Disentanglement-wise Image Dehazing through Cross-Domain Manifold Consensus

Supplementary Material

Abstract

This supplementary material first presents a theoretical justification of the Cross-domain Invariant Manifold (CIM) under the atmospheric scattering model. It further elaborates the mathematical derivation of our physically-grounded HSV constraint and provides comprehensive details of the complete network architecture. The document also includes extensive experimental analyses, encompassing domain ablation studies and additional visual comparisons that demonstrate the superiority of CIM-D across diverse hazy scenarios. Finally, we conclude the broader impacts of our proposed approach.

A. Theoretical Justification of the Cross-domain Invariant Manifold (CIM)

Setting and notation. Let $\boldsymbol{I}_s^i \in \mathcal{X}$ denote the i-th image with state $s \in \{h, c\}$, where h and c correspond to hazy and clear, respectively. $\mathcal{X}$ denotes the training dataset containing hazy and clear images. For each of the five domains $k \in \{1, \ldots, 5\}$ in CIM-D (namely spatial (SPA), frequency (FRE), non-local (NOL), diffusion (DIF), and compressed-sensing(CS)) a domain-specific encoder Φ_k processes $\boldsymbol{I}_s^i$ and outputs a global feature vector

$$\boldsymbol{f}_s^{i,k} = \Phi_k(\boldsymbol{I}_s^i) \in \mathbb{R}^{D_k}_{\text{feat},k}, \tag{18}$$

where D_k is the feature dimensionality of domain k, and $\mathbb{R}^{D_k}_{\text{feat},k}$ denotes the corresponding feature space. Then, a shared domain-translation network $\mathcal{P}$ maps each domain feature to a unified CIM representation:

$$\boldsymbol{z}_s^{i,k} = \mathcal{P}\big(\boldsymbol{f}_s^{i,k}\big) \in \mathcal{M}, \tag{19}$$

where $\mathcal{M}$ denotes the cross-domain invariant manifold, i.e., a low-dimensional latent space that contains the embeddings $\boldsymbol{z}_s^{i,k}$ produced by $\mathcal{P}$.

A.1 ASM scattering semantic on CIM

Although haze manifests differently across domains, its formation is universally governed by the Atmospheric Scattering Model (ASM) [42]:

$$\boldsymbol{I}_h = \boldsymbol{J} \cdot t + \boldsymbol{A} \cdot \big(1 - t\big), \tag{20}$$

where $\boldsymbol{I}_h$ is the hazy observation, $\boldsymbol{J}$ is the scene radiance, t is the transmission map, and $\boldsymbol{A}$ is the global atmospheric light. A clear image $\boldsymbol{I}_c$ corresponds to the limiting case $t \to 1$, causing the scattering term to vanish.

To formally connect the CIM with the Atmospheric Scattering Model (ASM), we recognize that haze formation is governed exclusively by the transmission map t and the atmospheric light $\boldsymbol{A}$, while the scene radiance $\boldsymbol{J}$ represents the underlying content, independent of the scattering process. (**This principle has been empirically validated by recent work,** PHAT [50], which demonstrates that the scattering parameters $(t, \boldsymbol{A})$ extracted from hazy images can be transferred to arbitrary clear images to synthesize new hazy scenes. This confirms that the scattering process is indeed separable from scene content and can be shared across different images.) This physical distinction motivates the definition of an abstract *scattering state* for an image $\boldsymbol{I}_s$ as a function of the scattering parameters:

$$\eta(\boldsymbol{I}_s) = \phi(t, \boldsymbol{A}) \in \mathbb{R}^p_{\text{scat}}, \tag{21}$$

where $\mathbb{R}^p_{\text{scat}}$ denotes the p-dimensional scattering-parameter space of $(t, \boldsymbol{A})$. ϕ maps the physical scattering parameters to an abstract state representation $\eta(\boldsymbol{I}_s)$. Critically, the scene radiance $\boldsymbol{J}$ is **excluded** from this state. It is precisely the entanglement between the inherent scene attributes $\boldsymbol{J}$ and the scattering variables $(t, \boldsymbol{A})$ that causes conventional methods to misidentify haze characteristics. The scattering state $\eta(\boldsymbol{I}_s)$ serves as a physical anchor to disentangle this ambiguity, encoding only the scattering condition irrespective of scene content.

Note that the scattering state $\eta(\boldsymbol{I}_s)$ is an abstract representation, as real-world haze involves more complexities than two global parameters. We do not require the CIM to explicitly reconstruct $(t, \boldsymbol{A})$; instead, we design the learning such that the dominant variations in $\mathcal{M}$ correlate with changes in the scattering state and are invariant to scene content $\boldsymbol{J}$. This leads to a manifold structured by scattering semantics.

A.2 Cross-domain Consensus Learning on CIM

The extraction of scattering states fundamentally relies on the regularity of atmospheric scattering effects as images transition from clear to hazy conditions. This regularity inherently encodes atmospheric scattering semantics. In our work, we select five distinct domains—spatial, frequency, non-local, diffusion, and compressed-sensing—as our processing targets precisely because haze exhibits discernible and consistent patterns across these domains, as visually demonstrated in Fig. 9:

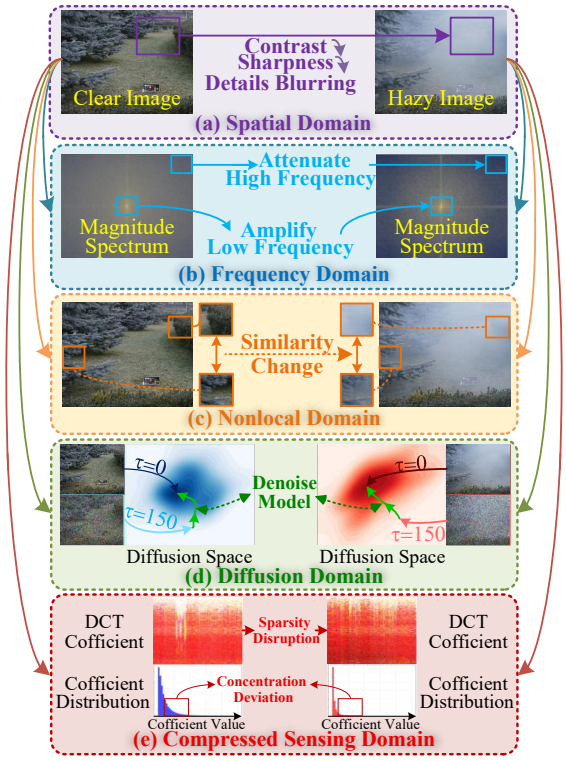

Figure 9. The impact of haze across different domains

- **Spatial Domain:** Haze systematically reduces image contrast and sharpness, enhancing low-frequency components while diminishing high-frequency details.
- **Frequency Domain:** Haze disturbs the magnitude spectrum distribution, causing attenuation of high-frequency coefficients and concentration of low-frequency energy.
- **Non-local Domain:** Haze disrupts inherent similarity relationships between image patches. Specifically, it reduces similarity between originally similar patches, while artificially increasing similarity between originally dissimilar patches due to atmospheric light homogenization—thereby interfering with non-local self-similarity priors.
- **Diffusion Domain:** Haze alters the feature evolution trajectory during the diffusion process, causing the denoising path to deviate from that of clear images.
- **Compressed-sensing Domain:** Haze impairs sparse representation in the compressed-sensing domain. For example, in the DCT [29] domain, haze disturbs the sparsity of transform coefficients and causes their distribution to deviate from the expected concentration.

In an ideal scenario, the projection from each domain would yield the same physics semantics embedding for a given image. However, due to domain-specific biases, we model the projection as:

$$\boldsymbol{z}_s^{i,k} = \bar{\boldsymbol{z}}_s^i + \delta_k, \tag{22}$$

where $\bar{\boldsymbol{z}}_s^i$ is the ideal consensus embedding depending only on $\eta(\boldsymbol{I}_s^i)$, and δ_k is the domain-specific residual.

To robustly estimate $\bar{\boldsymbol{z}}_s^i$ and suppress δ_k, we use the consensus density $\rho_s(\boldsymbol{z})$ (geometric mean of domain-wise densities). This product-of-experts formulation ensures that high density is assigned only to points with cross-domain agreement, thus filtering out domain-specific deviations.

Building upon the structured manifold, $\mathcal{L}_{ssa}$ (Eq. 8) reinforces consensus by pulling together high-density embeddings across domains while pushing away low-consensus-density samples $\boldsymbol{m}_-$, creating a self-reinforcing cycle that sharpens density peaks. Meanwhile, $\mathcal{L}_{dmt}$ guides the dehazed embedding $\boldsymbol{m}_d$ (Eq. 10) toward high-density clear regions ($\boldsymbol{\mu}_c$) and away from both hazy and low-consensus-density areas ($\boldsymbol{\mu}_h$ and $\boldsymbol{m}_-$). The clear and hazy prototypes $\{\boldsymbol{\mu}_c\}$ and $\{\boldsymbol{\mu}_h\}$, obtained via density-peak clustering of their respective consensus density fields, capture multiple modes of the scattering state distribution and provide diverse semantic anchors for the traversal process. The adaptive weight w from Eq. 10 ensures this traversal remains within physically plausible regions.

B. Derivation of Physically-Grounded HSV Constraint

HSV-RGB Conversion Given HSV values with $H \in [0,1]$, $S \in [0,1]$, $V \in [0,1]$, we can derive these components:

$$\begin{cases} C = VS, \\ X = C\big(1 - |(6H \bmod 2) - 1|\big), \\ m = V - C, \end{cases} \tag{23}$$

Let i be the integer part of $6H$ (i.e., $i = \lfloor 6H \rfloor$), which serves as the hue sector index. The RGB components are then given by

$$(R, G, B) = (R_i + m,\ G_i + m,\ B_i + m), \tag{24}$$

where

$$(R_i, G_i, B_i) = \begin{cases} (C, X, 0), & i = 0, \\ (X, C, 0), & i = 1, \\ (0, C, X), & i = 2, \\ (0, X, C), & i = 3, \\ (X, 0, C), & i = 4, \\ (C, 0, X), & i = 5. \end{cases} \tag{25}$$

For clarity of exposition, we first detail the derivation for hue sector $i = 0$; the final constraint is independent of i and

therefore holds for all sectors $i \in \{0, \dots, 5\}$. In this case, the spectral-balanced scene radiance $\boldsymbol{J^w} = (J_R^w, J_G^w, J_B^w)$ can be written as

$$\begin{cases} J_R^w = V_{J^w}, \\ J_G^w = V_{J^w} - V_{J^w} S_{J^w} \left|(6H_{J^w} \bmod 2) - 1\right|, \\ J_B^w = V_{J^w} - V_{J^w} S_{J^w}, \end{cases} \tag{26}$$

and similarly the hazy image $\boldsymbol{I}^w = (I_R^w, I_G^w, I_B^w)$ under a neutral atmospheric light $\boldsymbol{A}^w$ (with saturation $S_{A^w} \approx 0$) is given by

$$\begin{cases} I_R^w = V_{I^w}, \\ I_G^w = V_{I^w} - V_{I^w} S_{I^w} \left|(6H_{I^w} \bmod 2) - 1\right|, \\ I_B^w = V_{I^w} - V_{I^w} S_{I^w}. \end{cases} \tag{27}$$

Under neutral airlight, the hue is preserved, so $H_{I^w} = H_{J^w}$ in this case.

Remark on $\boldsymbol{J}$, $\boldsymbol{J}^w$ and HSV Constraint Derivation. In the classical atmospheric scattering model, the symbol $\boldsymbol{J}$ denotes a haze-free scene radiance under an atmospheric light $\boldsymbol{A}$. In that formulation $\boldsymbol{A}$ is written as a global $3 \times 1 \times 1$ vector. However, subsequent works have shown that, especially for images with large sky regions or non-uniform haze, it is more realistic to model the atmospheric light as a smoothly varying *airlight field* $\boldsymbol{A}$ or airlight map rather than a single constant vector [35, 43, 49]. In this supplement we adopt this more general view and treat $\boldsymbol{A}$ as a spatially varying map.

Based on the airlight map $\boldsymbol{A}$, one [26] can conceptually factor the clear radiance as $\boldsymbol{J} = \boldsymbol{R} \odot \boldsymbol{A}$, where $\boldsymbol{R}$ is the intrinsic surface reflectance and $\boldsymbol{A}$ plays the role of illumination. When $\boldsymbol{A}$ is chromatically biased, the corresponding $\boldsymbol{J}$ also inherits color cast casused by atmospheric map $\boldsymbol{A}$. To work in a better-conditioned domain, our method introduces a learnable spectral-balance matrix $\boldsymbol{W} \in \mathbb{R}^{3\times \mathrm{H}\times \mathrm{W}}$ that acts elementwise on each pixel, and we define the spectrally balanced quantities:

$$\boldsymbol{A}^w = \boldsymbol{A} \odot \boldsymbol{W}, \tag{28}$$

$$\boldsymbol{J}^w = \boldsymbol{J} \odot \boldsymbol{W} \tag{29}$$

$$= \boldsymbol{R} \odot \boldsymbol{A} \odot \boldsymbol{W} \tag{30}$$

$$= \boldsymbol{R} \odot \boldsymbol{A}^w, \tag{31}$$

where $\odot$ denotes element-wise multiplication. The spectrally balanced clear image $\boldsymbol{J}^w$ removes the color bias of $\boldsymbol{A}$ while preserving the underlying reflectance $\boldsymbol{R}$.

In the derivation of the HSV constraint below, we consistently use $\boldsymbol{J}^w$ to denote the spectrally balanced clear radiance, which aligns with the output of our disentanglement-wise HSV network. For notational simplicity throughout the main paper, we reuse the symbol $\boldsymbol{J}$ (without superscript) to denote the dehazed image predicted by our network.

Then, based on spectrally balanced output $\boldsymbol{J}^w$, we substitute the HSV expressions into the atmospheric scattering model for each channel:

$$\boldsymbol{I}^w = \boldsymbol{J}^w t + \boldsymbol{A}^w (1 - t), \tag{32}$$

one can collect terms in V and SV to obtain the following two scalar equations:

$$\begin{aligned} V_{I^w} &= V_{J^w}\, t + V_{A^w}(1 - t), \\ S_{I^w} V_{I^w} &= S_{J^w} V_{J^w}\, t + S_{A^w} V_{A^w}(1 - t), \end{aligned} \tag{33}$$

with $S_{A^w} \to 0$ for neutral airlight. Eliminating the transmission t from these equations yields the spectrally balanced scattering constraint

$$\frac{V_{I^w} S_{I^w}}{V_{A^w} - V_{I^w}} = \frac{V_{J^w} S_{J^w}}{V_{A^w} - V_{J^w}}, \tag{34}$$

where, for notational simplicity, $(V_{I^w}, S_{I^w}, V_{A^w})$ are understood in the spectrally balanced domain. By the piecewise-linear structure of the HSV↔RGB mapping, this derivation holds for all hue sectors $i \in \{0, \dots, 5\}$. At this point, we have thus derived the constraint term employed in Eq. (15) of the main text.

C. Network Architecture and Implementation Details

C.1 Domain-specific feature extractors

CIM-D employs five lightweight domain-specific feature extractors $\{\Phi_k\}_{k=1}^5$ that operate on the same input image but encode complementary scattering effects induced by the atmospheric scattering model (ASM).

Spatial feature extractor. We adopt a VGG-style convolutional encoder [47] to capture local structural information affected by haze, such as contrast decrease and edge degradation. This design choice is motivated by VGG's proven effectiveness in representing local image structures, which are critically altered during haze formation. The encoder outputs a spatial feature vector $\boldsymbol{f}_s^{i,1} \in \mathbb{R}^{D_1}_{\text{feat},1}$ that encodes local contrast variations and structural information degradation due to atmospheric scattering.

Frequency Feature Extractor. We employ a frequency-domain encoder based on Fast Fourier Convolution (FFC) [8] to explicitly model the frequency-dependent effects of atmospheric scattering. This design is motivated by the physical insight that haze simultaneously suppresses high-frequency components (corresponding to texture and edge

details) while enhancing low-frequency components (associated with the atmospheric veil). By analyzing the spectral distribution, the encoder captures this characteristic frequency shift induced by scattering phenomena. The resulting frequency feature vector $\boldsymbol{f}_s^{i,2} \in \mathbb{R}^{D_2}_{\text{feat},2}$, encodes the distinctive spectral signature of haze, providing complementary information to spatial-domain representations.

Non-local feature extractor. We employ a non-local attention encoder [52] to capture long-range dependencies altered by haze scattering. This design is motivated by the observation that atmospheric scattering disrupts natural self-similarity patterns across image regions, making global affinity relationships crucial for characterizing haze distribution. The encoder outputs a non-local feature vector $\boldsymbol{f}_s^{i,3} \in \mathbb{R}^{D_3}_{\text{feat},3}$, that encodes changes in long-range correlations and self-similarity patterns induced by haze.

Diffusion feature extractor. For diffusion domain, our implementation uses a simplified forward-diffusion plus denoising pipeline inspired by DDPM-style denoisers [23]. This design is motivated by the observation that haze fundamentally alters the natural image distribution, shifting it away from clear image statistics. The diffusion process inherently models these distributional shifts, allowing us to extract features that characterize how haze perturbs the image's trajectory through the diffusion space. Given an input image, we sample a small set of timesteps $\{\tau_\ell\}_{\ell=1}^{L}$ from a fixed noise schedule and generate noisy versions x_{τ_ℓ} via the forward diffusion process. A lightweight U-Net denoiser with shared weights across timesteps then predicts the clean image from each x_{τ_ℓ}. We compute residuals between the noisy and denoised images, fuse them with corresponding timestep embeddings, and employ a final linear layer to produce the diffusion feature vector $\boldsymbol{f}_s^{i,4} \in \mathbb{R}^{D_4}_{\text{feat},4}$. This feature vector captures how the image responds to iterative diffusion-based denoising under haze, reflecting the statistical deviations induced by scattering effects.

Compressed-sensing feature extractor. We implement a compressed-sensing based encoder to capture how haze degradation affects the sparse representations of natural images. This approach is motivated by the observation that haze introduces structured noise that disrupts the inherent sparsity of clear images in transform domains. Our implementation follows non-iterative CNN reconstructor designs such as ReconNet [30]. Specifically, the input image is vectorized and projected using a fixed random Gaussian measurement matrix (with orthonormalized rows) at a specified sampling ratio, producing compressed measurements $\boldsymbol{y}$. These measurements are processed by a shallow CNN-based reconstructor that generates a coarse reconstruction $\hat{\boldsymbol{x}}$. We compute the reconstruction residual $\boldsymbol{r} = \boldsymbol{x} - \hat{\boldsymbol{x}}$ and concatenate $\hat{\boldsymbol{x}}$ with $\boldsymbol{r}$ along the channel dimension. A convolutional head followed by global pooling and a linear layer then produces the compressed-sensing feature vector $\boldsymbol{f}_s^{i,5} \in \mathbb{R}^{D_5}_{\text{feat},5}$, which encodes how haze perturbs structured sparsity and reconstruction behavior.

Notably, the backbones of all feature extractors are pretrained as provided in their original publications and remain fixed throughout the training of CIM-D.

C.2 Domain Translation Network and Spectral-Balance Network

Domain translation network. The domain translation network transforms domain-specific features into a unified cross-domain invariant manifold through a two-stage architecture, as illustrated in Fig. 10.

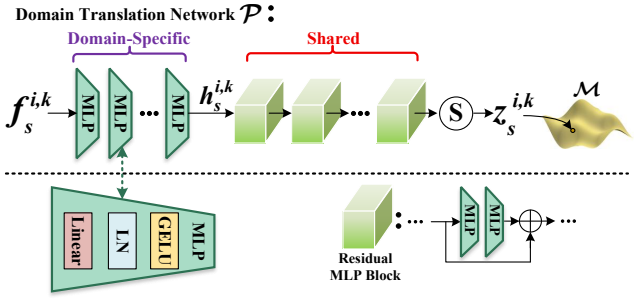

Figure 10. The Structure of Domain translation network $\mathcal{P}$, where "LN" indicate the layer normalization.

(1) Domain-specific projection heads: Each domain feature $\boldsymbol{f}_s^{i,k} \in \mathbb{R}^{D_k}_{\text{feat},k}$ first passes through a dedicated 4-layer MLP projection head, with each layer comprising linear transformation, layer normalization, and GELU activation. These domain-specific projections serve to unify feature dimensions and perform initial coordinate transformation while preserving domain characteristics.

(2) Shared Manifold Mapping: The projected features $\boldsymbol{h}_s^{i,k}$ are then transformed by a shared network composed of 4 residual MLP blocks. This network, followed by Sigmoid activation function, shared mapping enforces cross-domain alignment by projecting all features into the common CIM manifold $\mathcal{M}$, where points from different domains representing the same physical scene cluster in high-density regions. The resulting coordinates $\{\boldsymbol{z}_s^{i,1}, \ldots, \boldsymbol{z}_s^{i,5}\}$ on $\mathcal{M}$ capture consistent scattering semantics regardless of their original domain.

Spectral-balance network. Inspired by [1], the spectral-balance module implements the color compensation described in Sec. 3.2 through a lightweight U-Net architecture $\mathcal{B}\psi$. This network maps the hazy image $\boldsymbol{I}$ to its spectrally balanced version $\boldsymbol{I}^w = \mathcal{B}\psi(\boldsymbol{I}) \cdot \boldsymbol{I}$, which is subsequently

utilized in the ASM and HSV constraints. The U-Net follows a symmetric encoder-decoder design, consisting of multiple downsampling and upsampling stages, ReLU activations, and skip connections between corresponding encoder and decoder levels. We initialize $\mathcal{B}_\psi$ to approximate a basic white-balancing function based on the color consistency dataset [2], providing a reasonable starting point for color correction. Nevertheless, such hand-crafted priors cannot handle severely degraded scenes (e.g., dense haze or sandstorms). Hence, during training, the network learns residual spectral corrections under the guidance of the physical losses in Eq. 15-16, progressively refining its output toward the ideal balanced radiance $\boldsymbol{J}^w$.

C.3 Training Details

This section provides additional training parameter details that were omitted from the main paper due to space constraints. In Eq. (3), the bandwidth parameter σ is set to 0.15 to control the neighborhood similarity scale in contrastive learning. In Eq. (9), the scaling factor σ_g is set to 0.45 to regulate the similarity scaling in positive pair distribution. Both parameters are annealed during training: they start from larger values (1 and 1.9, respectively) and gradually decay to their final values following a cosine schedule with a total of 40 steps, enabling the model to first establish coarse cross-domain consensus before refining fine-grained similarity. We enhance each MConv block by integrating a Mixed Attention Module at its end (see Fig .5), which adaptively recalibrates the feature responses. For all convolutional layers, we adopt ReLU variants (specifically Leaky ReLU with a negative slope of $\alpha = 0.01$ in our implementation) as the activation function. The final output is obtained via a Tanh activation followed by linear scaling to [0,1]. To ensure stable training, the atmospheric light intensity V_A is estimated robustly as the average of the top 0.1% brightest pixels in V_I. For the channel decoupling loss $\mathcal{L}_{cdr}$ in Eq. (11), the Gaussian parameters $\mu_{i,j}$ and $\sigma_{i,j}$ are initialized from the statistical analysis of clear images and jointly optimized during training, allowing the model to adaptively refine the target distribution. To ensure stable training, we follow common practice by incorporating standard numerical safeguards (e.g., adding a small epsilon to denominators or restricting numerical ranges) into our loss functions; we omit further elaboration in the main text for brevity.

Inspired by [54], we also employ real-world clear images as additional inputs to our network architecture during training, which significantly improves training stability and convergence behavior. This approach provides stronger physical constraints and helps the model learn more robust feature representations for haze removal. To ensure fair comparison, all competing methods in our experiments are evaluated using either their publicly available pre-trained weights or our re-implementations following the original papers.

D Experiment Results

D.1 Progressive Domain Analysis of CIM

To systematically evaluate the contribution of each domain in our CIM framework, we conduct a progressive domain ablation study. We start by removing all domain-specific extractors and gradually reintroduce them one by one, resulting in variants CIM-D$_{(1)}$ to CIM-D$_{(5)}$, as illustrated in Tab. 3.

Table 3. Progressive domain composition of CIM-D variants.

Variant	Φ_1	Φ_2	Φ_3	Φ_4	Φ_5
CIM-D$_{(1)}$	✓	×	×	×	×
CIM-D$_{(2)}$	✓	✓	×	×	×
CIM-D$_{(3)}$	✓	✓	✓	×	×
CIM-D$_{(4)}$	✓	✓	✓	✓	×
CIM-D$_{(5)}$	✓	✓	✓	✓	✓

Qualitative Analysis: The visual comparisons presented in Fig. 11 reveal a progressive improvement across model variants: while CIM-D$_{(1)}$ demonstrates basic dehazing capability, it suffers from blurred texture details and residual haze in distant regions; CIM-D$_{(2)}$ significantly enhances high-frequency details yet occasionally amplifies noise in certain scenarios and introduces uneven color distribution in some areas; CIM-D$_{(3)}$ improves structural consistency across large regions and enhances long-range dependency modeling, though severe detail blurring persists in far-field areas; CIM-D$_{(4)}$ brings improved naturalness and better handles complex haze distributions through generative priors, yet introduces certain artifacts in sky regions; ultimately, CIM-D$_{(5)}$ achieves the optimal balance between detail preservation, color fidelity, and structural consistency.

Quantitative Analysis: We evaluate the CIM-D variants on the SOTS dataset and present the performance metrics in Figure 12. The results reveal a diminishing marginal improvement trend, where the performance gain from CIM-D$_{(1)}$ to CIM-D$_{(3)}$ is substantial, while improvements from CIM-D$_{(4)}$ to CIM-D$_{(5)}$ become more modest. This pattern suggests that the core haze semantics are effectively captured by the spatial, frequency, and non-local domains, while diffusion and compressed-sensing domains provide complementary refinements. Despite the diminishing returns, the complete CIM-D$_{(5)}$ framework achieves the best overall performance, justifying the inclusion of all five domains for robust dehazing across diverse scenarios. Considering the trade-off between computational complexity and performance gains as more domains are incorporated, we strategically adopt these five domains to maintain an optimal balance.

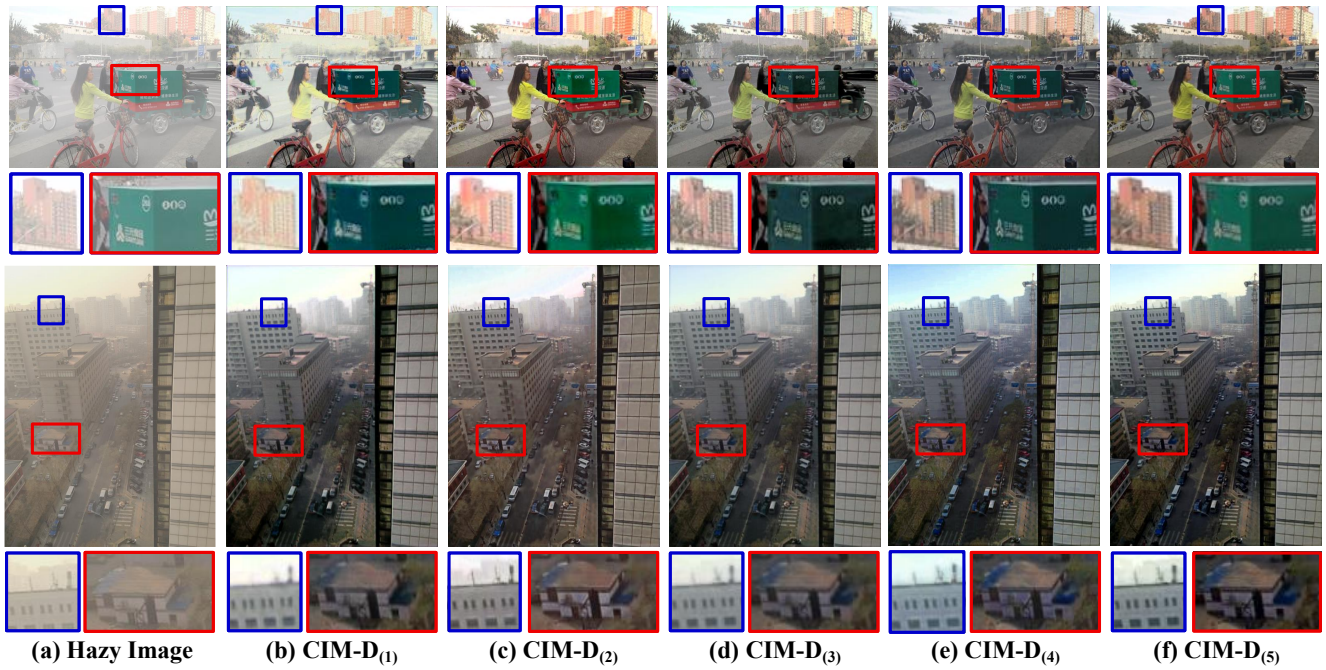

Figure 11. Visual comparison of CIM-D variants.

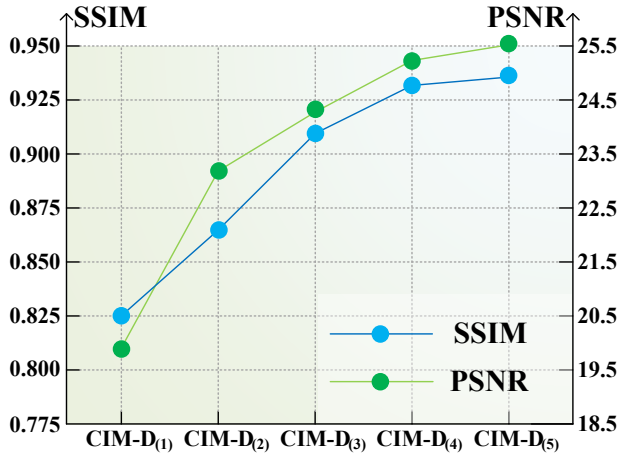

Figure 12. Quantitative scores between different variants of CIM-D on SOTS.

D.2 Further Limitation Analysis

Our method exhibits two limitations: halo artifacts near depth edges and over-smoothing in distant regions, as illustrated in Fig. 13. These issues can be attributed to two main factors: (1) the limited receptive field of the CNN-based decoupler, which hinders the recovery of fine texture details under severe degradation. This limitation is particularly evident near depth discontinuities, where abrupt transmission changes cannot be fully captured by local convolutions. (2) the global formulation of our physical constraints, which tends to over-smooth fine local discontinuities. In low-contrast distant regions, the strong regularization imposed by the ASM-based losses suppresses high-frequency details, leading to loss of texture. Future work will focus on developing edge-aware physical constraints and more powerful networks for chromatic separation.

D.3 Generalization Analysis

Additionally, our method shows robustness in nighttime haze with mild non-uniformity, but may exhibit artifacts near strong point light sources, as shown in Fig. 13. This is likely due to the inherent assumption of global atmospheric light uniformity in the ASM, which is violated by localized intense illumination. Future extensions could incorporate spatially varying atmospheric light estimation to handle such challenging scenarios.

D.4 Additional Visual Results

We provide extended visual comparisons to evaluate the effectiveness of our CIM-D framework against seven state-of-the-art dehazing methods, including IDE [27], C2P [58], KA-Net [18], PTTD [6], FSDGN [57], IPC [19], and UCL [54]. Fig. 14 illustrates the results on four challenging scenes with varying haze densities and degrees of color entanglement, while Fig. 15 presents additional visual comparisons across a wider range of scenarios. Our approach consistently delivers superior dehazing outcomes, characterized by more thorough haze removal, higher color fidelity, and better preservation of structural details. These

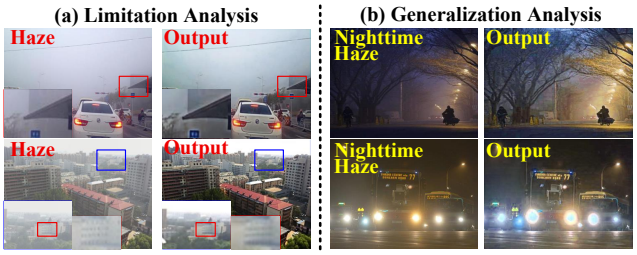

Figure 13. Limitation and generation analysis. **Zoom in for best view.**

extensive comparisons underscore the effectiveness of our cross-domain learning strategy and spectral balancing technique in handling diverse hazy conditions.

E. Broader Impacts

Our CIM-D algorithm demonstrates strong performance on real-world datasets by achieving unified physical consensus and effective color-aware disentanglement. This capability enables robust dehazing under diverse challenging conditions, including complex haze distributions, colored atmospheric effects (such as yellow sand haze and blue/purple haze from specific particle scattering), and standard white haze. The method's strong generalization across varied haze types ensures reliable performance in adverse imaging conditions, benefiting downstream applications like autonomous driving, aerial photography, surveillance, and computational photography. By advancing physically-grounded image restoration, we believe our work will have positive impacts on both academic research and industrial applications.

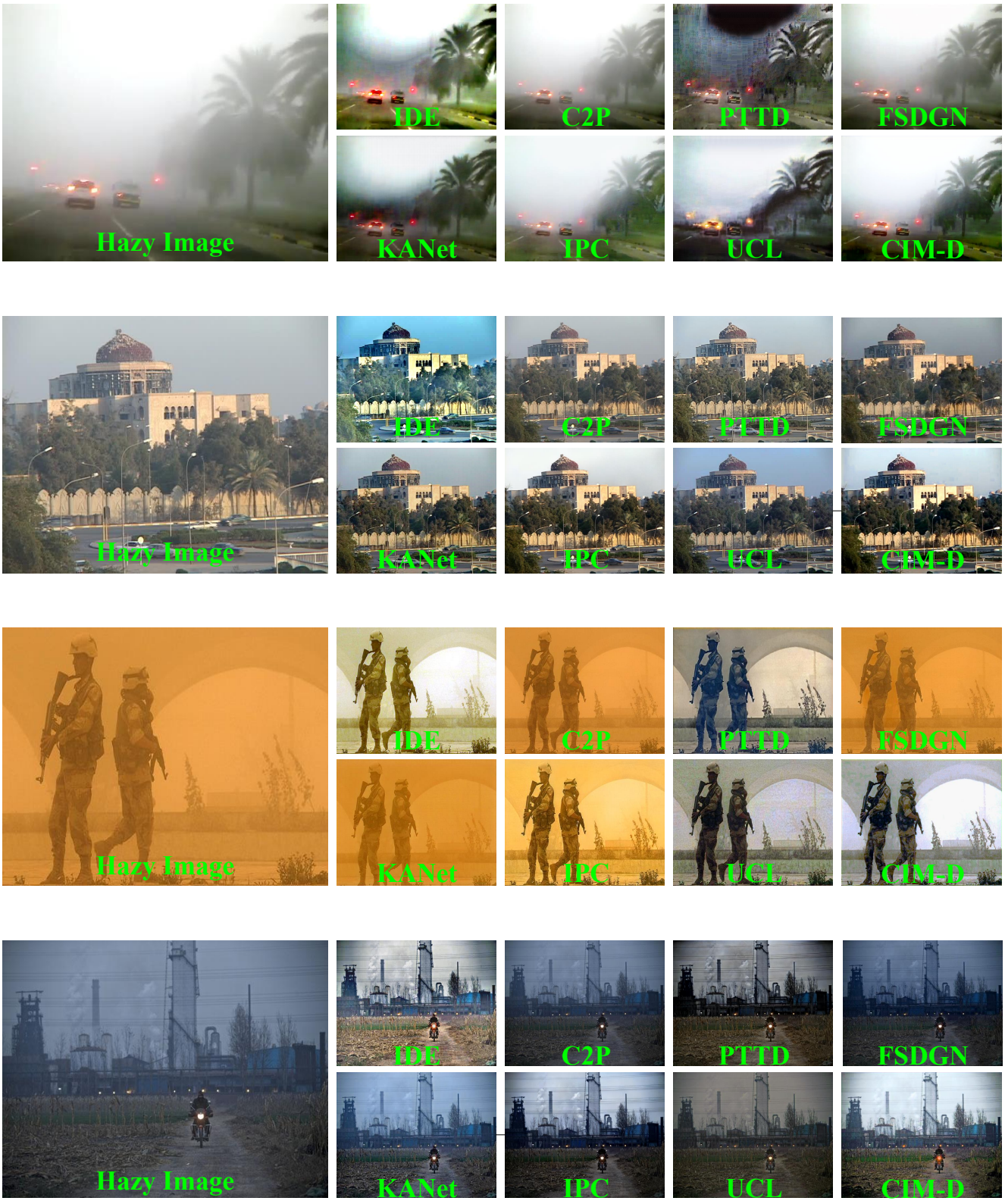

Figure 14. Visual comparison on challenging hazy scenes with varying concentrations and color shifts. **Zoom in for best view.**

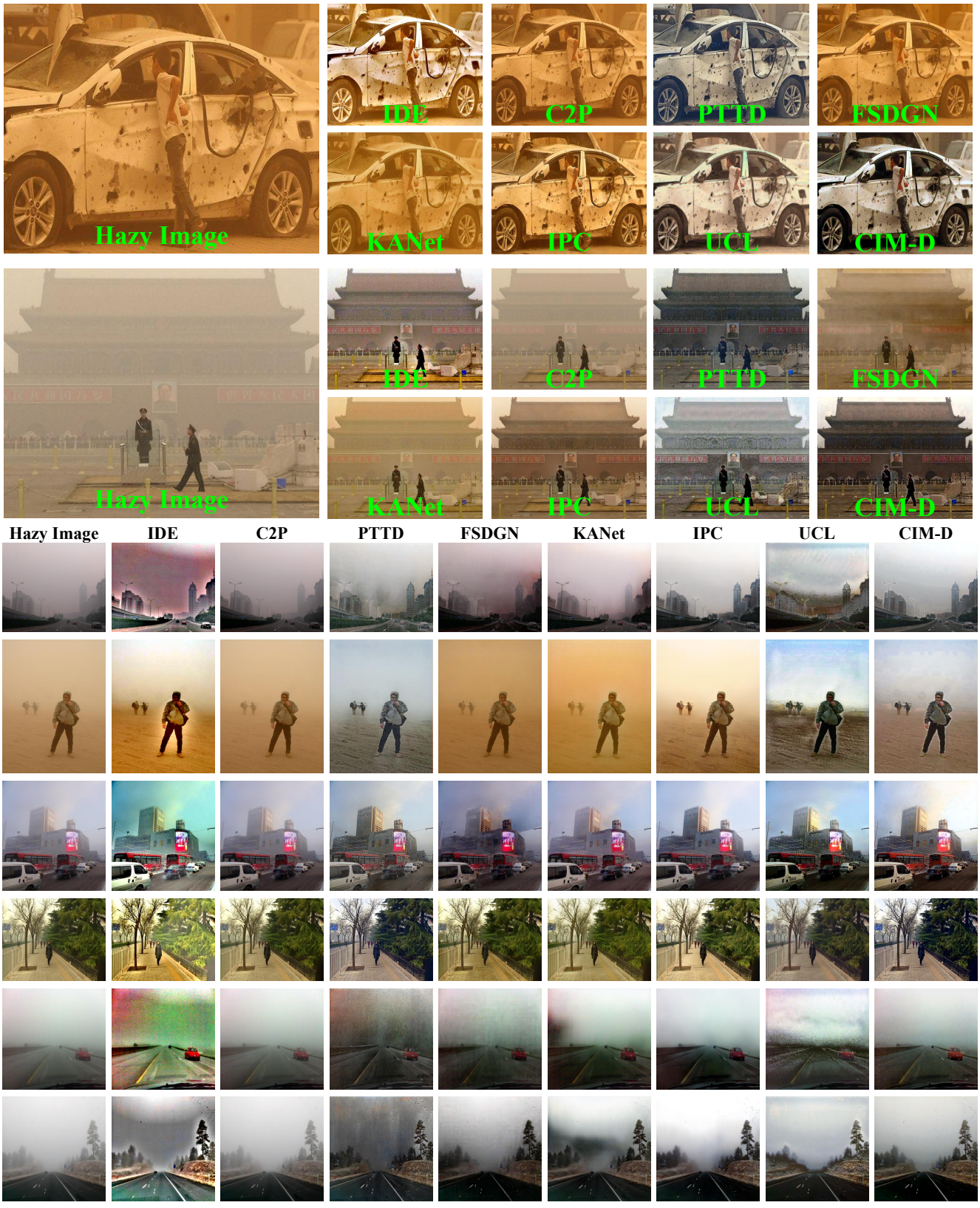

Figure 15. Extended visual comparisons across diverse real-world scenes. **Zoom in for best view.**

References

[1] Mahmoud Afifi and Michael S Brown. Deep white-balance editing. In *CVPR*, pages 1394–1403, 2020. 4

[2] Mahmoud Afifi, Brian Price, Scott Cohen, and Michael S Brown. When color constancy goes wrong: Correcting improperly white-balanced images. In *CVPR*, pages 1535–1544, 2019. 5

[3] Mohamed Ishmael Belghazi, Aristide Baratin, Sai Rajeshwar, Sherjil Ozair, Yoshua Bengio, Aaron Courville, and Devon Hjelm. Mutual information neural estimation. In *ICML*, pages 531–540. PMLR, 2018. 6

[4] Dana Berman, Tali Treibitz, and Shai Avidan. Non-local image dehazing. In *CVPR*, pages 1674–1682, 2016. 1, 2

[5] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE TIP*, 25(11):5187–5198, 2016. 2

[6] Zixuan Chen, Zewei He, Ziqian Lu, Xuecheng Sun, and Zhe-Ming Lu. Prompt-based test-time real image dehazing: A novel pipeline. In *ECCV*, pages 432–449, Cham, 2024. Springer Nature Switzerland. 1, 7, 6

[7] Zixuan Chen, Zewei He, and Zhe-Ming Lu. Dea-net: Single image dehazing based on detail-enhanced convolution and content-guided attention. *IEEE TIP*, 33:1002–1015, 2024. 1

[8] Lu Chi, Borui Jiang, and Yadong Mu. Fast fourier convolution. In *NIPS*, pages 4479–4488, Red Hook, NY, USA, 2020. Curran Associates Inc. 3

[9] Zewen Chi, Li Dong, Furu Wei, Nan Yang, Saksham Singhal, Shuming Wang, Kaitao Song, Changliang Mao, Lingxiao Liu, Heyan Huang, Ming Zhou, and Yue Zhang. XLM-E: Cross-lingual language model pre-training via ELECTRA. In *ACL*, pages 6170–6182, Dublin, Ireland, 2022. Association for Computational Linguistics. 5

[10] Lark Kwon Choi, Jaehee You, and Alan C. Bovik. Referenceless prediction of perceptual fog density and perceptual image defogging. *IEEE TIP*, 24(11):3888–3901, 2015. 7

[11] Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Image restoration via frequency selection. *IEEE TPAMI*, 46 (2):1093–1108, 2024. 1

[12] Yuning Cui, Wenqi Ren, and Alois Knoll. Omni-kernel network for image restoration. In *AAAI*, pages 1426–1434, 2024. 1

[13] Wei Dong, Han Zhou, Ruiyi Wang, Xiaohong Liu, Guangtao Zhai, and Jun Chen. Dehazedct: Towards effective non-homogeneous dehazing via deformable convolutional transformer. In *CVPRW*, pages 6405–6414, 2024. 2

[14] Alexey Dosovitskiy and Thomas Brox. Inverting visual representations with convolutional networks. In *CVPR*, pages 4829–4837, Las Vegas, NV, USA, 2016. 4

[15] Junkai Fan, Jiangwei Weng, Kun Wang, Yijun Yang, Jianjun Qian, Jun Li, and Jian Yang. Driving-video dehazing with non-aligned regularization for safety assistance. In *IEEE CVPR*, pages 26109–26119, 2024. 1

[16] Chengyu Fang, Chunming He, Fengyang Xiao, Yulun Zhang, Longxiang Tang, Yuelin Zhang, Kai Li, and Xiu Li. Real-world image dehazing with coherence-based pseudo labeling and cooperative unfolding network. In *NeurIPS*, pages 97859–97883. Curran Associates, Inc., 2024. 1

[17] Wenxuan Fang, JunKai Fan, Yu Zheng, Jiangwei Weng, Ying Tai, and Jun Li. Guided real image dehazing using ycbcr color space. In *AAAI*, pages 2906–2914, 2025. 7

[18] Yuxin Feng, Long Ma, Xiaozhe Meng, Fan Zhou, Risheng Liu, and Zhuo Su. Advancing real-world image dehazing: Perspective, modules, and training. *IEEE TPAMI*, 46(12): 9303–9320, 2024. 1, 7, 6

[19] Jiayi Fu, Siyu Liu, Zikun Liu, Chun-Le Guo, Hyunhee Park, Ruiqi Wu, Guoqing Wang, and Chongyi Li. Iterative predictor-critic code decoding for real-world image dehazing. In *CVPR*, pages 12700–12709, 2025. 1, 7, 6

[20] Ashwinkumar Ganesan, Francis Ferraro, and Tim Oates. Learning a reversible embedding mapping using bi-directional manifold alignment. In *ACL-IJCNLP*, pages 3132–3139, Online, 2021. Association for Computational Linguistics. 3

[21] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. In *CVPR*, pages 1956–1963, 2009. 2

[22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, Las Vegas, NV, USA, 2016. 8

[23] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *NIPS*, pages 6840–6851, 2020. 4

[24] Pratik Jawanpuria, Arjun Balgovind, Anoop Kunchukuttan, and Bamdev Mishra. Learning multilingual word embeddings in latent metric space: A geometric approach. *TACL*, 7:107–120, 2019. 3

[25] Zhi Jin, Yuwei Qiu, Kaihao Zhang, Hongdong Li, and Wenhan Luo. Mb-taylorformer v2: Improved multi-branch linear transformer expanded by taylor formula for image restoration. *IEEE TPAMI*, 47(7):5990–6005, 2025. 1

[26] Mingye Ju, Can Ding, Charles A. Guo, Wenqi Ren, and Dacheng Tao. Idrlp: Image dehazing using region line prior. *IEEE TIP*, 30:9043–9057, 2021. 3

[27] Mingye Ju, Can Ding, Wenqi Ren, Yi Yang, Dengyin Zhang, and Y. Jay Guo. Ide: Image dehazing and exposure using an enhanced atmospheric scattering model. *IEEE TIP*, 30: 2180–2192, 2021. 7, 6

[28] Mingye Ju, Can Ding, Wenqi Ren, and Yi Yang. Idbp: Image dehazing using blended priors including non-local, local, and global priors. *IEEE TCSVT*, 32(7):4867–4871, 2022. 2

[29] Mingye Ju, Chunming He, Can Ding, Wenqi Ren, Lin Zhang, and Kai-Kuang Ma. All-inclusive image enhancement for degraded images exhibiting low-frequency corruption. *IEEE TCSVT*, 35(1):838–856, 2025. 1, 2

[30] Kamalakar Kulkarni, Suhas Lohit, Pavan Turaga, Ronan Kerviche, and Amit Ashok. Reconnet: Non-iterative reconstruction of images from compressively sensed measurements. In *CVPR*, pages 449–458, 2016. 4

[31] N. Kwak and Chong-Ho Choi. Input feature selection by mutual information based on parzen window. *IEEE TPAMI*, 24(12):1667–1671, 2002. 6

[32] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE TIP*, 28(1):492–505, 2018. 7

[33] Boyun Li, Yuanbiao Gou, Shuhang Gu, Jerry Liu, Joey Tianyi Zhou, and Xi Peng. You only look yourself: Unsupervised and untrained single image dehazing neural network. *IJCV*, 129:1754 – 1767, 2021. 1, 2

[34] Xiaotian Li, Baojie Fan, Jiandong Tian, and Huijie Fan. Gafusion: Adaptive fusing lidar and camera with multiple guidance for 3d object detection. In *CVPR*, pages 21209–21218, 2024. 1

[35] Yu Li, Robby T. Tan, and Michael S. Brown. Nighttime haze removal with glow and multiple light colors. In *ICCV*, pages 226–234, 2015. 3

[36] Chengxu Liu, Lu Qi, Jinshan Pan, Xueming Qian, and Ming-Hsuan Yang. Frequency domain-based diffusion model for unpaired image dehazing. In *CVPR*, pages 7538–7547, 2025. 1

[37] Zhiyu Lyu, Yan Chen, and Yimin Hou. Mcpnet: Multi-space color correction and features prior fusion for single-image dehazing in non-homogeneous haze scenarios. *PR*, 150:110290, 2024. 2

[38] Long Ma, Yuxin Feng, Yan Zhang, Jinyuan Liu, Weimin Wang, Guang-Yong Chen, Chengpei Xu, and Zhuo Su. Coa: Towards real image dehazing via compression-and-adaptation. In *CVPR*, pages 11197–11206, 2025. 1

[39] Aravindh Mahendran and Andrea Vedaldi. Understanding deep image representations by inverting them. In *CVPR*, pages 5188–5196, Boston, MA, USA, 2015. 4

[40] Anish Mittal, Anush K. Moorthy, and Alan C. Bovik. No-reference image quality assessment in the spatial domain. *IEEE TIP*, 21(12):4695–4708, 2012. 7

[41] Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Making a "completely blind" image quality analyzer. *IEEE SPL*, 20(3):209–212, 2013. 7

[42] S.G. Narasimhan and S.K. Nayar. Contrast restoration of weather degraded images. *IEEE TPAMI*, 25(6):713–724, 2003. 1

[43] Dubok Park, David K. Han, and Hanseok Ko. Nighttime image dehazing with local atmospheric light and weighted entropy. In *IEEE ICIP (ICIP)*, pages 2261–2265, 2016. 3

[44] Yuwei Qiu, Kaihao Zhang, Chenxi Wang, Wenhan Luo, Hongdong Li, and Zhi Jin. Mb-taylorformer: Multi-branch efficient transformer expanded by taylor formula for image dehazing. In *ICCV*, pages 12756–12767, 2023. 1

[45] Hao Shen, Henghui Ding, Yulun Zhang, Zhong-Qiu Zhao, and Xudong Jiang. Spatial frequency modulation network for efficient image dehazing. *IEEE TIP*, 34:3982–3996, 2025. 1

[46] Tom Sherborne and Mirella Lapata. Meta-learning a cross-lingual manifold for semantic parsing. *TACL*, 11:49–67, 2023. 3

[47] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014. 3

[48] Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *IEEE TIP*, 32:1927–1941, 2023. 1

[49] Wei Sun, Hao Wang, Changhao Sun, Baolong Guo, Wenyan Jia, and Mingui Sun. Fast single image haze removal via local atmospheric light veil estimation. *Comput. Electr. Eng.*, 46:371–383, 2015. 3

[50] Fu-Jen Tsai, Yan-Tsung Peng, Yen-Yu Lin, and Chia-Wen Lin. Phatnet: A physics-guided haze transfer network for domain-adaptive real-world image dehazing. In *ICCV*, pages 5591–5600, 2025. 1

[51] Jing Wang, Songtao Wu, Zhiqiang Yuan, Qiang Tong, and Kuanhong Xu. Frequency compensated diffusion model for real-scene dehazing. *Neural Netw.*, 175:106281, 2024. 1, 2

[52] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. In *CVPR*, pages 7794–7803, 2018. 4

[53] Yaushian Wang, Ashley Wu, and Graham Neubig. English contrastive learning can learn universal cross-lingual sentence embeddings. In *EMNLP*, pages 9122–9133, Abu Dhabi, United Arab Emirates, 2022. Association for Computational Linguistics. 5

[54] Yongzhen Wang, Xuefeng Yan, Fu Lee Wang, Haoran Xie, Wenhan Yang, Xiao-Ping Zhang, Jing Qin, and Mingqiang Wei. Ucl-dehaze: Toward real-world image dehazing via unsupervised contrastive learning. *IEEE TIP*, 33:1361–1374, 2024. 1, 7, 5, 6

[55] Zhaofeng Wu, Xinyan Yu, Dani Yogatama, Jiasen Lu, and Yoon Kim. The semantic hub hypothesis: Language models share semantic representations across languages and modalities. In *ICLR*, pages 53705–53723, 2025. 2

[56] Zizheng Yang, Hu Yu, Bing Li, Jinghao Zhang, Jie Huang, and Feng Zhao. Unleashing the potential of the semantic latent space in diffusion models for image dehazing. In *ECCV*, pages 371–389, Cham, 2025. Springer Nature Switzerland. 1

[57] Hu Yu, Naishan Zheng, Man Zhou, Jie Huang, Zeyu Xiao, and Feng Zhao. Frequency and spatial dual guidance for image dehazing. In *ECCV*, pages 181–198, Cham, 2022. Springer Nature Switzerland. 1, 2, 7, 6

[58] Yu Zheng, Jiahui Zhan, Shengfeng He, Junyu Dong, and Yong Du. Curricular contrastive regularization for physics-aware single image dehazing. In *CVPR*, pages 5785–5794, 2023. 1, 7, 6

[59] Shihao Zhou, Jinshan Pan, Jinglei Shi, Duosheng Chen, Lishen Qu, and Jufeng Yang. Seeing the unseen: A frequency prompt guided transformer for image restoration. In *ECCV*, page 246–264, 2024. 1

[60] Qingsong Zhu, Jiaming Mai, and Ling Shao. A fast single image haze removal algorithm using color attenuation prior. *IEEE TIP*, 24(11):3522–3533, 2015. 2