SLRNet: Super Lightweight Residual Network for Real-Time Image Dehazing

Guanheng Qu, Fan Jiang and Jiangming Liu✉

School of Information Science and Engineering, Yunnan University, Kunming, China
jiangmingliu@ynu.edu.cn

Abstract. Image dehazing aims to generate the haze-free images from the hazy observation images. While recent deep learning approaches achieve impressive restoration quality, they suffer from excessive computational complexity and model size, hindering practical applications for real-world deployment on resource-constrained edge devices. To address the limitation, lightweight models are proposed to this end but compromise on dehazing performance. To bridge this gap, we propose SLRNet, Super Lightweight Residual Network, a high efficient-yet-effective end-to-end dehazing architecture. SLRNet integrates a novel Adaptive Feature Unit that automatically adjusts channel-wise features through a lightweight gating mechanism, coupled with compact residual blocks to preserve critical structural information. Unlike standard channel attention mechanisms that discard spatial information, our AFU employs an asymmetric split strategy to simultaneously preserve local texture details and capture global haze density. Our design emphasizes minimal parameter count and low latency without sacrificing perceptual quality. Experiments are carried out across standard benchmarks, showing that our proposed SLRNet demonstrates remarkable performance by achieving state-of-the-art efficiency-accuracy trade-offs compared to existing works, while maintaining robust generalization to real-world haze despite the synthetic-to-real domain gap. The codes are released in https://github.com/Unk1ndledAC/SLRNet.

Keywords: Image dehazing, Image restoration, Lightweight network, Real-time inference, Residual learning, Adaptive channel attention.

1 Introduction

Haze degradation caused by atmospheric scattering and absorption severely compromises visibility, color accuracy, and the performance of downstream vision tasks such as autonomous driving and surveillance. Image dehazing seeks to restore a clear scene from a hazy input. Early approaches relies on hand-crafted priors [5, 25], which offer interpretability but fail under complex or non-homogeneous haze conditions and exhibit limited adaptability. The rise of deep learning shifted the paradigm toward data-driven models that learn direct mappings from hazy to clean images. Recent methods have

✉ Corresponding Author

achieved impressive restoration quality by leveraging attention mechanisms [10, 13], contrastive regularization [19, 21], depth guidance [23], or transformer architectures [14], but they suffer from computational costs or depend heavily on synthetic training data, limiting their real-world applicability.

Efforts toward efficiency have yielded lightweight networks to alleviate the limitation [3, 8, 18], but at the expense of visual accuracy or generalization. So recent works attempt to bridge the synthetic-to-real gap through domain adaptation [12] or iterative decoding [4]. These methods face a serious problem of *high computational demand*.

To address the problem, we propose SLRNet, a Super Lightweight Residual Network that achieves both high dehazing quality and real-time efficiency without relying on complex modules or synthetic-to-real adaptation pipelines. It is built upon optimized residual blocks and enhanced with a novel *Adaptive Feature Unit* (AFU). AFU enables efficient, context-aware channel-wise features adjustment by selectively refining a subset of channels via depth-wise convolution and a global-context-driven sigmoid gate, facilitating effective fusion of shallow detail and deep semantic information.

We carry out the experiments across standard benchmarks, SLRNet achieves state-of-the-art performance among lightweight models by obtaining 28.78 dB PSNR and 0.9645 SSIM on SOTS-outdoor [9] at only 1.44 ms per image. Additionally, our model that is trained only on the coarsely-synthetic dataset RESIDE-6K [9] demonstrates the strong generalization to heterogenous settings referred to NH-HAZE [1]. This highlights the effectiveness and robustness of our compact architecture and AFU-based feature refinement.

2 Related Work

2.1 Prior-Based Dehazing

Classical dehazing methods rely on hand-crafted priors derived from observations of haze-free images. The Dark Channel Prior (DCP) [5] assumes that local patches in haze-free images contain some pixels with very low intensity in at least one color channel. Despite its elegance, DCP is computationally intensive and fails in regions violating the prior. The Color Attenuation Prior (CAP) [25] establishes a linear relationship between scene depth and the difference between brightness and saturation, offering faster processing but limited accuracy under complex lighting or non-uniform haze.

2.2 Learning-Based Dehazing

Early deep models like DehazeNet [2] used shallow CNNs to estimate transmission maps, while MSCNN [15] employed multi-scale inputs for better context. With the help of deep learning, AOD-Net [8] reformulated dehazing as a unified estimation problem, enabling a very compact network but yielding suboptimal results. To enhance feature representation, attention mechanisms have been widely adopted. FFA-Net [13] combines channel and spatial attention across multi-scale features. GridDehazeNet [10] uses a grid network with channel attention for multi-scale processing, improving

efficiency over FFA-Net. RIDCP [22] leverages codebook priors for real-image enhancement. However, these works have high computational complexity.

Recently, MSBDN [3] introduces a boosted multi-scale decoder but has similar latency constraints. UCL-Dehaze [18] employs unsupervised contrastive learning for domain adaptation. AECR-Net [21] uses contrastive regularization to help the dehazing model preserve information features adaptively. However, these methods often employ deep or wide architectures, making them unsuitable for edge devices deployment.

2.3 Residual Learning and Activation Functions

Residual connections, mitigate vanishing gradients and ease training of deep networks [6]. They have been successfully adapted to dehazing [24]. Motivated by this, our work leverages the efficiency of compact residual blocks to achieve a strong speed-accuracy trade-off.

3 Methodology

In this section, we present the architecture of SLRNet, a super lightweight yet high-performance network for single image dehazing. SLRNet is attributed to *minimal computational overhead* through efficient operators and compact blocks, and *effective feature refinement* via adaptive channel recalibration.

3.1 Atmospheric Scattering Model

The formation of a hazy image $I(x)$ is commonly described by the physical model [5]:

$$I(x) = J(x)t(x) + A(1 - t(x)), \tag{1}$$

where $J(x)$ is a haze-free image, A is the global atmospheric light, and $t(x)$ is the transmission map indicating the portion of light that reaches the camera without scattering. The goal of dehazing is to estimate $J(x)$ from $I(x)$. Traditional methods often estimate A and $t(x)$ separately, but end-to-end deep learning models like ours directly learn the mapping $I \mapsto J$.

3.2 Network Architecture Overview

As shown in the top of Figure 1, the pipeline of SLRNet proceeds as follows:

- Shallow Feature Extraction: A single 3 × 3 convolution projects the input image into a 32-channel feature space.
- Adaptive Feature Refinement: The feature map is first enhanced by an adaptive feature unit, then fed into a compact backbone composed of optimized residual blocks, followed by a second adaptive feature unit for late-stage recalibration.

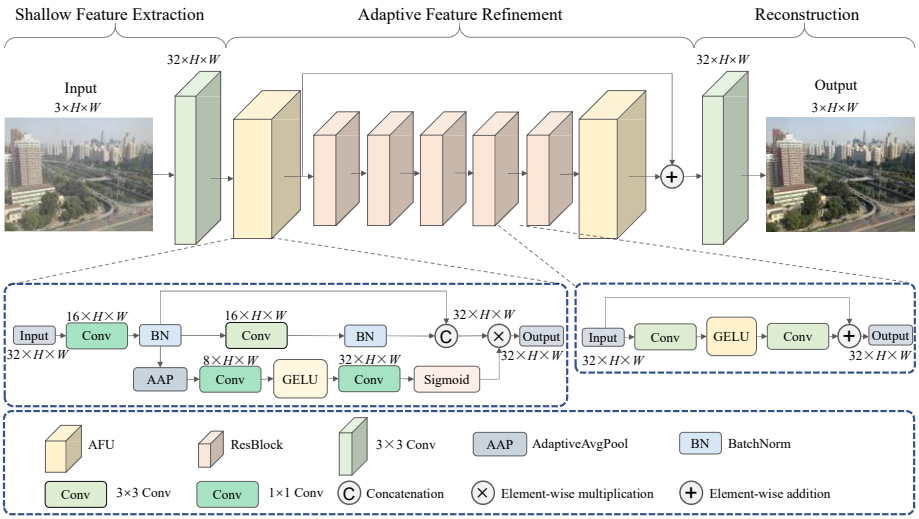

Fig. 1. Overall architecture of SLRNet.

- Skip Connection & Reconstruction: The refined 32-channel features are element-wise added to the output of the first adaptive feature unit, and a final 3 × 3 convolution compresses the result back to 3-channel feature space.

3.3 Adaptive Feature Unit

Adaptive Feature Unit (AFU) is a lightweight module that performs dynamic channel-wise feature recalibration. Given an input feature tensor $\boldsymbol{X} \in \mathbb{R}^{C\times H\times W}$, AFU operates as follows:

Channel Splitting We split $\boldsymbol{X}$ into two parts along the channel dimension:

$$\boldsymbol{X}_1, \boldsymbol{X}_2 = \text{Split}(\boldsymbol{X}),$$
$$\boldsymbol{X}_1 \in \mathbb{R}^{C_1\times H\times W}, \boldsymbol{X}_2 \in \mathbb{R}^{C_2\times H\times W}. \tag{2}$$

where $C_1 = \alpha C$ and $C_2 = (1-\alpha)C$ with $\alpha = 0.5$ by default. Only $\boldsymbol{X}_1$ undergoes further transformation.

Context-Aware Gating To generate a channel attention map $\boldsymbol{M} \in \mathbb{R}^{C_2\times 1\times 1}$ for $\boldsymbol{X}_2$, we compute global context via average pooling:

$$z = GAP(\boldsymbol{X}) = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}\boldsymbol{X}(i,j,:) \in \mathbb{R}^{C}. \tag{3}$$

Then, a lightweight fully-connected 1×1 convolution layer followed by a sigmoid function produces the gate:

$$\boldsymbol{M} = \sigma(W_g z + b_g), \tag{4}$$

where $\boldsymbol{W}_g \in \mathbb{R}^{C_2 \times C}$, $\boldsymbol{b}_g \in \mathbb{R}^{C_2}$, and $\sigma(\cdot)$ is the sigmoid function. This gate adaptively scales channels of $\boldsymbol{X}_2$ based on global image statistics.

Feature Transformation and Fusion $\boldsymbol{X}_1$ is refined through two depth-wise convolution blocks DWConv and PWConv with GELU activation:

$$\boldsymbol{X}_1' = \mathrm{PWConv}_{C_1 \to C_1}(\mathrm{GELU}(\mathrm{DWConv}_{k=3}(X_1))). \tag{5}$$

Finally, the transformed $\boldsymbol{X}_1'$ and gated $\boldsymbol{X}_2$ are concatenated and passed through a point-wise convolution to restore the original channel dimension:

$$\mathrm{AFU}(\boldsymbol{X}) = \mathrm{PWConv}_{C \to C}([\boldsymbol{X}_1'; \boldsymbol{M} \odot \boldsymbol{X}_2]), \tag{6}$$

where $[\cdot;\cdot]$ denotes channel concatenation and $\odot$ is element-wise multiplication. This design ensures that only half the channels are processed by DWConv, while the other half are adaptively weighted, striking an optimal balance between efficiency and representational power.

3.4 Residual Block

Residual Block (ResBlock) in our architecture is designed for minimal computational latency while preserving representational capacity. Each block follows the classic residual formulation:

$$\mathrm{Block}(\boldsymbol{X}) = \boldsymbol{X} + \mathcal{T}(\boldsymbol{X}), \tag{7}$$

where $\mathcal{T}(\cdot)$ denotes a learnable transformation. However, unlike the original ResNet blocks [6], we employ the Gaussian Error Linear Unit (GELU) [7] as the activation function to enable smoother gradient flow during optimization:

$$\mathrm{GELU}(x) = x\Phi(x), \tag{8}$$

with $\Phi(x)$ being the cumulative distribution function of the standard normal distribution.[1] Our ResBlock consists of two sequential 3 × 3 convolutional layers with GELU activation in between:

$$\boldsymbol{Y} = \mathrm{Conv}_{3\times3}(\mathrm{GELU}(\mathrm{Conv}_{3\times3}(\boldsymbol{X}))), \tag{9}$$

and

$$\mathrm{ResBlock}(\boldsymbol{X}) = \boldsymbol{X} + \boldsymbol{Y}, \tag{10}$$

preserving the input channel dimensionality.

[1] We use the widely adopted approximation: $\mathrm{GELU}(x) \approx 0.5x(1 + \tanh[\sqrt{2/\pi}(x + 0.044715x^3)])$.

3.5 Loss Function

The loss function consists of follows:

- $\mathcal{L}_{\text{L1}}$ enforces pixel-wise reconstruction accuracy:

$$\mathcal{L}_{\text{L1}} = \frac{1}{N}\sum_{i=1}^{N} \| J_i - \hat{J}_i \|_1, \tag{11}$$

  where J_i and $\hat{J}_i$ denote the ground-truth and predicted clear images for the i-th sample, respectively.
- $\mathcal{L}_{\text{perc}}$ captures perceptual similarity through deep features by extracting mid-level features. $\phi(\cdot)$ obtain the features from the first 16 layers of VGG-19 [16]:

$$\mathcal{L}_{perc} = \frac{1}{N}\sum_{i=1}^{N} \| \phi(\hat{J}_i) - \phi(J_i) \|_2^2. \tag{12}$$

- $\mathcal{L}_{\text{freq}}$ aligns images in the frequency domain by minimizing the discrepancy between the magnitude spectra of the reconstructed and target images:

$$\mathcal{L}_{freq} = \frac{1}{N}\sum_{i=1}^{N} \| |\mathcal{F}(\hat{J}_i)| - |\mathcal{F}(J_i)| \|_2^2 \tag{13}$$

  where $\mathcal{F}(\cdot)$ denotes the 2-D real-valued discrete Fourier transform with orthogonal normalization.
- $\mathcal{L}_{\text{cont}}$ encourages consistent global feature representations via contrastive learning. Given feature maps $\phi(\hat{J}_i)$ and $\phi(J_i)$, we apply spatial average pooling to obtain global descriptors $\bar{f}_i = GAP(\phi(\hat{J}_i))$ and $\bar{g}_i = GAP(\phi(J_i))$. The contrastive loss is formulated as:

$$\mathcal{L}_{cont} = \frac{1}{N}\sum_{i=1}^{N}(1 - sim(\bar{f}_i, \bar{g}_i)), \tag{14}$$

  where sim(·) is the cosine similarity.

The weighted summation of these losses is taken as the final loss for model training:[2]

$$\mathcal{L} = \mathcal{L}_{L1} + \lambda_{perc}\mathcal{L}_{perc} + \lambda_{freq}\mathcal{L}_{freq} + \lambda_{cont}\mathcal{L}_{cont}. \tag{15}$$

4 Experiments

We conduct extensive experiments to validate the effectiveness and efficiency of SLR-Net. This section details our experimental setup, presents quantitative and qualitative comparisons with state-of-the-art methods, and includes ablation studies to analyze the contribution of each component.

[2] The weights are set according to the performances on validation data, with ranges $\lambda_{perc} \in [0.1,0.2]$, $\lambda_{freq} \in [0.005,0.02]$, $and\ \lambda_{cont} \in [0.01,0.1]$.

4.1 Setup

Datasets All models are trained on RESIDE-6K dataset [9], which contains 6,000 hazy-clear image pairs. For evaluation, we use three benchmarks:

- RESIDE-6K-Test: The official test set (1,000 images) from the same domain as training.
- SOTS-Outdoor: A standard synthetic test set with 500 outdoor scenes [9].
- NH-HAZE: A real-world non-homogeneous haze dataset with 55 outdoor scenes [1], used to assess generalization to complex haze conditions.

Baselines We compare against a wide range of methods, including prior-based approaches, DCP [5] and CAP [25], as well as various deep learning models: DehazeNet [2], AOD-Net [8], GridDehazeNet (GDN) [10], FFA-Net [13], MSCNN [15], MSBDN [3], AECR-Net [21], RIDCP [22], and UCL-Dehaze [18].[3]

Evaluation Metrics We use Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) [20] for quantitative assessment. Inference time for all models is measured on a NVIDIA RTX 4070 Ti Super GPU.

Implementation Details The weighting coefficients of the loss function are set to $\lambda_{perc} = 0.2$, $\lambda_{freq} = 0.02$, and $\lambda_{cont} = 0.01$. To ensure fair comparison, we reimplement and train all competing models under identical settings (100 epochs, AdamW [11] optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, same data augmentation). We employ the OneCycle learning rate schedule [17] for SLRNet with a maximum learning rate of 1×10^{-3}. All models are trained for 100 epochs on RESIDE-6K [9] with a batch size of 32. Input images are randomly cropped to 256×256 patches during training. All experiments are conducted on a single NVIDIA RTX 4070 Ti Super GPU with 16GB of VRAM.

4.2 Results

Table 1 summarizes the performance of all compared methods across three benchmark datasets, and Table 2 summarizes the parameters and computational complexity of each model. SLRNet consistently outperforms the state-of-the-art in both PSNR and SSIM, demonstrating its superior performance and generalization capabilities. Our SLRNet consistently outperforms the state-of-the-art on all datasets, demonstrating its superior performance and generalization capabilities, highlighting the trade-off between accuracy, model efficiency, and generalization.

On SOTS-Outdoor, our SLRNet achieves a PSNR of 28.78 dB and SSIM of 0.9645, ranking second only to AECR-Net (29.34 dB) in PSNR but surpassing it in SSIM. Crucially, SLRNet uses only 96.5K parameters, over 26 × fewer than AECR-Net (2.55M), yet runs at 1.44 ms, making it significantly more efficient than heavy models like FFA-Net (75.87 ms) or UCL-Dehaze (2.94 ms). This demonstrates that SLRNet strikes an excellent balance between high accuracy and computational efficiency among modern deep dehazing networks.

[3] DCP and CAP are non-learning methods and thus have no training phase.

Table 1. Results on SOTS-Outdoor, RESIDE-6K-Test, and NH-HAZE datasets. The best result is bolded, and the second best is underlined.

Method	SOTS-Outdoor			RESIDE-6K-Test			NH-HAZE		
	PSNR (dB)	SSIM	Time (ms)	PSNR (dB)	SSIM	Time (ms)	PSNR (dB)	SSIM	Time (ms)
DCP* [5]	17.20	0.8482	19.18	16.90	0.8326	19.40	12.08	0.4698	208.86
CAP* [25]	23.35	0.9277	53.10	20.18	0.8700	58.50	**13.12**	**0.5061**	524.45
DehazeNet [2]	22.30	0.8487	<u>1.14</u>	21.22	0.8118	**0.53**	11.54	0.4460	<u>2.35</u>
AOD-Net [8]	22.51	0.8969	**0.89**	20.70	0.8482	<u>0.70</u>	11.61	0.4548	**2.31**
GridDehazeNet [10]	24.00	0.9094	7.79	22.97	0.8727	6.93	11.83	0.4712	12.84
MSCNN [15]	17.92	0.8335	44.50	15.42	0.7679	37.10	11.57	0.4412	91.87
MSBDN [3]	23.67	0.8583	1.15	22.44	0.8133	0.93	11.67	0.4494	5.82
FFA-Net [13]	27.94	0.9557	75.87	26.10	0.9363	81.36	12.07	0.4682	442.88
AECR-Net [21]	**29.34**	<u>0.9607</u>	3.47	**27.23**	**0.9453**	2.72	12.18	0.4747	8.84
RIDCP [22]	24.01	0.8700	1.52	23.10	0.8271	1.28	11.67	0.4357	3.53
UCL-Dehaze [18]	27.54	0.9476	2.94	25.33	0.9346	2.44	12.05	0.4638	7.30
SLRNet (Ours)	<u>28.78</u>	**0.9645**	1.44	<u>26.43</u>	<u>0.9379</u>	1.41	<u>12.25</u>	<u>0.4753</u>	3.50

Table 2. Comparison of parameter count and computational cost among different models.[4]

Method	Params	FLOPs
DehazeNet [2]	7,459	370,900,992
AOD-Net [8]	1,761	87,607,296
GridDehazeNet [10]	955,747	16,406,347,776
MSCNN [15]	265,153	2,986,676,224
MSBDN [3]	136,131	6,820,724,736
FFA-Net [13]	4,455,913	220,142,647,552
AECR-Net [21]	2,551,908	36,247,844,352
RIDCP [22]	742,659	9,569,566,720
UCL-Dehaze [18]	11,388,035	43,586,486,272
SLRNet (Ours)	96,515	4,791,206,688

On RESIDE-6K-Test, SLRNet attains 26.43 dB PSNR and 0.9379 SSIM, outperforming most lightweight models (e.g., DehazeNet, AOD-Net, RIDCP) and approaching the performance of much larger architectures like AECR-Net (27.23 dB), while being over 26 × smaller in parameter count. Inference time (1.41 ms) of SLRNet remains competitive with the fastest baselines, reinforcing its suitability for real-time applications without sacrificing reconstruction quality.

On the NH-HAZE dataset, a real-world, non-homogeneous hazy benchmark where models are evaluated without fine-tuning, SLRNet achieves 12.25 dB PSNR and 0.4753 SSIM, surpassing all methods except CAP (scores 13.12 dB), due to the timeconsuming handcrafted priors of CAP. It is worth noting that the absolute metric values on NH-

[4] Note that DCP and CAP are not included since they are non-learning methods.

HAZE are inherently limited by the well-known synthetic-to-real domain gap, as our model, like most supervised methods, is trained exclusively on synthetic data due to the scarcity of paired real-world hazy images. This performance drop is a universal challenge rather than a specific flaw of our architecture. Crucially, despite this domain shift, SLRNet still outperforms several significantly larger competitors (e.g., AECR-Net (12.18 dB) and FFA-Net (12.07 dB)) on this dataset. This relative superiority underscores the strong robustness and generalization capability of our lightweight design when deployed in complex, unseen real-world environments without specific domain adaptation, demonstrating the effectiveness of our architectural design in capturing generalizable dehazing cues.

To further evaluate the model complexity, we analyze the parameter count and computational cost detailed in Table 2. SLRNet exhibits exceptional compactness with only 96.51K parameters, which is significantly fewer than heavy-weight models like UCL-Dehaze (11.39M) and FFA-Net (4.46M). In terms of computational complexity, SLRNet requires approximately 4.79 GFLOPs. While this is moderately higher than extremely shallow networks like AOD-Net (0.09 GFLOPs), it remains substantially lower than complex architectures like GridDehazeNet (16.41 GFLOPs) and significantly lower than FFA-Net (220.14 GFLOPs). This confirms that SLRNet achieves state-of-the-art performance not by blindly increasing computational burden, but through an efficient architectural design that balances model capacity with inference speed.

4.3 Ablation Study

To validate the effectiveness of our architectural design and loss formulation in SLRNet, we conduct comprehensive ablation studies on the SOTS-Outdoor dataset. All models are trained for 10 epochs on the RESIDE-6K dataset under identical protocols and evaluated using PSNR and SSIM metrics.

Our ablation study is specifically designed to explore the optimal performance boundary *within* a strict lightweight constraint suitable for edge devices. We acknowledge that recent state-of-the-art works [4,12] have increasingly focused on maximizing restoration quality by incorporating computationally heavy modules, such as large-scale Transformers, iterative decoding pipelines, or complex domain adaptation strategies. However, these approaches fundamentally prioritize absolute accuracy over efficiency, often resulting in model sizes and latencies that are orders of magnitude larger than what is feasible for real-time edge deployment.

Consequently, directly integrating or comparing against such heavy components in our ablation study would be methodologically inconsistent with our core research objective: maximizing performance under severe parameter and latency constraints. Unlike studies that validate modules by scaling up model capacity, our goal is to isolate the contribution of each proposed component while maintaining the overall ultra-lightweight architecture. This ensures that the observed gains are directly attributable to our specific efficient design choices (e.g., AFU) rather than simply increased model capacity or the inclusion of computationally prohibitive modules prevalent in recent literature. This design choice allows us to fairly evaluate the true value of our contributions for the targeted resource-constrained scenarios.

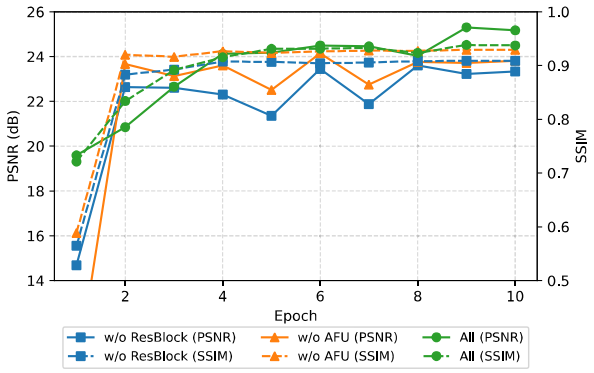

Fig. 2. Training dynamics of architectural variants.

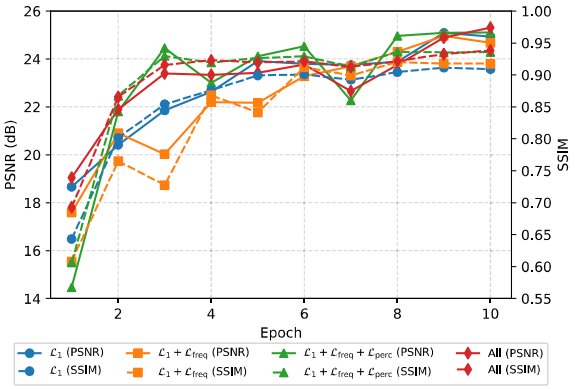

Fig. 3. Training dynamics on different training set.

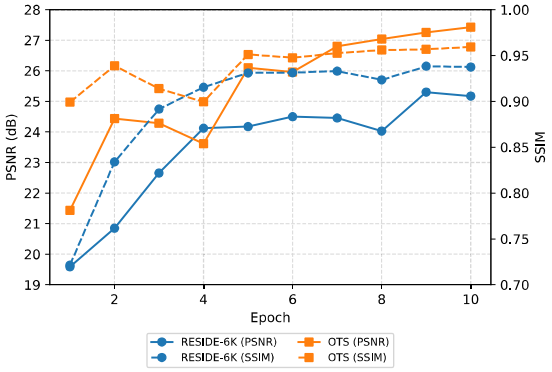

Fig. 4. Training dynamics on different training set.

Table 3. Ablation study

Architectural Variant	PSNR (dB)	SSIM
SLRNet	**25.3043**	**0.9383**
w/o ResBlock	23.6029	0.9088
w/o AFU	24.1415	0.9294
Loss Configuration		
$\mathcal{L}_1$	25.0877	0.9112
$\mathcal{L}_1 + \mathcal{L}_{freq}$	24.9558	0.9203
$\mathcal{L}_1 + \mathcal{L}_{freq} + \mathcal{L}_{perc}$	25.1037	0.9353

Table 4. Performance comparison of SLRNet trained on different datasets.

Training Datase	PSNR (dB)	SSIM
RESIDE-6K	25.3043	0.9383
OTS	27.4267	0.9593

Architectural Component Analysis We investigate the contribution of the two core modules in our network: AFU and the residual backbone (ResBlock). As shown in Table 2, removing either component significantly degrades performance: the variant without ResBlock achieves only 23.60 dB PSNR, while removing AFU yields 24.14 dB. In contrast, the full model reaches **25.30 dB** PSNR and **0.9383** SSIM, demonstrating a gain of over 1.7 dB compared to either module alone. Curves in Figure 2 further reveal that both components are essential for stable convergence and high-accuracy reconstruction, highlighting strong synergy between adaptive feature modulation and deep residual learning.

Loss Function Ablation We analyze the impact of our multi-term loss objective by starting from the basic ℓ_1 reconstruction loss and incrementally adding the frequency-domain term ($\mathcal{L}_{freq}$), the perceptual term ($\mathcal{L}_{perc}$), and the contrastive feature alignment term ($\mathcal{L}_{cont}$). As shown in Table 3 and Figure 3, incorporating $\mathcal{L}_{freq}$ yields a slight improvement in SSIM but a tiny drop in PSNR; further adding $\mathcal{L}_{perc}$ consistently enhances both metrics; and finally, including $\mathcal{L}_{cont}$ in the full objective delivers slight yet consistent gains, validating the effectiveness of high-level feature alignment.

Generalization Across Training Datasets To further assess the adaptability of our model, we additionally train the full SLRNet architecture on the more challenging OTS (Outdoor Training Set) [9] dataset for only 10 epochs and evaluate it on the same SOTS-Outdoor test set. As summarized in Table 4 and Figure 4, training on OTS—whose domain closely matches that of SOTS-Outdoor—leads to substantially improved performance: PSNR increases from 25.30 dB to 27.43 dB, and SSIM rises from 0.9383 to 0.9593. This demonstrates that our architecture is not only effective but also highly responsive to domain-aligned training data, achieving state-of-the-art accuracy when trained on a suitable dataset.

Fig. 5. Examples of different dehazing methods on the SOTS dataset with highlighted details. Zoom in for best view.

4.4 Visual Comparison

Figure 5 shows visual results on diverse hazy images from SOTS-Outdoor and NH-HAZE. SLRNet effectively removes haze while preserving fine details and color accuracy, avoiding common artifacts like color distortion (seen in DCP/CAP) or over-smoothing (seen in AOD-Net/LightClearNet). On the challenging NH-HAZE image with non-uniform haze, SLRNet produces a clearer and more natural result than several recent deep models.

5 Conclusion

In this work, we have presented SLRNet, a super lightweight residual network for image dehazing that achieves a strong trade-off between computational efficiency and restoration quality. By introducing the Adaptive Feature Unit that performs dynamic channel-wise feature recalibration through asymmetric processing and global context gating, SLRNet enables effective haze removal with minimal overhead. Comprehensive experiments on synthetic and real-world benchmarks show that it attains competitive PSNR and SSIM scores while maintaining ultra-low inference latency. By demonstrating competitive results even under the challenging synthetic-to-real domain gap, SLRNet proves that minimalistic architectures can achieve both high fidelity and practical deployability. These characteristics make it suitable for real-time applications on resource-constrained devices.

Acknowledgments. This work was partially supported by Yunnan Fundamental Research Projects (grant No. 202401CF070189) and the National Natural Science Foundation of China (62566064).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ancuti, C.O., Ancuti, C., Timofte, R.: NH-HAZE: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 1798–1805 (2020).
2. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: An end-to-end system for single image haze removal. IEEE Transactions on Image Processing 25(11), 5187–5198 (2016).
3. Dong, H., Pan, J., Xiang, L., Hu, Z., Zhang, X., Wang, F., Yang, M.H.: Multi-scale boosted dehazing network with dense feature fusion. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2154–2164 (2020).
4. Fu, J., Liu, S., Liu, Z., Guo, C.L., Park, H., Wu, R., Wang, G., Li, C.: Iterative predictor-critic code decoding for real-world image dehazing. In: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12700–12709 (2025).
5. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(12), 2341–2353 (2011).
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016).
7. Hendrycks, D., Gimpel, K.: Gaussian error linear units (gelus). arXiv: Learning. arXiv preprint arXiv:1606.08415 (2016).
8. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: Aod-net: All-in-one dehazing network. In: 2017 IEEE International Conference on Computer Vision. pp. 4780–4788 (2017).
9. Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z.: Benchmarking single-image dehazing and beyond. IEEE Transactions on Image Processing 28(1), 492–505 (2019).
10. Liu, X., Ma, Y., Shi, Z., Chen, J.: Griddehazenet: Attention-based multi-scale network for image dehazing. In: 2019 IEEE/CVF International Conference on Computer Vision. pp. 7313–7322 (2019).
11. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (2017).
12. Ma, L., Feng, Y., Zhang, Y., Liu, J., Wang, W., Chen, G.Y., Xu, C., Su, Z.: CoA: Towards Real Image Dehazing via Compression-and-Adaptation . In: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11197–11206 (2025).
13. Qin, X., Wang, Z., Bai, Y., Xie, X., Jia, H.: FFA-Net: Feature Fusion Attention Network for Single Image Dehazing. Proceedings of the AAAI Conference on Artificial Intelligence 34(07), 11908–11915 (2020).
14. Qiu, Y., Zhang, K., Wang, C., Luo, W., Li, H., Jin, Z.: Mb-taylorformer: Multi-branch efficient transformer expanded by taylor formula for image dehazing. In: 2023 IEEE/CVF International Conference on Computer Vision. pp. 12756–12767 (2023).
15. Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., Yang, M.H.: Single image dehazing via multi-scale convolutional neural networks with holistic edges. International Journal of Computer Vision 128(1), 240–259 (2020).
16. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: 3rd International Conference on Learning Representations. pp. 1–14 (2015).
17. Smith, L.N., Topin, N.: Super-convergence: Very fast training of neural networks using large learning rates. arXiv preprint arXiv:1708.07120 (2018).
18. Wang, Y., Yan, X., Wang, F.L., Xie, H., Yang, W., Zhang, X.P., Qin, J., Wei, M.: Ucl-dehaze: Toward real-world image dehazing via unsupervised contrastive learning. IEEE Transactions on Image Processing 33, 1361–1374 (2024).

19. Wang, Z., Zhao, H., Peng, J., Yao, L., Zhao, K.: Odcr: Orthogonal decoupling contrastive regularization for unpaired image dehazing. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 25479–25489 (2024).
20. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing 13(4), 600–612 (2004).
21. Wu, H., Qu, Y., Lin, S., Zhou, J., Qiao, R., Zhang, Z., Xie, Y., Ma, L.: Contrastive learning for compact single image dehazing. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10546–10555 (2021).
22. Wu, R.Q., Duan, Z.P., Guo, C.L., Chai, Z., Li, C.: Ridcp: Revitalizing real image dehazing via high-quality codebook priors. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 22282–22291 (2023).
23. Zhang, Y., Zhou, S., Li, H.: Depth information assisted collaborative mutual promotion network for single image dehazing. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2846–2855 (2024).
24. Zhou, J., Leong, C.T., Li, C.: Multi-scale and attention residual network for single image dehazing. In: 2021 6th International Conference on Intelligent Computing and Signal Processing. pp. 483–487 (2021).
25. Zhu, Q., Mai, J., Shao, L.: A fast single image haze removal algorithm using color attenuation prior. IEEE Transactions on Image Processing 24(11), 3522–3533 (2015).