (22)
AOD-Net	19.68/0.9151	15.02/0.7003	16.20/0.6686	13.50/0.7430	21.47/0.8927	17.27/0.7658
GCANet	20.31 /0.9201	16.22 / 0.7838	18.25/0.7530	25.19/0.9139	25.25/0.9241	19.90/ 0.8316
FFA-Net	16.60/0.8567	10.88/0.4561	13.86/0.5890	24.42/0.9339	17.72/0.7807	15.65/0.6987
DehazeFormer-S	19.52/ 0.9280	12.54/0.5652	15.00/0.6516	28.62 / 0.9609	17.57/0.7511	17.48/0.7563
ECA-LDNet	17.36/0.8797	14.96/0.7452	19.21 / 0.7569	28.58/0.9486	27.56 / 0.9705	19.93 /0.8262

Because this subset contains only 22 heterogeneous images, including two images each from I-HAZE, SOTS-Indoor, and SOTS-Outdoor, the results provide case-level evidence and are not used to infer dataset-level model rankings.

Across this small set, ECA-LDNet attains the highest overall mean PSNR (19.93 dB), marginally above GCANet (19.90 dB), while GCANet attains the highest overall mean SSIM (0.8316) with ECA-LDNet second (0.8262). ECA-LDNet is strongest on the outdoor scenes, leading both metrics on SOTS-Outdoor (27.56 dB / 0.9705) and on O-HAZE (19.21 dB / 0.7569), and is close to the best on SOTS-Indoor. Its weakest subset is Haze-1K, where its PSNR and SSIM trail several baselines, consistent with a residual sensitivity to that domain’s color and density characteristics visible in the montage. Given the small, heterogeneous subset, these values remain case-level evidence rather than population-level rankings.

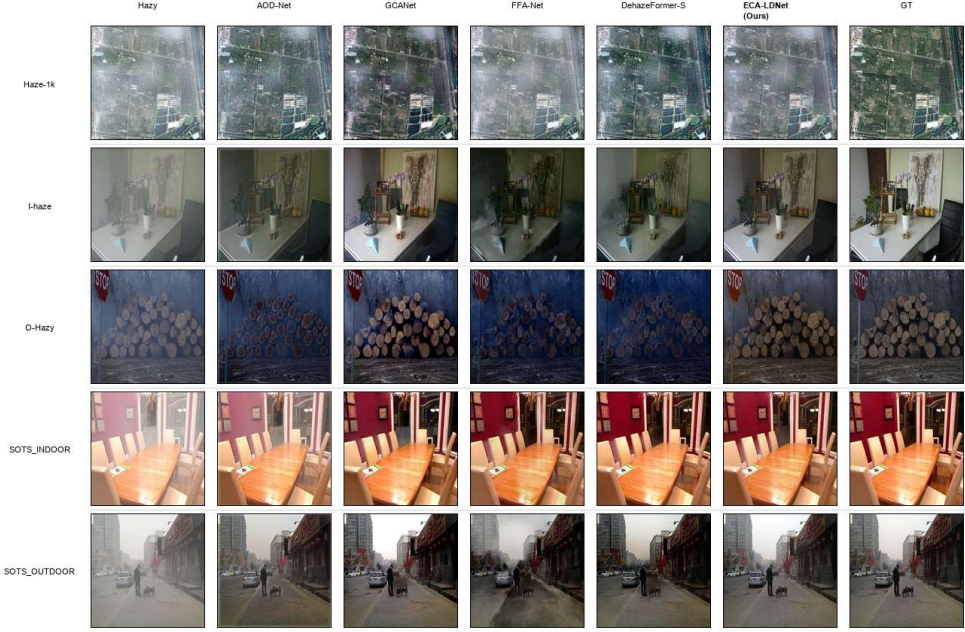


Fig. 3 Common-image comparison, one representative scene per dataset (top to bottom: Haze-1K, I-HAZE, O-HAZE, SOTS-Indoor, SOTS-Outdoor). Columns show the hazy input, AOD-Net, GCANet, FFA-Net, DehazeFormer-S, ECA-LDNet, and the ground truth. All methods are run with their public pretrained weights and the input normalization their implementations require, and are evaluated under one identical pipeline; no model-specific output rescaling is applied to any method. *Takeaway:* ECA-LDNet restores strong contrast on the outdoor scenes (O-HAZE, SOTS-Outdoor) and is weakest on the Haze-1K example, consistent with Table 6.

4.9 Efficiency Analysis

The verified efficiency evidence consists of the exact 1,481,871-parameter model count and device-specific latency. On an NVIDIA Tesla P100 using TensorFlow 2.16.1, FP32, batch size 1, and 256×256 inputs, ECA-LDNet requires approximately 72–82 ms per image after a 10-image warm-up, corresponding to 12–14 frames/s over 500 images. This measurement is framework- and hardware-specific and should not be interpreted as edge-device latency or as a universal real-time guarantee. Fig. 2 situates ECA-LDNet’s reported operating point against the published accuracy–parameter trade-offs of prior methods on SOTS-Indoor and RESIDE-6K; because training data, preprocessing, and evaluation implementations differ across these methods, the plot is a contextual visualization rather than a controlled Pareto frontier.

We additionally report an analytical multiply–accumulate (MAC) count computed by a released layer-wise counter that sums the standard per-layer cost of every convolutional and dense layer in the built model. At 256×256 the model requires 2.171 GMACs (4.343 GFLOPs); 93.6% of this cost lies in the pointwise and standard convolutions and 6.4% in the depthwise convolutions. The same counter reproduces the exact 1,481,871-parameter total, confirming that the model is analyzed correctly. We

report this as a hardware-independent structural measure of the depthwise-separable design, and we deliberately do not convert it into cross-paper MAC ratios or “lowest compute” claims, because published baseline MAC values are produced by different tools under different counting conventions. Table 5 separately quantifies the exact parameter contribution of each proposed module.

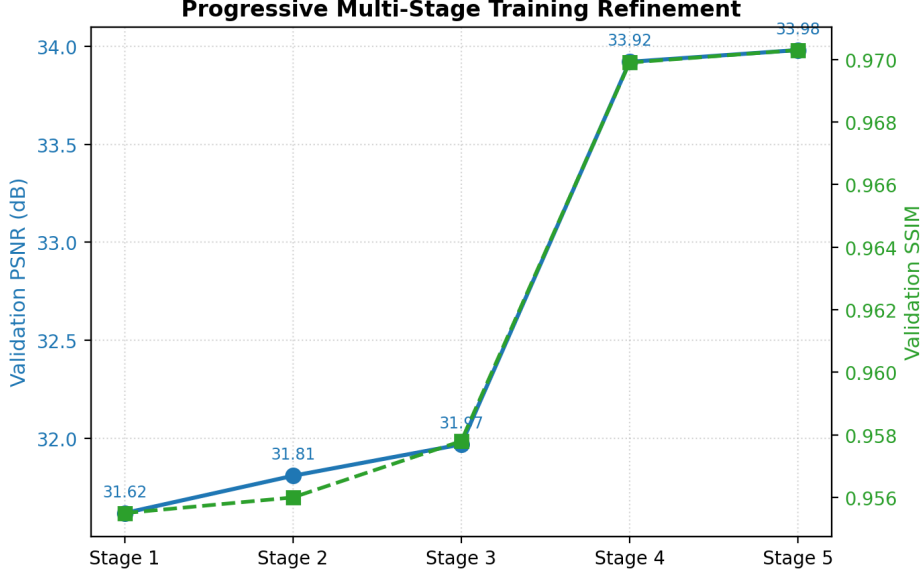


Fig. 4 Progressive multi-stage training. Validation PSNR/SSIM improve across stages, with the largest gain at the S3→S4 transition that introduces EMA and perceptual-heavy refinement. *Take-away:* most of the quality gain comes from the late EMA and perceptual-heavy refinement (S4–S5) rather than the early high-learning-rate stages. Reported as a training-strategy study, not a component ablation.

4.10 Training-Strategy Study: Multi-Stage Progression

To characterize the contribution of our multi-stage training curriculum, we record validation performance at the conclusion of each stage in Table 7, and Fig. 4 plots the multi-stage validation progression. We emphasize that this is a training-strategy study on the mixed validation split, **not** a controlled component ablation: it isolates the effect of the staged learning-rate annealing and transfer-learning refinement rather than the effect of any individual architectural component.

The progression shows two regimes. The initial three stages (S1→S3) yield gradual, monotonic gains (31.62 → 31.97 dB) as the learning rate is annealed under the primary loss configuration. The largest single improvement occurs at the S3→S4 transition (+1.95 dB), where fine-tuning with EMA weight averaging (decay 0.999) and the perceptual-heavy polish loss is introduced. The final stage (S5) provides a small additional gain to 33.98 dB / 0.9703, indicating convergence. This pattern is consistent with a dedicated late refinement phase being valuable: most of the quality gain

Table 7 Multi-stage training progression on the mixed validation split. Reported as a training-strategy study, not a component ablation. *Takeaway:* most of the improvement appears in the late EMA and perceptual-heavy refinement stages (S4–S5), with the single largest jump at S3→S4.

Stage	Description	Val PSNR (dB) $\uparrow$	Val SSIM $\uparrow$
S1	60 ep, LR 1e-3, warmup 5	31.62	0.9555
S2	30 ep, LR 1e-4	31.81	0.9560
S3	20 ep, LR 5e-5, batch 8	31.97	0.9578
S4	Fine-tuning, EMA, perceptual-heavy	33.92	0.9699
S5	Fine-tuning continued	33.98	0.9703

is recovered in the late stages rather than the early high-learning-rate stages. We caution, however, that the S3→S4 step changes several factors at once (EMA weight averaging, the perceptual-heavy loss reweighting, and the learning-rate schedule), so the +1.95 dB gain cannot be attributed to any single one of them; isolating their individual effects would require controlled experiments.

We stress that the values in Table 7 are measured on the mixed validation split, which follows a different distribution from the SOTS-Indoor test set used in Table 1. That the final-stage validation PSNR (33.98 dB) exceeds the reported SOTS-Indoor test PSNR (32.53 dB) therefore reflects this distribution difference and is not an internal contradiction.

4.11 Qualitative Results

The representative outputs show that ECA-LDNet can recover contrast and local texture in indoor and outdoor scenes. These images are illustrative examples rather than evidence that any one architectural component caused the observed behavior.

The atmospheric-model estimate enters through the 8% blend while the learned CNN branch governs most of the output. The predicted transmission-like and atmospheric-light-like variables are accessible as intermediate outputs, but they were not evaluated against physical ground truth. We therefore do not infer physical accuracy or interpretability from their visual appearance.

Fig. 5 shows representative single-model outputs. The broader cross-model montage in Fig. 3 includes both favorable examples and visible color/luminance failures.

Limitation on synthetic evaluation. All quantitative results above use synthetically hazed RESIDE images, which assume a homogeneous single-scattering model. Real-world haze exhibits spatially varying density, illumination, and multiple-scattering effects not captured by the ASM training distribution. Evaluation on real hazy images using no-reference perceptual metrics (NIQE, FAD, BRISQUE) is identified as a necessary next step; the frozen model can be applied to any real hazy image and these metrics computed without retraining. We acknowledge this as a significant open gap between demonstrated and deployed performance.

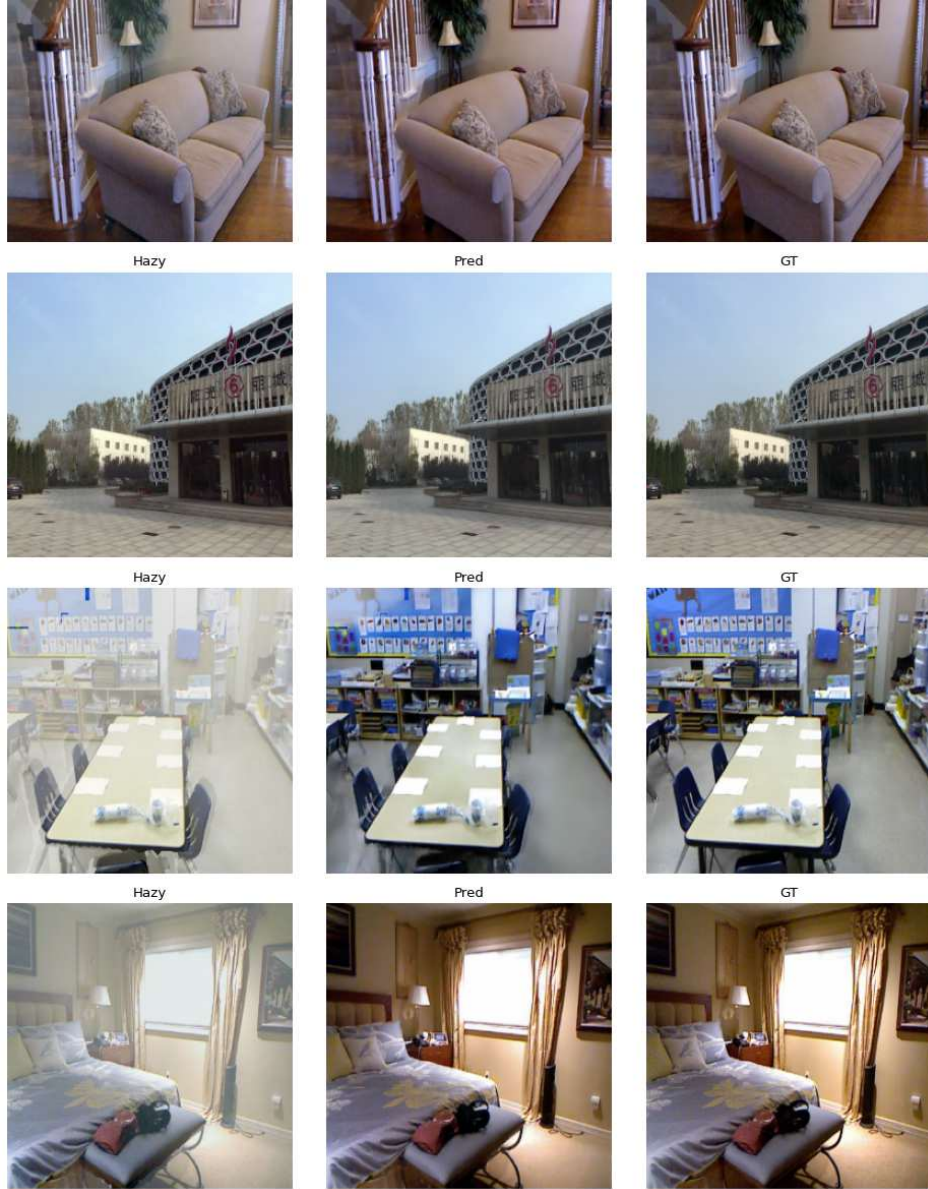


Fig. 5 Representative ECA-LDNet outputs on indoor and outdoor validation scenes. Each row shows the hazy input, restoration, and clean reference. *Takeaway:* ECA-LDNet recovers contrast and texture while staying close to the clean reference. These examples illustrate restoration behavior and are not a cross-method comparison.

5 Discussion and Limitations

The strongest evidence in this study is the controlled component analysis rather than the cross-paper leaderboard comparison. ECA contributes substantial PSNR gains at negligible parameter cost. Atmospheric-model guidance and pixel attention also affect restoration, but their effects depend on the metric: both favor PSNR while their removal slightly improves SSIM and MS-SSIM in several settings. These results argue for reporting component trade-offs rather than treating every module as uniformly beneficial.

5.1 Where ECA-LDNet Succeeds

Compact architecture with measured component costs. ECA-LDNet contains 1.482M parameters. ECA accounts for only 27 parameters, pixel attention for 109,433 parameters, and atmospheric-model guidance for 51,588 parameters. The device-specific latency is approximately 72–82 ms per 256×256 image on an NVIDIA Tesla P100, and a released layer-wise counter reports 2.171 GMACs at this resolution. We treat the MAC count as a hardware-independent structural measure and do not convert it into cross-paper ratios or a lowest-compute claim, because baseline MAC values follow different counting conventions.

Protocol-bounded external positioning. ECA-LDNet obtains 30.20 dB/0.9641 SSIM on RESIDE-6K. The listed DehazeFormer variants remain ahead on both metrics, including variants with similar or fewer parameters. The comparison is not controlled because ECA-LDNet additionally uses ITS and different preprocessing. The common-image analysis controls the evaluated files, output handling, and metric code, and applies no model-specific output rescaling, but it still does not control training data or the input normalization each public model requires.

5.2 Where ECA-LDNet Falls Short

ECA-LDNet does not lead the reported benchmarks. On SOTS-Indoor, FFA-Net and all listed DehazeFormer variants exceed its PSNR and SSIM; DehazeFormer-T also uses fewer parameters. On the common-image set it is weakest on the Haze-1K subset, where its PSNR and SSIM trail several baselines, indicating sensitivity to that domain’s color and density characteristics. The model should therefore be positioned as a compact architecture with characterized internal trade-offs, not as a new accuracy or efficiency frontier.

5.3 Limitations

Mixed, non-standard training protocol. ECA-LDNet is trained on RESIDE-6K plus ITS and uses gamma correction with per-channel min–max normalization. The published baselines use different training and preprocessing protocols, preventing causal architecture comparisons.

Single-run point estimates. Training and ablation results use one fixed seed. They establish directional effects under the tested run but do not estimate variance or statistical significance.

Shortened ablation protocol. The ablation checkpoints use a shortened training schedule and are not directly comparable with the final EMA checkpoint. Consequently, the ablations establish relative component effects only under their shared shortened protocol.

Training on synthetic haze. ECA-LDNet is trained and evaluated entirely on synthetically hazed data, with testing on SOTS-Indoor, SOTS-Outdoor, and the RESIDE-6K test split [25]. The Atmospheric Scattering Model used to synthesize these benchmarks assumes homogeneous, single-scattering haze, whereas real-world haze, fog, and smoke exhibit spatially varying density, non-uniform illumination, and multiple-scattering effects. Our reported metrics therefore characterize performance under idealized scattering conditions and should not be read as a direct measure of real-world generalization. Bridging this synthetic-to-real gap is an active research direction in its own right (for example, RIDCP [32] revitalizes real-image dehazing through high-quality codebook priors), and evaluating ECA-LDNet on, or adapting it to, real hazy imagery is an important avenue we leave to future work.

Structural-fidelity gap. ECA-LDNet’s SSIM is below the listed high-fidelity baselines. The ablations further show that pixel attention and atmospheric-model guidance can improve PSNR while reducing structural metrics.

Auxiliary blend weight not sensitivity-tested. The 8% ASM-derived blend ($\hat{J} = 0.92 \cdot \hat{J}_{\text{raw}} + 0.08 \cdot J_{\text{asm}}$) was chosen empirically and not validated through a systematic sensitivity study. Whether lower or higher blend weights improve accuracy remains unknown. The transmission-like and atmospheric-light-like latent variables have not been independently validated against physical ground truth.

Fixed operating resolution. Training, evaluation, latency, and the reported MAC count all use 256×256 inputs. The analytical MAC value (2.171 GMACs) is specific to this resolution and is computed by the released layer-wise counter; full-resolution behavior and its compute scaling are not characterized here.

Small common-image set. The common-image comparison uses one identical evaluation pipeline for all five models with no model-specific output rescaling, but it covers only 22 heterogeneous images and does not control training data or each model’s required input normalization. It therefore provides case-level evidence rather than a population-level ranking.

5.4 Future Work

Several concrete directions follow directly from the limitations above.

The immediate priorities are multi-seed validation to attach variance to the ablation, evaluation on larger real-image sets with no-reference metrics, and a sensitivity analysis of the 8% blend weight. A unified retraining study, in which the compared architectures share one training corpus and preprocessing pipeline, would be required before making causal accuracy–efficiency comparisons with external architectures.

6 Conclusion

This paper presented ECA-LDNet, a 1.482M-parameter single-image dehazing CNN combining a depthwise-separable U-Net, ECA-then-pixel-attention residual blocks,

and a softly blended atmospheric-model branch. The model achieves 32.53 dB/0.9717 SSIM on SOTS-Indoor, 31.94 dB/0.9769 on SOTS-Outdoor, and 30.20 dB/0.9641 on RESIDE-6K. Controlled ablations show that ECA provides the largest and most consistent PSNR contribution per parameter, while pixel attention and atmospheric-model guidance introduce PSNR–structural-similarity trade-offs. Published cross-method results remain protocol-unmatched, and a fair common-image comparison under one identical evaluation pipeline gives ECA-LDNet the highest overall PSNR on a small 22-image set without a controlled same-training comparison. ECA-LDNet is therefore best understood as a compact architecture with transparent component behavior, not as a state-of-the-art fidelity or compute result.

Declarations

Ethics approval.

NA.

Consent to participate.

NA.

Consent to publish.

NA.

Data availability.

The datasets used during the current study (RESIDE, SOTS, Haze-1K, I-HAZE, O-HAZE) are publicly available in standard research repositories. Code availability: Code, model weights, training configurations, and evaluation scripts are available at <https://github.com/manpreet28111995/Image-Dehazing-using-ECA-LDNet>.

Funding.

The authors received no specific funding for this work.

Competing interests.

The authors declare no competing interests.

Use of AI-assisted technologies.

LLMs were only used for copyediting the work.

Author Contributions.

M.S.: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing (original draft, review & editing). Rohith Reddy Bellibatlu: Methodology, Software, Investigation, Data curation, Visualization, Writing (original draft, review & editing). Sai Deekshith Lekkala: Software, Validation, Investigation, Visualization, Writing (original draft, review & editing). Rahul Joshi: Conceptualization, Resources, Supervision, Project administration, Writing (original draft, review & editing).

Co-first authorship.

Manpreet Singh, Rohith Reddy Bellibatlu, and Sai Deekshith Lekkala contributed equally to this work as co-first authors. Order was determined alphabetically.

References

- [1] S. G. Narasimhan and S. K. Nayar, “Vision and the Atmosphere,” *International Journal of Computer Vision (IJCV)*, vol. 48, no. 3, pp. 233-254, 2002.
- [2] K. He, J. Sun, and X. Tang, “Single Image Haze Removal Using Dark Channel Prior,” *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 33, no. 12, pp. 2341-2353, 2011.
- [3] Y. Song, Z. He, H. Qian, and X. Du, “Vision Transformers for Single Image Dehazing (DehazeFormer),” *IEEE Transactions on Image Processing (TIP)*, vol. 32, 2023.
- [4] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” *arXiv preprint arXiv:1704.04861*, 2017.
- [5] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, “ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020.
- [6] 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.*, 2020.
- [7] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image Quality Assessment: From Error Visibility to Structural Similarity,” *IEEE Transactions on Image Processing (TIP)*, vol. 13, no. 4, pp. 600-612, 2004.
- [8] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual Losses for Real-Time Style Transfer and Super-Resolution,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016.
- [9] R. Fattal, “Single Image Dehazing,” *ACM Transactions on Graphics (Proc. SIGGRAPH)*, vol. 27, no. 3, pp. 1-9, 2008.
- [10] Q. Zhu, J. Mai, and L. Shao, “A Fast Single Image Haze Removal Algorithm Using Color Attenuation Prior,” *IEEE Transactions on Image Processing (TIP)*, vol. 24, no. 11, pp. 3522-3533, 2015.
- [11] D. Berman, T. Treibitz, and S. Avidan, “Non-Local Image Dehazing,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 1674-1682, 2016.
- [12] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, “Efficient Image Dehazing with Boundary Constraint and Contextual Regularization,” in *Proc. IEEE Int.*

- Conf. Comput. Vis. (ICCV), pp. 617-624, 2013.
- [13] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, “DehazeNet: An End-to-End System for Single Image Haze Removal,” *IEEE Transactions on Image Processing (TIP)*, vol. 25, no. 11, pp. 5187-5198, 2016.
 - [14] 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. Eur. Conf. Comput. Vis. (ECCV)*, pp. 154-169, 2016.
 - [15] 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.
 - [16] 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/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 3253-3261, 2018.
 - [17] J. Hu, L. Shen, and G. Sun, “Squeeze-and-Excitation Networks,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018.
 - [18] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “CBAM: Convolutional Block Attention Module,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018.
 - [19] X. Liu, Y. Ma, Z. Shi, and J. Chen, “GridDehazeNet: Attention-Based Multi-Scale Network for Image Dehazing,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2019.
 - [20] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, “Restormer: Efficient Transformer for High-Resolution Image Restoration,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 5728-5739, 2022.
 - [21] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li, “Uformer: A General U-Shaped Transformer for Image Restoration,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 17683-17693, 2022.
 - [22] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, “SwinIR: Image Restoration Using Swin Transformer,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, pp. 1833-1844, 2021.
 - [23] Y. Qiu, K. Zhang, C. Wang, W. Luo, H. Li, and Z. Jin, “MB-TaylorFormer: Multi-Branch Efficient Transformer Expanded by Taylor Formula for Image Dehazing,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, pp. 12802-12813, 2023.
 - [24] L.-A. Tran and D.-C. Park, “Distilled Pooling Transformer Encoder for Efficient Realistic Image Dehazing,” *arXiv:2412.14220*, 2024.

- [25] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, “Benchmarking Single-Image Dehazing and Beyond (RESIDE),” *IEEE Transactions on Image Processing (TIP)*, vol. 28, no. 1, pp. 492-505, 2019.
- [26] H. Zhang and V. M. Patel, “Densely Connected Pyramid Dehazing Network,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 3194-3203, 2018.
- [27] Y. Qu, Y. Chen, J. Huang, and Y. Xie, “Enhanced Pix2pix Dehazing Network,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 8160-8168, 2019.
- [28] Z. Chen, Z. He, and Z.-M. Lu, “DEA-Net: Single Image Dehazing Based on Detail-Enhanced Convolution and Content-Guided Attention,” *IEEE Transactions on Image Processing (TIP)*, vol. 33, pp. 1002-1015, 2024.
- [29] Y. Song, Y. Zhou, H. Qian, and X. Du, “Rethinking Performance Gains in Image Dehazing Networks,” *arXiv:2209.11448*, 2022.
- [30] 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)*, pp. 5785-5794, 2023.
- [31] Y. Cui, W. Ren, X. Cao, and A. Knoll, “Revitalizing Convolutional Network for Image Restoration,” *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 46, no. 12, pp. 9423-9438, 2024.
- [32] 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)*, pp. 22282-22291, 2023.
- [33] H. Wu, Y. Qu, S. Lin, J. Zhou, R. Qiao, Z. Zhang, Y. Xie, and L. Ma, “Contrastive Learning for Compact Single Image Dehazing (AECD-Net),” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 10551-10560, 2021.
- [34] C.-L. Guo, Q. Yan, S. Anwar, R. Cong, W. Ren, and C. Li, “Image Dehazing Transformer with Transmission-Aware 3D Position Embedding (Dehamer),” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 5812-5820, 2022.
- [35] D. Chen, M. He, Q. Fan, J. Liao, L. Zhang, D. Hou, L. Yuan, and G. Hua, “Gated Context Aggregation Network for Image Dehazing and Deraining,” in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, pp. 1375-1383, 2019.