This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Disentanglement-wise Image Dehazing through Cross-Domain Manifold Consensus

Tianyi Lyu[1] Mingye Ju[1*] Kai-Kuang Ma[2]

[1]Nanjing University of Posts and Telecommunications
[2]Nanjing University of Aeronautics and Astronautics

Abstract

Current dehazing methods face two intertwined challenges: (1) the misidentification of haze-related features due to domain-specific interference in both single-domain and empirically integrated multi-domain approaches, and (2) severe chromatic distortion caused by haze-induced color entanglement. To overcome these limitations, we propose a unified framework centered on a ***Cross-domain Invariant Manifold*** *(CIM), which aligns multi-domain features into a unified latent space through shared scattering semantics. The manifold is optimized via* ***consensus-density-driven contrastive learning****, effectively enhancing cross-domain consistency while eliminating domain-specific biases. Building upon this structured foundation, we further introduce a disentanglement-wise architecture, i.e.,* ***Physics-Guided HSV Decomposition Network****, that explicitly separates entangled color components to ensure robust color fidelity. Comprehensive experiments demonstrate that our CIM-D framework achieves state-of-the-art performance, effectively eliminating haze-induced color shifts and restoring natural scene appearance.*

1. Introduction

Image dehazing remains a challenging task in computer vision, crucial for applications such as autonomous driving [12] and recognition [28]. While deep learning [4, 5, 8, 9, 13, 15, 16, 21, 27, 36, 39, 42, 46] has advanced the field, existing methods face fundamental limitations in handling the complex physical nature of atmospheric degradation. Since image formation under haze follows the atmospheric scattering model (ASM) [35], the observed hazy image $\boldsymbol{I}$ can be mathematically expressed as:

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

where $\boldsymbol{J}$, $\boldsymbol{A}$, and t denote the clean scene radiance, global atmospheric light, and transmission map, respectively. To the best of our knowledge, most existing dehazing methods have been developed in diverse representation domains, including spatial [15, 36, 46], frequency [37, 47], non-local [2, 42], diffusion [44], and compressed-sensing [24, 31] domains. Despite their encouraging progress, these methods still suffer from two drawbacks.

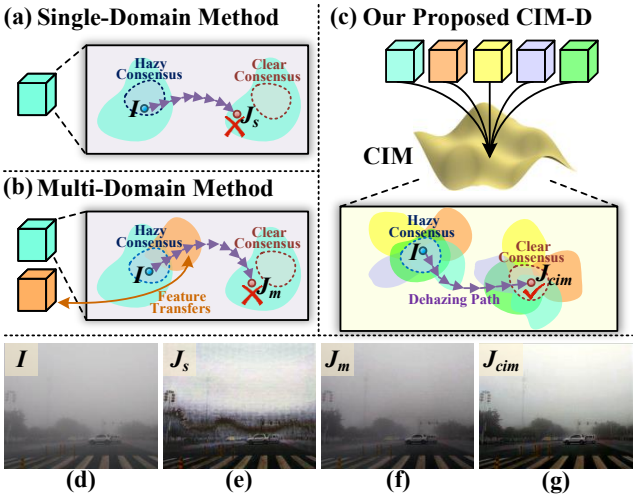

Figure 1. Comparative overview of dehazing methodologies. **(a)**-**(c)**: Single-domain, multi-domain, and our processing pipelines, with colored cubes representing features from different perceptual domains; **(d)**-**(f)**: Visual comparisons between hazy input $\boldsymbol{I}$, and corresponding dehazing result $\boldsymbol{J}_s$ [42], $\boldsymbol{J}_m$ [45], and $\boldsymbol{J}_{cim}$.

Drawback 1: These methods treat the corresponding domains as independent statistical spaces (see Fig. 1 (a)), which limits effective cross-domain interaction and the joint exploitation of complementary haze-related cues. Even though some methods [29, 40, 45] attempt to leverage such correlations, they still rely on manually designed feature transfers or domain-specific learned representations, which fail to fully account for the inherent scattering consistency across perceptual spaces (see Fig. 1 (b)). As a result, these approaches struggle to accurately identify haze characteristics, often misclassifying domain-specific features as haze

[1]Corresponding author.

attributes, leading to suboptimal restoration and noticeable color distortions, particularly under complex haze conditions, as illustrated in Fig. 1 (e) & (f).

Motivation for drawback 1: A well-established finding in multilingual natural language processing (NLP) is that large language models learn a shared semantic space across languages. This aligns with the "Semantic Hub Hypothesis" [43], which suggests that sentences conveying the same meaning, despite superficial differences in vocabulary or grammar, map to nearby points in a common semantic space. Drawing on this insight, we hypothesize: for the task of image dehazing, could the diverse representations of the same hazy scene across different perceptual domains (e.g., spatial, frequency) also contain a unified, domain-invariant "scattering semantic core"? ***In other words, despite the varied representations in different domains, is there a shared latent space that contains the common features necessary for dehazing?*** Based on above, we introduce the cross-domain invariant manifold (CIM), where features from different perceptual domains naturally converge and align, forming a compact representation that captures the intrinsic scattering properties of the given image.

Drawback 2: It is well-known that achieving photometrically accurate dehazing further requires addressing haze-induced chromatic distortions. Under challenging conditions such as sandstorms or dense fog, hazy images exhibit pronounced color perceptual coupling in HSV space—a marked departure from the relative channel independence observed in clear images (as detailed in Sec. 4). This haze-induced entanglement complicates color recovery by introducing spurious interdependencies between channels, thereby necessitating dedicated perceptual disentanglement to ensure faithful restoration.

Motivation for drawback 2: Inspired by the channel independence observed in clear images within the HSV space, we estimate haze-induced channel coupling by quantifying gradient direction consistency and design a learnable residual correction mechanism to progressively disentangle color channels on a continuous HSV representation, thereby achieving physically accurate color restoration.

Motivated by these facts, we propose the CIM-D framework, which extends the CIM by introducing a dedicated pathway for color perceptual disentangling. This ensures both structural and color fidelity in the dehazed image. Unlike sequential designs, CIM-D integrates these complementary objectives: the manifold provides domain-invariant physical model consensus to guide the decomposition process, while the HSV network's physical constraints—including channel decoupling regularization and spectrally-balanced scattering constraints—in turn, regularize manifold learning toward photometrically valid geometries. This interaction ensures that manifold convergence and chromatic decomposition mutually reinforce each other, creating a physically consistent restoration paradigm where physical consensus unification and color perceptual separation operate in harmony. Our main contributions are:

- **Cross-Domain Invariant Manifold**: A theoretically grounded representation space that unifies multi-domain features through atmospheric scattering physics, enabling consistent cross-domain feature alignment.
- **Physics-Guided HSV Decomposition**: A chromatic separation network that exploits intrinsic perceptual color space properties to decouple haze-induced chromatic entanglement, ensuring scattering fidelity via explicitly formulated physical constraints.
- **Synergistic Optimization**: A synergistic learning framework that concurrently optimizes the unification of physical consensus and chromatic separation, outperforming state-of-the-art methods under diverse haze conditions.

2. Related Works

Dehazing in Single Domain: Early dehazing approaches predominantly operate within a single perceptual space—typically the spatial domain—relying on either hand-crafted priors [2, 18, 23, 48] or end-to-end network architectures [3, 10]. While effective in constrained scenarios, these single-domain methods tend to conflate haze characteristics with intrinsic scene content (e.g., sky regions), resulting in over-enhanced artifacts. Such domain-specific constraints further limit generalization across varied imaging environments.

Dehazing via Multi-Domain Fusion: To mitigate single-domain bias, recent methods explore complementary information from multiple domains. For instance, Yu *et al.* [45] propose a unified framework that concurrently exploits frequency and spatial domains for haze removal, while Wang *et al.* [40] incorporate mid-/high-frequency components to enhance diffusion-based dehazing. However, most multi-domain methods rely on empirically designed transfers or domain-specific heads, failing to establish a physics-grounded inter-domain alignment; this often introduces domain conflicts and degraded color fidelity when confronted with shifting haze statistics.

Dehazing in Perceptual Color Spaces: Recent advancements have investigated perceptual color spaces to guide restoration processes. Li *et al.* [27] leverage luminance-saturation discrepancies as dehazing cues, and Lyu *et al.* [30] integrate priors from both HSV and YCbCr color spaces within a multi-color-space network. Yet, these methods typically employ HSV space as supplementary guidance rather than a primary representation space, and crucially, they fail to explicitly disentangle the haze-induced coupling among Hue, Saturation, and Value components. This oversight manifests as persistent color casts and unstable hue reproduction, especially in challenging scenarios

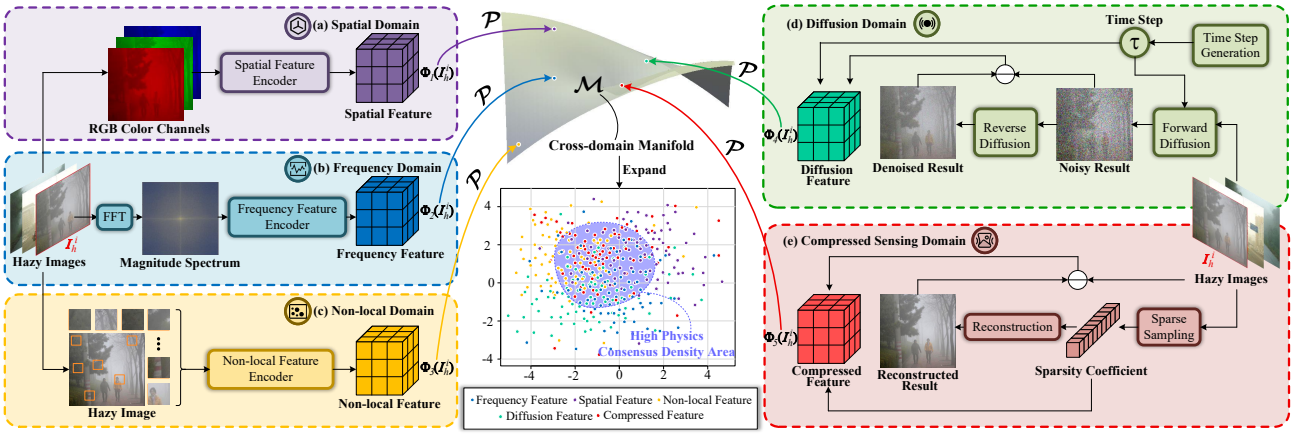

Figure 2. Schematic illustration of the cross-domain invariant manifold (CIM).

involving colored haze, like sandstorms. Our approach directly addresses these limitations by establishing physics-consistent cross-domain feature unification and implementing explicit disentanglement within the HSV color space.

3. Cross-Domain Invariant Manifold

Single-domain dehazing methods suffer from inherent domain ambiguity, where haze-induced degradations are often indistinguishable from intrinsic scene attributes (e.g. low contrast in the spatial domain may arise from haze attenuation or intrinsic surface reflectance property). Such ambiguity limits feature discriminability and hinders generalization under complex scenarios.

To address this limitation, we revisit the fundamental physics of haze formation. Although haze presents differently across perceptual domains, the resulting degradations all stem from the same atmospheric scattering process (Eq. 1). Inspired by manifold learning in multilingual NLP, where sentences sharing the same semantics despite different linguistic forms are projected onto a common low-dimensional manifold [17, 20, 38], we hypothesize that scattering features from different domains may also reside in a shared manifold, termed the cross-domain invariant manifold (CIM). The theoretical justification is provided in Supp. A.

Building upon this hypothesis, we realize the CIM as a unified latent space $\mathcal{M}$, where a domain translation network $\mathcal{P}$ is introduced to align multi-domain features by exploiting their shared scattering properties (see Fig. 2). Formally, for the i-th image $\boldsymbol{I}_s^i$ with state $s \in \{h, c\}$ (denoting hazy and clear), five domain-specific encoders $\Phi_k{}_{k=1}^{5}$ extract features from spatial, frequency, non-local, diffusion, and compressed-sensing domains (the rationale for selecting these five domains is provided in Supp. A). These features are then mapped to the unified manifold $\mathcal{M}$ through $\mathcal{P}$, which can be expressed as:

$$\boldsymbol{z}_s^{i,k} = \mathcal{P}(\Phi_k(\boldsymbol{I}_s^i)), \boldsymbol{z}_s^{i,k} \in \mathcal{M}. \tag{2}$$

To reveal the intrinsic geometry of $\mathcal{M}$, we must distinguish true physical consensus from features that are salient only in individual domains (i.e., domain-specific noise). Thus, we introduce a consensus density, which quantifies the agreement across all perceptual domains at any point on the manifold. For each state $s \in \{h, c\}$, the domain-wise feature density at $\boldsymbol{z} \in \mathcal{M}$ is expressed as:

$$\rho_s^k(\boldsymbol{z}) = \frac{1}{N_k} \sum_{i=1}^{N_k} e^{-\frac{|\boldsymbol{z} - \mathcal{P}(\Phi_k(\boldsymbol{I}_s^i))|^2}{2\sigma^2}}, \tag{3}$$

where σ controls the kernel bandwidth, N_k is the sample count for domain k. To prioritize cross-domain consensus over single-domain saliency, we define the overall consensus density as the geometric mean of domain-wise densities:

$$\rho_s(\boldsymbol{z}) = \left(\prod_{k=1}^{K} \rho_s^k(\boldsymbol{z}) \right)^{\frac{1}{K}}. \tag{4}$$

This formulation identifies regions with consistent cross-domain responses and establishes the core structure of the CIM through the density fields $\rho_h(\boldsymbol{z})$ and $\rho_c(\boldsymbol{z})$, whose peaks correspond to the modes of hazy and clear feature distributions while suppressing domain-specific deviations. Unlike conventional latent clustering based on semantic similarity, this organization arises from cross-domain physical consensus. Within the resulting self-organized manifold, haze-dominated patterns and clear-scene features progressively converge through our cross-domain consensus learning (Sec. 4.2 and Fig. 4(a)), forming a continuous scattering semantic field with smooth state transitions. This geometry emerges autonomously under the principles of atmospheric scattering and enables physically interpretable restoration through manifold traversal.

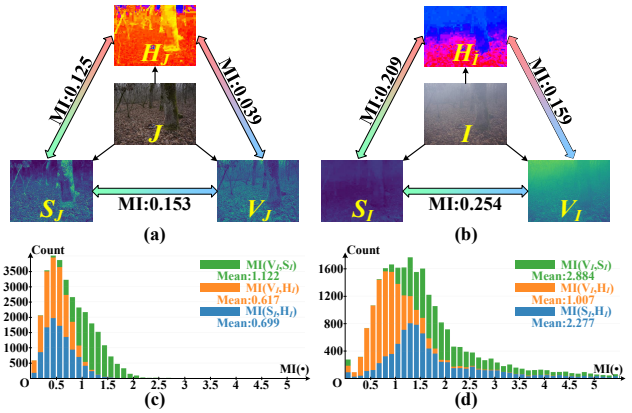

Figure 3. Analysis of haze-induced chromatic entanglement. (a),(b): Visualization of inter-channel entanglement in HSV space for clear and hazy image, respectively; (c),(d): Distribution histograms of entanglement degrees (measured by MI) computed from 10,000 clear images and 10,000 hazy images, demonstrating the significant increase in chromatic coupling caused by haze.

4. Method

While the CIM provides an effective scheme for extracting atmospheric-scattering-related features through consensus density, its operation on entangled RGB inputs inherently limits its ability to address severe *chromatic distortions* (**Drawback 2**). These distortions originate from *color perceptual entanglement*, whereby haze, especially colored haze, destroys the native independence of the H (Hue), S (Saturation), and V (Value) channels in HSV space by introducing strong nonlinear couplings. We quantify this effect via mutual information (MI), which reveals substantially increased statistical dependencies and correlated channel variations in hazy images, in contrast to the relatively independent variations observed in clear images (Fig. 3). This insight motivates us to explicitly reverse such haze induced coupling through a dedicated architectural design.

To this end, we propose the CIM-D framework, which integrates the CIM with a novel decomposition-wise network to establish a dual-perspective paradigm that: (1): the CIM resolves *feature ambiguity* by distilling domain-invariant representations through cross-domain consensus learning, identifying precise scattering characteristics; (2): the decomposition network resolves *chromatic distortion* by explicitly reversing haze-induced channel coupling through physics-guided disentanglement in HSV space. Following subsections will detail the network architecture, and the learning approach that ensure effective disentanglement.

4.1. Disentanglement-wise HSV Network

Direct use of raw HSV channels as network input is problematic due to the circular nature of the hue component H_I and its instability in low-saturation regions. To overcome this, we convert the HSV space into a continuous Cartesian representation:

$$\begin{cases} D_x = (S_I \cos(2\pi H_I) + 1)/2, \\ D_y = (S_I \sin(2\pi H_I) + 1)/2, \\ D_z = V_I, \end{cases} \tag{5}$$

where D_x, D_y, D_z constitute the stabilized input to the network, and $H_I, S_I, V_I \in [0, 1]$ denote the normalized HSV components.

As illustrated in Fig.4(b), the network follows a U-Net architecture, consisting of an encoder and a decoder, each containing three multi-scale convolution blocks (MConv blocks) and three residual decoupling blocks (see fig.5). Symmetric connections link the encoder and decoder, facilitating multi-scale representation learning while preserving spatial accuracy. Each residual decoupling block mitigates haze-induced channel dependencies through gradient-based correlation analysis. To enable explicit quantification and suppression of inter-channel coupling, we design an inversion pathway that leverages the established capability of CNN intermediate layers to reconstruct input-space representations [11, 32]. Within this pathway, high-dimensional features are compressed via a 1×1 convolution into three interpretable components D_x, D_y, D_z, which are subsequently mapped back into the original HSV space to facilitate perceptual disentanglement.

Gradients ∇H, ∇S, and ∇V are computed for each channel using the Sobel operator. The absolute cosine similarities $|\mathrm{Sim}(\nabla S, \nabla V)|$, $|\mathrm{Sim}(\nabla S, \nabla H)|$, and $|\mathrm{Sim}(\nabla H, \nabla V)|$ are denoted as $\mathcal{S}_\nabla(S, V)$, $\mathcal{S}_\nabla(S, H)$, and $\mathcal{S}_\nabla(H, V)$, respectively. These values quantify the haze-induced chromatic entanglement. To suppress such coupling, we apply the following adaptive correction:

$$\begin{aligned} H_{dec} &= H - \mathcal{F}\big(\mathrm{cat}[\mathcal{S}_\nabla(S, H), \mathcal{S}_\nabla(H, V)]\big) \cdot \mathcal{W}, \\ S_{dec} &= S - \mathcal{F}\big(\mathrm{cat}[\mathcal{S}_\nabla(S, H), \mathcal{S}_\nabla(S, V)]\big) \cdot \mathcal{W}, \\ V_{dec} &= V - \mathcal{F}\big(\mathrm{cat}[\mathcal{S}_\nabla(S, V), \mathcal{S}_\nabla(H, V)]\big) \cdot \mathcal{W}, \end{aligned} \tag{6}$$

where $\mathcal{F}$ denotes lightweight convolutional layers followed by ReLU-based activation functions and $\mathcal{W}$ is an adaptive weighting term that balances the contribution of each channel. The weight $\mathcal{W}$ is estimated via an intensity-weighted fusion of gradient magnitudes:

$$\mathcal{W} = \mathrm{Sigmoid}\big(\mathcal{F}(\mathrm{cat}[\|\nabla H\|, \|\nabla S\|, \|\nabla V\|])\big), \tag{7}$$

which suppresses unnecessary disentanglement in uninformative, low-gradient regions (e.g. sky area). The decoupled features $H_{dec}, S_{dec}, V_{dec}$ are then remapped into D_x, D_y, D_z and reintegrated via residual connections, progressively suppressing haze-induced chromatic entanglement throughout the network to produce the final disentangled result $\boldsymbol{J}$.

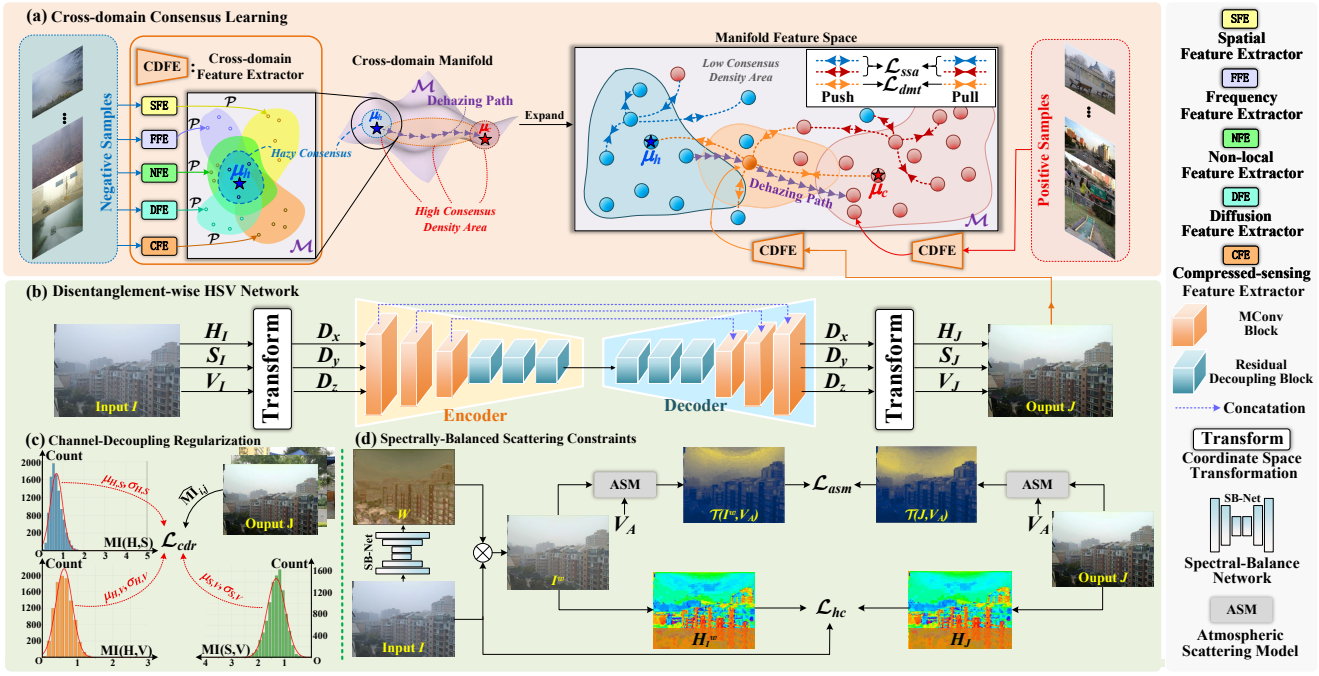

Figure 4. Overview of our proposed CIM-D.

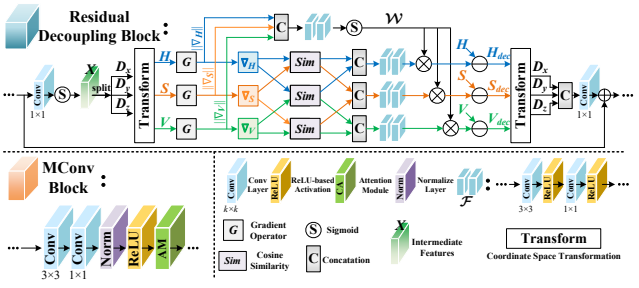

Figure 5. The structure of our multi-scale convolution block (MConv block) and residual decoupling block.

4.2. Cross-domain Consensus Learning

Scattering Semantics Alignment: Building upon the cross-domain invariant manifold, we develop a consensus density-driven contrastive learning framework to establish physical consensus across perceptual domains. Drawing inspiration from semantic alignment in multilingual NLP [6, 41], our approach leverages the density field ρ_s to guide optimization, ensuring the manifold intrinsically encodes scattering semantics:

$$\mathcal{L}_{ssa} = -\mathbb{E}_{(\boldsymbol{m}_s^i, \boldsymbol{m}_s^j)\sim p_{\text{pos}}} \left[\log \frac{\mathcal{D}(\boldsymbol{m}_s^i, \boldsymbol{m}_s^j)}{\mathcal{D}(\boldsymbol{m}_s^i, \boldsymbol{m}_s^j) + \mathcal{R}(\boldsymbol{m}_s^i)} \right], \tag{8}$$

where $\mathcal{D}(u, v) = e^{\text{Sim}(u,v)/\tau}$ measures exponential similarity, $\boldsymbol{m}_s^i$ and $\boldsymbol{m}_s^j$ form positive pairs (same state) from distribution p_{pos}, and $\mathcal{R}(\boldsymbol{m}_s^i) = \sum_{n=1}^{N} \mathcal{D}(\boldsymbol{m}_s^i, \boldsymbol{m}_-^n)$ aggregates negative similarities from low-density regions. The positive pair distribution incorporates manifold geometry:

$$p_{\text{pos}}(\boldsymbol{m}_s^i, \boldsymbol{m}_s^j) \propto \rho_s(\boldsymbol{m}_s^i)\rho_s(\boldsymbol{m}_s^j) e^{-\frac{\|\boldsymbol{m}_s^i - \boldsymbol{m}_s^j\|^2}{2\sigma_g^2}}, \tag{9}$$

where σ_g controls neighborhood similarity scale. The resulting density field ρ_s thus captures the intrinsic scattering semantics, with high-density regions corresponding to physically consistent hazy or clear states.

Dehazing via Manifold Traversal: To translate the learned manifold geometry into effective restoration guidance, we formulate a contrastive dehazing loss that directly leverages the intrinsic density field:

$$\mathcal{L}_{dmt} = -\log \frac{w \cdot \mathcal{D}(\boldsymbol{m}_d, \boldsymbol{\mu}_c)}{w \cdot [\mathcal{D}(\boldsymbol{m}_d, \boldsymbol{\mu}_c) + \mathcal{D}(\boldsymbol{m}_d, \boldsymbol{\mu}_h)] + (1-w) \cdot \mathcal{R}(\boldsymbol{m}_d)}, \tag{10}$$

where $\boldsymbol{m}_d$ denotes the manifold position of the dehazed result $\boldsymbol{J}$, $\boldsymbol{\mu}_c$ and $\boldsymbol{\mu}_h$ represent high consensus-density clear and hazy prototypes obtained through density-peak clustering, $w = 1 - e^{-(\rho_h(\boldsymbol{m}_d) + \rho_c(\boldsymbol{m}_d))/2}$ serves as the consensus-based weighting factor, and $\mathcal{R}(\boldsymbol{m}_d)$ aggregates negative similarities from low-density regions. This formulation conceptualizes dehazing as a manifold traversal process, where the restoration is guided toward high-density clear regions while simultaneously being repelled from both hazy prototypes and physically implausible low-density states, ensuring photometrically consistent results.

Table 1. Quantitative comparison between CIM-D and dehazing methods on different datasets. (Data highlighted in **bold** represents the best performance, while data in blue signifies the second-best.)

Dataset	Method	IDE	C2P	KA-Net	PTTD	FSDGN	IPC	UCL	CIM-D
	Metric	TIP 2021	CVPR 2023	TPAMI 2024	ECCV 2024	ECCV 2022	CVPR 2025	TIP2024	-
Raw2ah	PSNR ↑	16.06	17.26	16.45	15.48	16.23	13.23	16.47	**17.89**
	SSIM ↑	0.492	0.553	0.542	0.509	0.537	0.483	0.529	**0.585**
SOTS	PSNR ↑	21.25	**27.22**	19.48	23.12	25.09	21.52	23.33	25.51
	SSIM ↑	0.834	**0.955**	0.665	0.864	0.929	0.924	0.910	0.935
RTTS	FADE ↓	1.575	2.061	0.873	0.827	1.266	0.976	0.824	**0.795**
	Brisque ↓	37.55	47.42	19.88	17.43	24.26	**15.25**	21.51	16.32
	NIQE ↓	4.605	5.666	4.344	3.887	4.765	4.026	4.734	**3.844**
Runtime (s) ↓		0.652	0.575	0.088	0.438	0.356	3.291	0.128	**0.062**
Parameter (M) ↓		-	7.17	55.25	**2.02**	2.73	18.78	11.38	2.38

4.3. Physically-Grounded Constraints Loss

Channel-Decoupling Regularization: Natural clear images exhibit statistical independence among HSV channels, a property disrupted by haze-induced chromatic coupling. We statistically analyze 10,000 clear images and find that the mutual information between channel pairs follows Gaussian distributions. Based on this observation, we propose a channel-decoupling regularization formulated as:

$$\mathcal{L}_{cdr} = -\frac{1}{B}\sum_{b=1}^{B}\sum_{(i,j)} \log \mathcal{N}\left(\widehat{\mathrm{MI}}_{i,j}(\boldsymbol{J}_b); \mu_{i,j}, \sigma_{i,j}^2\right), \quad (11)$$

where B is the batch size, $\widehat{\mathrm{MI}}_{i,j}(\boldsymbol{J}_b)$ denotes the differentiable MI (e.g., [1, 25])estimate for the b-th dehazed image, (i,j) represents channel pairs $\in \{(H,S),(H,V),(S,V)\}$, $\mu_{i,j}$ and $\sigma_{i,j}$ are the mean and standard deviation of the mutual information between channel pair (i,j) computed from the statistical analysis, and $\mathcal{N}(u;\mu,\sigma^2)$ represents the Gaussian probability density function evaluated at u.

Spectrally-Balanced Scattering Constraints: To ensure photometric fidelity under the atmospheric scattering model (ASM) and reduce color distortion caused by haze, we introduce HSV-domain constraints derived from a spectrally-balanced scattering formulation. We employ a learnable spectral balance matrix $\boldsymbol{W}$ to enable adaptive color compensation through end-to-end optimization:

$$\boldsymbol{W}\boldsymbol{I} = \boldsymbol{W}\boldsymbol{J}\cdot t + \boldsymbol{W}\boldsymbol{A}\cdot(1-t), \quad (12)$$

$$\Longrightarrow \boldsymbol{I}^w = \boldsymbol{J}^w\cdot t + \boldsymbol{A}^w\cdot(1-t), \quad (13)$$

where $\boldsymbol{I}^w$, $\boldsymbol{J}^w$, and $\boldsymbol{A}^w$ denote the spectrally-balanced versions of the hazy image, scene radiance, and atmospheric light, respectively. Crucially, $\boldsymbol{W}$ evolves through the joint optimization with our decomposition network and physical constraints, ensuring that color adaptation emerges as a consequence of our architectural innovations rather than predefined corrections.

Mapping this adaptively balanced model into HSV (derivation in supp.B) yields two actionable constraints. First, for the value and saturation components:

$$\begin{aligned} V_{I^w} &= V_{J^w}\,t + V_{A^w}(1-t), \\ S_{I^w}V_{I^w} &= S_{J^w}V_{J^w}\,t + S_{A^w}V_{A^w}(1-t). \end{aligned} \quad (14)$$

Since $S_{A^w} \approx 0$ for spectrally-balanced atmospheric light, eliminating t produces an invariant ratio between the input and restored frames. We therefore define:

$$\mathcal{L}_{asm} = \mathbb{E}[\|\mathcal{T}(\boldsymbol{I}^w, V_A) - \mathcal{T}(\boldsymbol{J}, V_A)\|_2^2], \quad (15)$$

where $\mathcal{T}(\boldsymbol{u}, V_A) = S_u V_u/(V_A - V_u)$ is structure regularization operator after eliminating t. $\mathbb{E}$ denotes averaging over the image domain. This term maintains the physical constraint between saturation and value components during scattering for color-consistent dehazing.

Besides, hue remains stable in regions of low brightness and high saturation, as these areas experience minimal atmospheric scattering interference. We impose a hue-consistency loss on pixel pairs $(a,b) \in \mathcal{Q} = \{(\boldsymbol{I},\boldsymbol{I}^w),(\boldsymbol{I},\boldsymbol{J}),(\boldsymbol{J},\boldsymbol{I}^w)\}$ to penalize significant hue drift:

$$\mathcal{L}_{hc} = \frac{1}{3\mathrm{HW}}\sum_{(a,b)\in\mathcal{Q}} \left|\left|\Omega\cdot\sin(2\pi H_a - 2\pi H_b)\right|\right|_2^2, \quad (16)$$

where H and W are the height and width of the image. $\Omega = e^{-V_I/2S_I}$ is a weight matrix that down-weights the loss in regions where hue is unreliable.

In summary, the total loss of our CIM-D is as below:

$$\mathcal{L}_{total} = \lambda_1\,\mathcal{L}_{ssa} + \lambda_2\,\mathcal{L}_{dmt} + \lambda_3\,\mathcal{L}_{cdr} + \lambda_4\,(\mathcal{L}_{asm} + \mathcal{L}_{hc}), \quad (17)$$

where $\lambda_{1..4}$ balance scattering semantics alignment, manifold traversal dehazing, channel decoupling, and photometric consistency. Optimization under (17) establishes mutual reinforcement between manifold consensus and chromatic disentanglement.

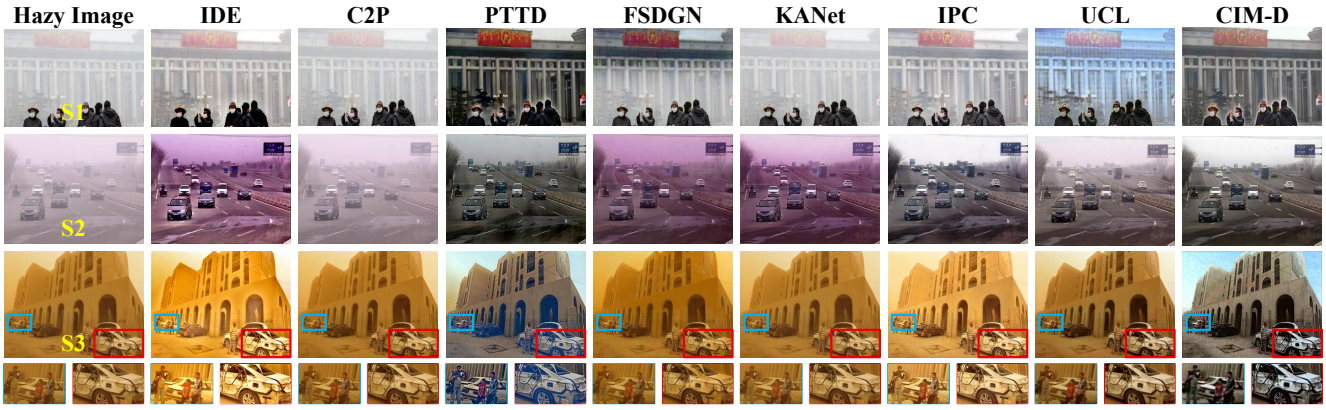

Figure 6. Visual comparison on real-world datasets. Zoom in for best view.

Figure 7. Visual comparison on synthetic datasets. Zoom in for best view.

5. Experiment

Datasets. In our experiments, the training data consist of 2,500 real-world hazy images selected from the RTTS [26] and URHI [26] datasets, and 1,800 haze-free images randomly chosen from OTS [26]. All hazy and clean samples are unpaired, ensuring that no explicit pixel-wise correspondence exists between the two domains. For evaluation, we employ both the synthetic benchmark Raw2ah [14], SOTS [26], and the real-world RTTS dataset to comprehensively assess generalization performance under diverse conditions.

Implementation Details. During the training procedure, we adopt the AdamW optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and a learning rate of 1×10^{-5}. The loss weights λ_1, λ_2, λ_3, and λ_4 are set to 0.1, 0.5, 0.05, and 1, respectively, in all experiments. For the contrastive module, the temperature coefficient is fixed at $\tau = 0.07$. The training process lasts for 100 epochs with a batch size of 16. The models were trained on two NVIDIA RTX 4090 GPUs, and all evaluations were performed on an NVIDIA RTX 3080Ti GPU.

5.1. Comparison with State-of-the-Art Methods

We compare the proposed CIM-D against seven state-of-the-art dehazing methods, including IDE [22], C2P [46], KA-Net [15], PTTD [4], FSDGN [45], IPC [16], and UCL [42]. Comprehensive evaluations are conducted across both synthetic and real-world benchmarks.

Quantitative Comparisons. We evaluate the proposed method using both full-reference metrics (PSNR, SSIM) on synthetic datasets and no-reference metrics (FADE [7], BRISQUE [33], NIQE[34]) on real-world data. As shown in Tab. 1, our method demonstrates competitive performance across various benchmarks. On the synthetic Raw2ah dataset, CIM-D achieves the best results in both PSNR and SSIM. For the SOTS dataset, while C2P attains the highest PSNR and SSIM, our method ranks second with 25.51 dB PSNR and 0.935 SSIM. On the real-world RTTS dataset, CIM-D obtains the best FADE and NIQE scores, and achieves the second-best BRISQUE value. Notably, our method exhibits superior efficiency with the fastest inference time (0.062s) and the second-lightest parameter count, highlighting its advantages for real-time applications.

Qualitative Comparisons. We conduct qualitative comparisons on both real-world and synthetic datasets to validate our method's effectiveness under diverse hazy conditions. From the RTTS dataset, we select three representative images with varying color cast levels: no cast (S1), mild cast (S2), and severe cast (S3). Visual results (Fig. 6) demonstrate that our method consistently removes color casts and recovers vivid details across all scenarios. In the challenging severe color cast case (S3), compared methods includ-

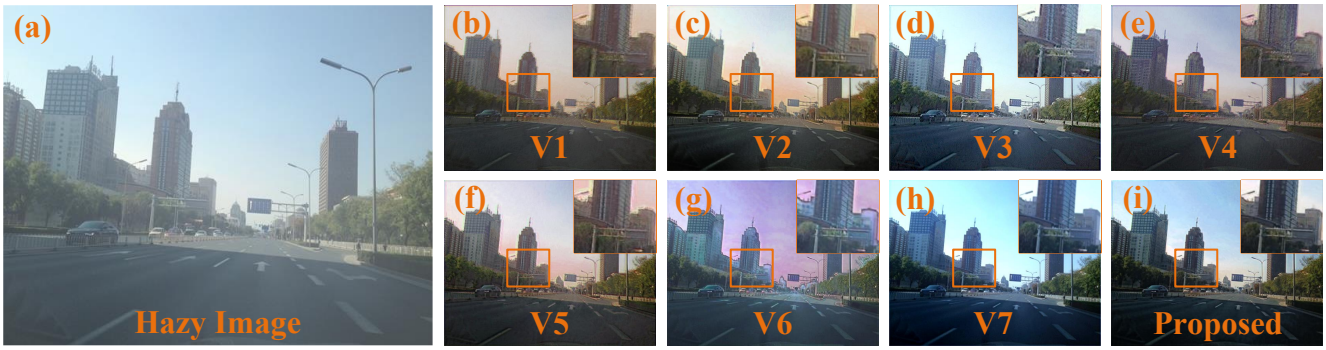

Figure 8. Ablation results of CIM-D framework, where (a) Hazy images; (b) V1: w/o $\mathcal{L}_{dmt}$; (c) V2: w/o $\mathcal{L}_{ssa}$; (d) V3: w/o $\mathcal{L}_{cdr}$; (e) V4: w/o $\mathcal{L}_{asm}$; (f) V5: w/o $\mathcal{L}_{hc}$; (g) V6: with raw HSV inputs; (h) V7: with standard residual blocks; (i) Our complete CIM-D.

Table 2. Ablation study results on SOTS datasets.

Variant	Ablated Component	PSNR↑	SSIM↑
V1	w/o $\mathcal{L}_{dmt}$	21.15	0.875
V2	w/o $\mathcal{L}_{ssa}$	22.89	0.891
V3	w/o $\mathcal{L}_{cdr}$	23.26	0.910
V4	w/o $\mathcal{L}_{asm}$	20.57	0.804
V5	w/o $\mathcal{L}_{hc}$	23.82	0.906
V6	Cartesian representation	18.95	0.878
V7	Decoupling blocks	24.18	0.919
CIM-D	**-**	**25.51**	**0.935**

ing IDE, C2P, FSDGN, KA-Net, IPC, and UCL fail to effectively eliminate chromatic shifts. Such color entanglement significantly compromises their dehazing performance. Although PTTD partially restores natural colors, it retains considerable haze and introduces unnatural blue artifacts. Benefiting from the physical consensus learned through our Cross-Domain Invariant Manifold and the specialized chromatic separation design, our approach consistently removes haze from all regions while producing visually pleasing results with natural color restoration. In addition, we select two hazy images from a synthetic dataset for further comparison Fig. 7). The results show that while other methods struggle with color casts, our approach effectively disentangles the chromatic distortions induced by haze and achieves the most visually appealing dehazing outcome.

5.2. Ablation Studies

We conduct comprehensive ablation studies to validate the contribution of each component in our framework. All variants are evaluated on the SOTS dataset, with quantitative results presented in Tab. 2.

Loss Function Analysis. We systematically ablate each loss component while keeping others intact. V1 removes the manifold traversal dehazing loss $\mathcal{L}_{\text{dmt}}$, resulting in compromised structural coherence due to lack of explicit guidance toward clear regions. V2 excludes the scattering semantics alignment loss $\mathcal{L}_{\text{ssa}}$, leading to reduced feature discriminability from missing cross-domain consensus. V3 eliminates the channel decoupling regularization $\mathcal{L}_{\text{cdr}}$, causing persistent chromatic entanglement and color distortion. V4 removes the spectrally-balanced scattering constraint $\mathcal{L}_{\text{asm}}$, resulting in significant detail loss. V5 omits the hue consistency loss $\mathcal{L}_{\text{hc}}$, producing noticeable hue drift in sky regions.

Architecture Analysis. We further examine two architectural variants. V6 replaces our stable Cartesian representation with raw HSV inputs, suffering from severe instability due to the circular nature of hue channels and poor saturation handling. V7 substitutes our residual decoupling blocks with standard residual modules [19], yielding noticeable chromatic coupling despite moderate structural recovery, confirming the necessity of explicit gradient-based disentanglement. In contrast, our complete CIM-D achieves superior performance across all metrics, producing photo-realistic results with natural color reproduction and fine detail preservation. For more experimental results and more ablation study on the perceptual domains (refer to supp.D).

6. Discussion

Conclusion: We have presented CIM-D, a unified dehazing framework that addresses both structural and chromatic degradation through cross-domain consensus learning and physics-guided color disentanglement. By introducing the Cross-Domain Invariant Manifold for unified feature representation and the HSV Decomposition Network for explicit chromatic separation, our method achieves state-of-the-art performance across multiple benchmarks while ensuring photometrically accurate restoration.

Limitations: While achieving strong performance on standard benchmarks, our method occasionally produces halo artifacts near depth edges and may oversmooth complex multi-layer haze scenes due to the architecture design and physical constraints.(Detailed in Supp.D) Future work will focus on enhancing structural recovery while maintaining physical consistency.

7. Acknowledgements

This work was supported by National Natural Science Foundation of China (62471253), Natural Science Research of Jiangsu Higher Education Institutions of China (23KJB520027), and Natural Science Foundation of Jiangsu Province (BK20251873).

References

[1] 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

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

[3] 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

[4] 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

[5] 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

[6] 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

[7] 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

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

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

[10] 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

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

[12] 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

[13] 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

[14] 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

[15] 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

[16] 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

[17] 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

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

[19] 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

[20] 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

[21] 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

[22] 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

[23] 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

[24] 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

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

[26] 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

[27] 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

[28] 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

[29] 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

[30] 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

[31] 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

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

[33] 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

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

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

[36] 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

[37] 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

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

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

[40] 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

[41] 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

[42] 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

[43] 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

[44] 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

[45] 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

[46] 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

[47] 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

[48] 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