An Image Dehazing Method with Multi-Loss Constraints and VGG-Based Perceptual Learning

*Xinjie Chen**
College of Engineering
Yanbian University
Jilin, China
2983178521@qq.com

***Abstract*—Single-image dehazing aims to restore clear scenes from haze-degraded images; however, due to the spatially non-uniform distribution of haze and the unobservable scene depth, it remains a highly ill-posed problem. To enhance the structural realism, semantic consistency, and color naturalness of dehazed images, this paper proposes an end-to-end single-image dehazing framework based on generative adversarial networks. The generator adopts a multi-scale encoder–decoder architecture with cross-scale feature fusion to jointly model local texture details and global contextual information, while the discriminator introduces capsule representations with dynamic routing on top of convolutional features to strengthen structure-consistency discrimination from a spatial relationship perspective. During training, a unified optimization objective is constructed by combining adversarial loss, pixel-wise L1 loss, SSIM-based structural loss, VGG perceptual loss, and color consistency loss, together with a stage-wise dynamic loss weighting strategy and synchronized learning rate scheduling to stabilize adversarial training and improve visual quality. Experimental results on the RESIDE dataset demonstrate that the proposed method outperforms several representative dehazing approaches, including GMAN net for image dehazing using TensorFlow, in terms of PSNR and SSIM, while qualitative comparisons further show that it effectively alleviates color shift, over-saturation, and highlight expansion artifacts, producing more natural and visually realistic dehazed results.**

Keywords—Image dehazing; Generative adversarial networks; Multi-scale feature fusion; Capsule networks; Perceptual loss; Color consistency

I. INTRODUCTION

In outdoor scenes, atmospheric particles such as haze, smoke, and water vapor cause light scattering and absorption, leading to reduced contrast, blurred details, and color distortion in captured images. Such degradation not only degrades human visual perception but also significantly impairs downstream vision tasks, including object detection, semantic segmentation, autonomous driving, and remote sensing, where the loss of structural and color cues may result in severe performance degradation or safety-critical misjudgments. Therefore, image dehazing, which aims to recover scene radiance from degraded observations, has become a fundamental front-end task for robust real-world visual systems.

Despite its importance, image dehazing remains a highly ill-posed inverse problem due to the strong coupling among scene radiance, atmospheric scattering, and depth, resulting in spatially non-uniform and scene-dependent degradation. Variations in depth, illumination, and surface reflectance, especially under complex lighting conditions such as nighttime, backlighting, and high-dynamic-range environments, further complicate haze distributions and make it difficult to simultaneously preserve structural sharpness, photometric consistency [1], and color fidelity. Moreover, excessive haze removal often introduces artifacts and color distortion, whereas conservative strategies leave residual haze and limited clarity.

Existing dehazing methods still exhibit inherent limitations under such conditions. Prior-based and physics-driven approaches rely on fixed assumptions that frequently fail in non-uniform haze or complex illumination, while data-driven methods mainly optimize pixel-wise losses (e.g., L1/L2), which are insufficient to ensure global structural coherence and semantic consistency. In addition, most adversarial discriminators emphasize local texture realism while neglecting long-range structural dependencies, leading to locally sharp yet globally inconsistent results. These issues, together with the domain gap of synthetic training data, severely restrict generalization to real-world scenes.

To overcome these challenges, we propose an end-to-end GAN-based single-image dehazing framework motivated by the observation that pixel-level supervision alone is insufficient to preserve structural coherence and perceptual realism under spatially non-uniform haze. A multi-scale encoder–decoder generator with cross-scale feature fusion is designed to jointly model haze degradation and scene structures, enabling the simultaneous recovery of local texture details and global contextual consistency across varying haze densities. This design is motivated by the observation that baseline dehazing models supervised only by adversarial and L1 losses provide coarse evaluation signals, which are insufficient to consistently assess and guide haze removal quality across different spatial scales, especially in regions with large depth variations. Furthermore, a capsule-based discriminator with dynamic routing is introduced to explicitly model part–whole spatial relationships, thereby enhancing structural realism beyond conventional patch-wise discrimination.

The model is trained using a unified objective integrating adversarial, L1, SSIM, VGG perceptual, and color consistency [2] losses, This design is motivated by the limitation of baseline adversarial and L1 supervision, which primarily evaluates pixel fidelity and local realism, but fails to provide effective constraints on structural integrity, perceptual coherence, and

global color distribution. In addition, a stage-wise dynamic weighting strategy and stabilized adversarial training techniques [3] are employed to progressively shift the optimization focus from low-level reconstruction accuracy to high-level perceptual fidelity. Extensive experiments on the RESIDE dataset demonstrate that the proposed method consistently outperforms representative baselines in both quantitative performance and visual quality.

II. Related Work

Early image dehazing methods are mainly built upon the Atmospheric Scattering Model (ASM) [4], which formulates a hazy image as a linear combination of scene radiance, transmission, and atmospheric light:

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

where $I(x)$ denotes the observed hazy image, $J(x)$ is the scene radiance, $t(x)$ is the transmission map, and Arepresents the global atmospheric light. The transmission is typically related to scene depth $d(x)$ and the scattering coefficient βvia an exponential decay $t(x) = e^{-\beta d(x)}$; as the scene depth tends to infinity, $t(x)$ approaches zero and the image gradually degenerates to the atmospheric light A. Within this framework, the Dark Channel Prior (DCP) [5] estimates transmission from local minimum statistics and achieves promising results, but it often produces halos and color distortions in depth discontinuities, strong edges, and sky regions.

With the development of deep learning, neural networks have been introduced to replace handcrafted priors for learning physical model parameters. DehazeNet [6] predicts the transmission $t(x)$ [7] using CNNs and restores the clear image via ASM, while AOD-Net [8] unifies transmission and atmospheric light into a single variable for end-to-end inference, reducing error accumulation and model complexity. Despite improved robustness, these methods remain limited by physical model assumptions. To address the difficulty of obtaining paired hazy–clear images, unpaired learning has been explored; Cycle-Dehaze [9] learns bidirectional mappings through cycle-consistency and perceptual constraints without explicit physical modeling, but still suffers from texture degradation and detail loss.

More recently, dehazing has shifted toward purely data-driven end-to-end architectures. GCA-Net [10] and FFA-Net [11] enhance modeling of spatially varying haze through multi-scale feature fusion and attention mechanisms, where FFA-Net integrates channel and pixel attention to distinguish thin and dense haze. Furthermore, AECR-Net [12], FSDGN [13], RIDCP [14], and Transformer-based methods (e.g., SelfPromer [15]) continue to improve performance through context modeling and long-range dependency learning. However, when the training objective is dominated by adversarial and pixel-wise losses, even multi-scale feature representations may remain insufficiently constrained, often leading to visually plausible yet structurally inconsistent or chromatically distorted dehazing results under complex haze conditions.

Nevertheless, most existing image dehazing methods predominantly rely on pixel-wise loss functions, such as L1/L2 or SSIM, which tend to produce over-smoothed structures under complex haze conditions and thus lack sufficient perceptual modeling capability. The VGG network [16] is able to capture hierarchical visual features ranging from low-level textures to high-level semantics, including edges, contours, and global layouts, and has been widely adopted as a generic visual feature extractor. Consequently, VGG-based perceptual loss [17] constrains the reconstructed image and the ground truth in the deep feature space, encouraging semantic and structural similarity from a human visual perception perspective rather than enforcing strict pixel-wise correspondence.

Recent advances in representation learning suggest that capsule networks are capable of encoding part–whole relationships and spatial hierarchies through dynamic routing mechanisms, offering a more structured alternative to traditional convolutional discriminators. Although capsule-based discriminators have shown promise in tasks such as image generation and inpainting, their potential for enforcing structural consistency in image dehazing remains underexplored. This observation motivates the integration of structure-aware adversarial supervision for more faithful haze removal.

III. Methods

A. Overall Framework and System Architecture

This paper proposes an end-to-end single-image dehazing framework based on a Generative Adversarial Network (GAN). The framework follows the standard adversarial learning paradigm and consists of a dual-network optimization system composed of a generator (G) and a discriminator (D).

The generator is responsible for learning the dehazing mapping, i.e., transforming images from the hazy domain to the clear domain, while the discriminator performs binary classification to distinguish real clear images from generated ones. Through this adversarial interaction, the discriminator provides informative gradients that guide the generator toward producing more realistic dehazed results. During inference, only the generator is deployed, enabling efficient and practical haze removal.

B. Data Pre-processing and Augmentation

All input images are resized to a fixed resolution of 256×256pixels and linearly normalized to the range $[0, 1]$, which ensures numerical stability and accelerates convergence during network optimization.

The training set is constructed in the form of paired hazy–clear images. To improve robustness to different haze densities, multiple hazy counterparts are synthetically generated from each clean image using the atmospheric scattering model, forming a one-to-many mapping between a single clear image and diverse hazy observations. This strategy significantly increases the diversity of haze conditions and prevents overfitting to specific degradation patterns.

To further enhance generalization under real-world imaging variability, synchronized data augmentation is applied during training. Specifically, the input hazy images are randomly flipped horizontally and subjected to brightness perturbation $\Delta b \in [-0.2, 0.2]$and contrast adjustment $\Delta c \in [0.8, 1.2]$. These stochastic transformations simulate variations in illumination

and atmospheric conditions, enabling the network to learn more robust and invariant feature representations.

C. Generator Architecture and Computational Pipeline

To effectively model the highly non-uniform and scale-varying characteristics of haze, the generator is designed as a multi-scale encoder–decoder network, whose forward computation proceeds from feature encoding to final image synthesis in a unified manner.

Given an input hazy image I_{hazy}, it is first fed in parallel into K down-sampling branches ($K = 3$ in this work). Each branch employs strided convolutions to progressively reduce the spatial resolution with scale factors 2^i $(i = 0,1,2)$, while simultaneously increasing the number of feature channels. This multi-branch design enables the network to capture complementary representations at different scales: shallow branches focus on low-level details such as edges and textures, whereas deeper branches encode high-level semantic and structural information. Consequently, the receptive field is effectively extended from local neighborhoods to the global image domain, allowing robust modeling of spatially varying haze distributions.

After multi-scale encoding, the extracted features are progressively upsampled back to the original resolution via transposed convolution layers. To preserve fine structures that may be lost during down-sampling, cross-layer skip connections are introduced, where feature maps from the l-th encoder layer are concatenated with those of the symmetric decoder layer along the channel dimension. Through this mechanism, low-level spatial details and high-level semantic representations are jointly propagated, which not only alleviates gradient vanishing but also significantly improves texture recovery and edge sharpness in the reconstructed image. The decoded features from all scales are concatenated and fused via a 1×1 convolution to explicitly enable cross-scale interaction, allowing the generator to jointly leverage local details and global context for more accurate haze – background separation while reducing over-smoothing and residual haze artifacts.

Finally, the fused representation is mapped to the RGB space via a 3×3convolution, yielding the restored clear image

$$I_{\text{clear}} = G(I_{hazy}; \theta_G), \tag{2}$$

where $\theta_G$denotes the learnable parameters of the generator. This output serves as the candidate dehazed result that will be further evaluated by the discriminator under the adversarial framework.

D. Discriminator Network Architecture

Complementary to the generator, the discriminator is designed to assess whether the generated clear image is both visually realistic and consistent with the input hazy observation. To this end, a conditional PatchGAN [18] is adopted, in which the input is the channel-wise concatenation of the image pair (I_{hazy}, I_{target}), where $I_{target} \in \{I_{real}, I_{clear}\}$.

This conditional formulation forces the discriminator to judge realism in the context of the given hazy image, rather than evaluating the clear image in isolation.

The network consists of four convolutional layers with a kernel size of 4×4 and a stride of 2, which progressively extract hierarchical joint representations from the paired inputs. Each layer is followed by a LeakyReLU [19] activation with a negative slope of 0.2, ensuring stable gradient propagation during adversarial learning.

To further capture long-range spatial dependencies and structural consistency, the resulting feature maps are fed into a Capsule Network module. Unlike conventional patch-based discriminators used in baseline adversarial training, the capsule mechanism explicitly models part–whole relationships, providing structure-aware adversarial feedback that complements the multi-scale feature fusion in the generator.

$$D(I_{hazy}, I_{target}) \in [0,1], \tag{3}$$

which measures the authenticity of the image pair.

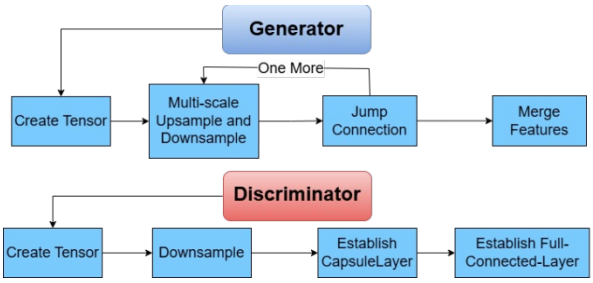

Figure 1. Architecture of the Generator–Discriminator Network

E. Joint Loss Function Design

While adversarial and L1 losses are commonly adopted in baseline dehazing methods, they provide limited evaluation signals for complex haze removal, as visually realistic results may still suffer from structural degradation, semantic inconsistency, or color over-saturation. Therefore, the generator is guided by a joint objective function that integrates five complementary constraints:

$$\mathcal{L}_G = \mathcal{L}_{adv} + \lambda_{L1}\mathcal{L}_{L1} + \lambda_{ssim}\mathcal{L}_{SSIM} + \lambda_{per}\mathcal{L}_{perceptual} + \lambda_{col}\mathcal{L}_{color}. \tag{4}$$

These losses jointly regulate the generator from pixel, structure, perception, and color distribution perspectives, ensuring both fidelity and visual realism.

Adversarial loss

The least-squares GAN (LSGAN) objective is adopted:

$$\mathcal{L}_{adv} = \mathbb{E}_{I_h \sim p_{hazy}}[(1 - D(I_h, G(I_h)))^2], \tag{5}$$

where $I_h$denotes a hazy input image and $G(\cdot)$and $D(\cdot)$represent the generator and discriminator, respectively.

L1 pixel-wise reconstruction loss

To enforce pixel-level consistency between the restored and ground-truth clear images:

$$\mathcal{L}_{L1} = \| I_{\text{real}} - I_{\text{clear}} \|_1, \tag{6}$$

SSIM structural similarity loss

Structural similarity is computed over an 11×11 sliding window:

$$\mathcal{L}_{SSIM} = 1 - \mathrm{SSIM}(I_{\mathrm{real}}, I_{\mathrm{clear}}), \quad (7)$$

VGG perceptual loss

High-level feature consistency is measured using a pretrained VGG-19 network. Feature maps are extracted from layers $\{2, 4, 8, 12, 16\}$after ReLU activations, and their squared Euclidean distance is minimized:

$$\mathcal{L}_{perc} = \sum_{j\in J} \| \phi_j(I_{\mathrm{real}}) - \phi_j(I_{\mathrm{clear}}) \|_2^2, \quad (8)$$

where $\phi_j(\cdot)$denotes the feature extraction function of the j-th VGG layer and all layers are equally weighted.

Color consistency loss

To constrain global color statistics and suppress color shifts, particularly preventing over-saturation and unnatural chromatic amplification in sky-dominant and high-illumination regions such as the sky and sun:

$$\mathcal{L}_{color} = \sum_{c\in\{R,G,B\}} | \mu_c(I_{\mathrm{real}}) - \mu_c(I_{\mathrm{clear}}) |, \quad (9)$$

where $\mu_c(\cdot)$denotes the mean value of channel c.

F. Dynamic Training Strategy

Although the above loss design provides comprehensive supervision, the optimization priorities should vary across training stages. Therefore, a dynamic training strategy is introduced to gradually shift the learning focus from low-level reconstruction to high-level perceptual realism.

A three-stage dynamic weighting scheme is employed to progressively shift the optimization focus from pixel fidelity to perceptual realism.

Let $p = \frac{t}{T}$ denote the normalized training progress, where t and T represent the current and total number of iterations, respectively. The loss weights are defined as

$$(\lambda_{L1}, \lambda_{SSIM}, \lambda_{perc}, \lambda_{color}) = \begin{cases} (80, 0.1, 0.05, 0.05), & 0 \le p < 0.3, \\ (50, 0.5, 0.3, 0.2), & 0.3 \le p < 0.7, \\ (20, 1.0, 0.8, 0.5), & 0.7 \le p \le 1. \end{cases} \quad (10)$$

In the first stage, a dominant pixel-wise constraint enforces stable low-level reconstruction. In the second stage, structural, perceptual, and color-consistency losses are strengthened to jointly optimize reconstruction fidelity and perceptual quality. In the final stage, perceptual and color constraints dominate, encouraging visually realistic and color-consistent dehazed results.

To further stabilize the adversarial optimization, the learning rates of both the generator and the discriminator are controlled by a performance-driven ReduceLROnPlateau scheduler based on the validation PSNR. When the PSNR does not improve for five consecutive epochs, the scheduler automatically reduces both learning rates by a factor of 0.5, which effectively suppresses oscillations and promotes convergence. To avoid numerical underflow and premature stagnation, the learning rate is lower-bounded by 1×10^{-6}.

Before full adversarial training, a short warm-up phase of five epochs is conducted to verify the convergence behavior of the proposed multi-term loss. During this phase, the learning rate is fixed at 1×10^{-4}, and the loss weights follow the first-stage configuration. The quantitative results obtained in this initialization stage serve as a baseline reference, ensuring that the network parameters are placed in a stable and meaningful region of the optimization landscape.

After the training process converges, only the generator with the optimized parameters $\theta_G^*$is retained for inference. Given any hazy input image I_{hazy}, the dehazed result is produced by a single forward pass:

$$I_{\mathrm{output}} = G(I_{hazy}; \theta_G^*) \quad (11)$$

No data augmentation or discriminator is involved at this stage, making the inference procedure both efficient and deterministic. The computational complexity is linear with respect to the image size, i.e., $O(n)$, where $n$denotes the number of pixels.

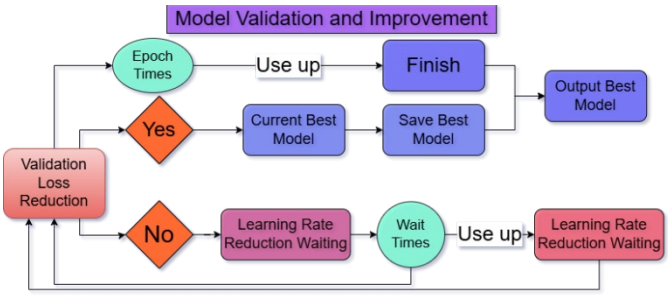

Figure 2. Framework of Validation-Driven Model Selection and Adaptive Learning Rate Adjustment

IV. Experiment

A. Dataset and Experimental Setup

The RESIDE dataset is adopted as the benchmark, where both training and testing sets are selected from the SOTS-Outdoor subset to ensure representative evaluation under real-world outdoor haze conditions. A total of 840 clean images are used for training and validation, and each clean image is associated with 32 synthetically generated hazy images with different haze levels. During each training epoch, three hazy images are randomly sampled from the 32 variants, which maintains data diversity while controlling computational cost. For testing, 75 clean–hazy image pairs from the SOTS-Outdoor test set are used to evaluate generalization performance.

All experiments are implemented using PyTorch 2.8.0 on Ubuntu 22.04 with Python 3.12. CUDA 12.8 is used for GPU acceleration. The hardware platform is equipped with a single NVIDIA RTX 4090 GPU with 24 GB memory.

B. Evaluation Metrics

PSNR (Peak Signal-to-Noise Ratio)

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{MAX^2}{\mathrm{MSE}}\right) \quad (12)$$

where MAX is the maximum possible pixel value (255 for 8-bit images), and MSE is the mean squared error between the restored image I and the ground truth I^{gt}.PSNR measures in

decibels (dB), quantifies the pixel-wise fidelity of reconstruction; higher PSNR indicates lower distortion.

SSIM (Structural Similarity Index)

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)} \quad (13)$$

SSIM evaluates luminance, contrast, and structural similarity, reflecting perceptual image quality.

C. Experimental results

Table 3. Quantitative performance evolution during two-stage training：

Phase	Epochs	Train PSNR	Train SSIM	Val PSNR	Val SSIM
I start	1	23.8	0.7974	25.36	0.8814
I end	5	28.33	0.9394	25.85	0.9137
Gain	–	4.53	0.142	0.49	0.0323
II start	6	26.15	0.9354	27.38	0.9456
II end	25	29.5	0.9624	30.18	0.9673
Gain	–	3.35	0.027	2.8	0.0217

Table 3 reports the performance evolution and gains on the training and validation sets during the two-stage training process.

Table 4. Final quantitative results on the test set：

Metric	I (5 epochs)	II (25 epochs)	Improvement
PSNR	21.82	26.48 ± 5.08	4.66
SSIM	0.8498	0.9332 ± 0.0545	0.0834

Table 4 compares the final quantitative performance on the test set after Stage I and Stage II training, along with the achieved improvements.

Table 5. Loss convergence and adversarial training stability:

Loss Term	Initial	Final	Change
L1 loss	0.0546	0.0297	−45.6%
Perceptual loss	5.2107	1.8979	−63.6%
Color loss	0.0001	0.0010	Constraint only
Total generator loss	10.2921	3.7856	−63.2%
Discriminator loss	1.30–1.40	1.30–1.40	Stable

Table 5 summarizes the convergence of individual loss terms and demonstrates the stability of adversarial training.

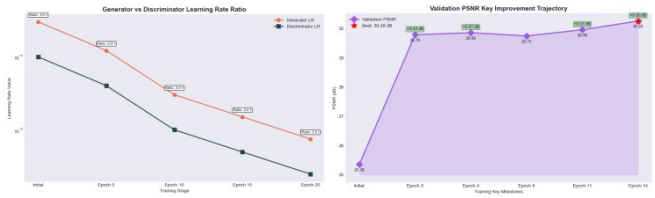

Figure 6. Coupled Evolution of Learning Rate Scheduling and Validation Performance

Figure 6 illustrates the coupled training dynamics of learning rate scheduling and validation performance. The left subfigure depicts the progressive decay of learning rates for the generator and discriminator across training stages, while the right subfigure shows the corresponding validation PSNR improvement at key training milestones. Overall, the results indicate that the adaptive learning rate strategy effectively stabilizes training and consistently enhances reconstruction performance.

Table 7.Comparison of Model Architectures and Training Strategies：

Aspect	Baseline Method	Proposed Method
Loss Function Composition	L1 loss + adversarial (GAN) loss	L1 + adversarial + SSIM + VGG perceptual + color consistency losses
Loss Weighting Strategy	Fixed weighting coefficients	Three-stage dynamic weighting (early / middle / late training)
Perceptual Constraint	Not used	VGG19-based perceptual loss for high-level feature alignment
Data Augmentation	Not applied	Random flipping, brightness and contrast perturbation
Discriminator Architecture	Capsule-based Patch GAN	Capsule-based conditional Patch GAN

Table 7 compares the baseline and proposed methods, indicating that the incorporation of multi-level perceptual constraints, dynamic optimization scheduling, and enhanced data diversity contributes to improved visual fidelity and more stable training behavior.

Table 8.Quantitative and Qualitative Performance Comparison：

Metric	Baseline Method	Proposed Method	Expected Gain
PSNR	22–25	25–28	+2–3 dB
SSIM	0.85–0.88	0.88–0.92	+0.03–0.04
Visual Quality	Residual haze, color distortion	Natural colors, clearer structures	Significant
Training Stability	Moderate	Improved (dynamic loss balancing)	Improved
Convergence Speed	Normal	Faster (adaptive weighting)	Slight

Computational Cost	Lower	Higher (perceptual features)	+20–30%

Table 8 indicates that the proposed method consistently improves both objective metrics and subjective visual quality, while incurring a moderate increase in computational complexity owing to perceptual feature modeling.

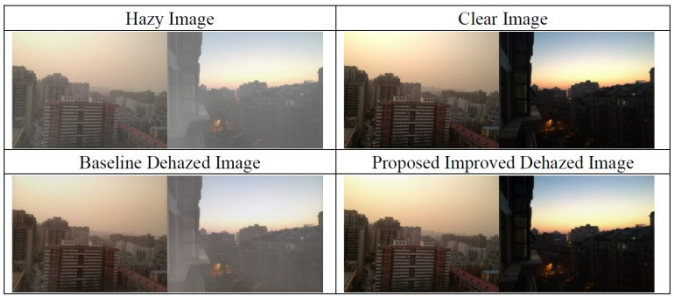

Figure 9. Visual comparison on representative outdoor hazy scenes

Beyond quantitative metrics, visual quality is a critical criterion for evaluating image dehazing performance. Figure 9 presents visual comparisons under varying haze densities, scene depths, and illumination conditions, including Hazy Image, the corresponding Clear Image, Baseline Dehazed Image, and the Proposed Improved Dehazed Image within a unified framework. This visual comparison highlights differences in structure recovery, detail reconstruction, and color fidelity, providing complementary evidence of the effectiveness and visual superiority of the proposed method.

V. Conclusions and Limitation

This paper proposes an end-to-end single-image dehazing framework based on generative adversarial networks (GANs), jointly optimizing adversarial, pixel-wise, structural, perceptual, and color-consistency losses. By integrating multi-level constraints, the method effectively preserves structural integrity, recovers textures, and maintains color fidelity under complex haze conditions. The generator employs a multi-scale encoder – decoder architecture with cross-scale feature fusion to balance local detail enhancement and global context modeling. A capsule-based discriminator with dynamic routing is introduced to capture spatial structure and geometric consistency beyond conventional real – fake discrimination, providing more structure-aware adversarial supervision.

Despite its effectiveness, the framework still faces limitations, including reliance on paired training data, limited color modeling under complex illumination, and high computational cost. Future work will focus on lightweight architectures, improved color representation, higher-resolution training, and joint optimization with downstream vision tasks such as object detection and semantic segmentation.

VI. References

[1] Yu, J. J., Forghani, F., Derpanis, K. G., & Brubaker, M. A. (2023). Long-term photometric consistent novel view synthesis with diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 7094-7104).

[2] Peng, S., Gao, T., Qu, S., Yu, Z., Wang, J., & Peng, J. (2025). An evaluation and study of detail contrast preservation and color consistency in decolorization. *Digital Signal Processing*, 105468.

[3] Bai, T., Luo, J., Zhao, J., Wen, B., & Wang, Q. (2021). Recent advances in adversarial training for adversarial robustness. *arXiv preprint arXiv:2102.01356*.

[4] Yan, W., & Cui, L. (2024). Image dehaze algorithm based on improved atmospheric scattering models. *IEEE Access*.

[5] Golts, A., Freedman, D., & Elad, M. (2019). Unsupervised single image dehazing using dark channel prior loss. *IEEE transactions on Image Processing*, *29*, 2692-2701.

[6] Srivastava, M. K., Ranjan, R., & Kedar, S. (2025). DehazeNet-assisted fog removal and channel estimation for robust visible light V2V communication. Journal of Optical Communications, (0).

[7] Li, B., Zhao, J., & Fu, H. (2020). DLT-Net: deep learning transmittance network for single image haze removal. *Signal, Image and Video Processing*, *14*(6), 1245-1253.

[8] Li, B., Peng, X., Wang, Z., Xu, J., & Feng, D. (2017). Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE international conference on computer vision* (pp. 4770-4778).

[9] Sopian, S. M., Satyawan, A. S., Haqiqi, M. M. E., Susilawati, H., Nurpalah, R., & Hamdani, N. A. (2025, July). Enhancing Hazy Image Dehazing Using a Modified CycleGAN. In 2025 International Seminar on Intelligent Technology and Its Applications (ISITIA) (pp. 266-270). IEEE.

[10] Zhang, X., Yu, B., Hua, L., Yi, R., He, P., & Yuan, X. (2025, July). GCA-Net: Global Contextual Attention Network with Lightweight Hierarchical Alignment for Text-Guided Fashion Image Retrieval. In International Conference on Intelligent Computing (pp. 284-295). Singapore: Springer Nature Singapore.

[11] Jing, H., Ren, Y., Han, M., Hu, J., Yang, W., & Zhang, M. (2025, April). PEFA-Net: A Parallel Enhanced Feature Attention Network for single image dehazing. In 2025 10th International Conference on Computer and Communication System (ICCCS) (pp. 361-366). IEEE.

[12] Song, X., Dai, J., Fu, Z., & Tan, D. (2024, July). Optimization Improvement of AECR-NET Image Dehazing Network. In 2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS) (pp. 1428-1436). IEEE.

[13] Yu, H., Zheng, N., Zhou, M., Huang, J., Xiao, Z., & Zhao, F. (2022, October). Frequency and spatial dual guidance for image dehazing. In *European conference on computer vision* (pp. 181-198). Cham: Springer Nature Switzerland.

[14] Wu, R. Q., Duan, Z. P., Guo, C. L., Chai, Z., & Li, C. (2023). Ridcp: Revitalizing real image dehazing via high-quality codebook priors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 22282-22291).

[15] Wang, C., Pan, J., Lin, W., Dong, J., Wang, W., & Wu, X. M. (2024, March). Selfpromer: Self-prompt dehazing transformers with depth-consistency. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 38, No. 6, pp. 5327-5335).

[16] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[17] An, T., Mao, B., Xue, B., Huo, C., Xiang, S., & Pan, C. (2023). Patch loss: A generic multi-scale perceptual loss for single image super-resolution. *Pattern Recognition*, *139*, 109510.

[18] Chen, G., Zhang, G., Yang, Z., & Liu, W. (2023). Multi-scale patch-GAN with edge detection for image inpainting. *Applied intelligence*, *53*(4), 3917-3932.

[19] Xu, J., Li, Z., Du, B., Zhang, M., & Liu, J. (2020, July). Reluplex made more practical: Leaky ReLU. In *2020 IEEE Symposium on Computers and communications (ISCC)* (pp. 1-7). IEEE.