IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 25 May 2026, accepted 11 June 2026, date of publication 17 June 2026, date of current version 30 June 2026.

Digital Object Identifier 10.1109/ACCESS.2026.3704688

RESEARCH ARTICLE

Semi-PhysHDNet: Orthogonal Content Haze Disentanglement and Uncertainty-Calibrated Learning for Dense Dehazing

HIRA KHAN[1], **(Graduate Student Member, IEEE), AND SUNG WON KIM**[2], **(Member, IEEE)**
[1]Department of Information and Communication Engineering, Yeungnam University, Gyeongsan-si 38541, Republic of Korea
[2]School of Computer Science and Engineering, Yeungnam University, Gyeongsan-si 38541, Republic of Korea

Corresponding author: Sung Won Kim (swon@yu.ac.kr)

This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant RS-2021-NR060120, in part by the BK21 FOUR of NRF funded by the Ministry of Education under Grant 2120251015517, and in part by the NRF Grant funded by Korean Government through MSIT under Grant RS-2022-NR069161.

ABSTRACT Single-image dehazing plays an important role in adverse-weather vision, where visibility degradation affects both human perception and downstream scene understanding. Dense and non-uniform haze remains a stimulating degradation for image restoration since it simultaneously obfuscates scene structure, distorts color, and weakens the reliability of purely supervised models under real-world domain shift. This paper presents Semi-PhysHDNet, a dehazing framework built on orthogonal content-haze disentanglement and uncertainty-calibrated semi-supervised learning. The network integrates a haze-aware hierarchical Swin encoder with subband-selective frequency refinement, a latent disentanglement stage that separates scene content from haze-related degradation, and a physics-constrained atmospheric reconstruction branch for transmission and atmospheric light estimation. To advance adaptation to real hazy data, a heteroscedastic uncertainty mechanism is presented to down-weight unreliable regions during semi-supervised optimization. This joint design enables more stable restoration under dense and spatially varying haze conditions. Experimental results on synthetic and real-world benchmarks demonstrate strong and consistent performance, including 35.25/0.988 on RESIDE Outdoor, 23.98/0.867 on O-HAZE, and 20.87/0.789 on NH-HAZE. The proposed method also improves downstream YOLOv8s detection on Foggy Cityscapes from 8.77 to 11.01 detections per image, indicating practical value beyond pixel-level restoration.

INDEX TERMS Image dehazing, dense haze, content-haze disentanglement, semi-supervised learning, uncertainty-calibrated learning, Swin transformer, atmospheric reconstruction.

I. INTRODUCTION

Image dehazing seeks to recover clear scene radiance from visibility-degraded observations and remains a fundamental problem in adverse-weather vision. Atmospheric scattering and haze attenuate scene radiance, weaken contrast, and blur structural cues, thereby degrading both image quality and the reliability of downstream tasks such as autonomous driving, surveillance, and remote sensing [1]. The impact extends beyond pixel-level restoration: visibility degradation under dense fog or haze directly compromises target detection and recognition in safety-critical applications, as evidenced by recent work on detection frameworks designed specifically for low-visibility foggy environments [2].

Early dehazing methods are largely built on hand-crafted physical priors. A representative example is the Dark Channel Prior (DCP) [3], which estimates transmission from local intensity statistics in haze-free outdoor images. Although such methods are physically interpretable and effective under mild or relatively uniform haze, their assumptions often break down in dense, spatially varying, or scene-dependent conditions, especially in bright regions, sky areas, and complex indoor environments.

The associate editor coordinating the review of this manuscript and approving it for publication was Jiachen Yang.

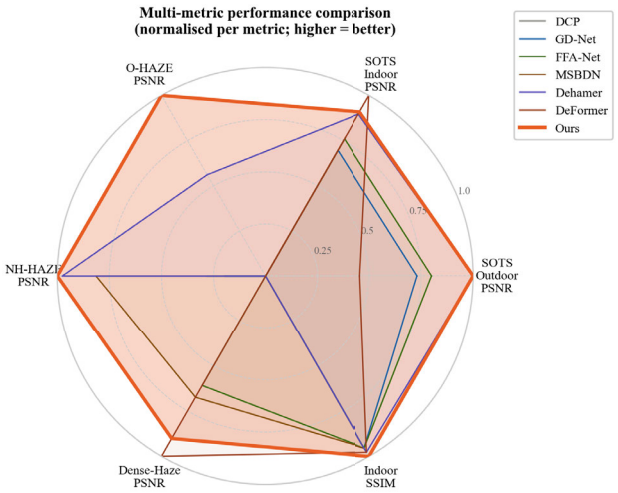

FIGURE 1. Multi-metric radar comparison against state-of-the-art methods.

FIGURE 2. Representative dehazing results on NH-HAZE: (a) hazy inputs and (b) outputs produced by the proposed method.

With the rise of deep learning, convolutional dehazing models [4], [5], [6], [7], [8], [9], [10], have shown clear gains by learning direct haze-to-clean mappings from paired synthetic data. These methods achieve strong performance on standard benchmarks, yet their robustness in real scenes is still constrained by the domain gap between synthetic haze generation and real atmospheric degradation. In practice, models trained only on synthetic pairs may leave residual haze, distort color, or produce unstable enhancement when applied to real-world images.

More recent transformer-based approaches, including Dehamer [11], DehazeFormer [12], Vision Transformer (ViT) [13], and USCFormer [14] further improve restoration quality by modeling long-range dependencies and global contextual information. However, most of these methods still formulate dehazing as a direct image restoration problem, without explicitly separating haze-related degradation from scene-preserving content or imposing sufficient physical constraints on the reconstruction process. This limitation becomes more evident on challenging real-world datasets such as NH-HAZE [15], where haze is dense, non-uniform, and difficult to represent using conventional paired training settings.

These observations point to several limitations that motivate the proposed framework. First, most existing dehazing networks operate primarily in the spatial domain and do not explicitly exploit frequency-selective representations, even though haze suppression and detail recovery affect different spectral components in different ways. As a result, fine structures and residual haze can remain insufficiently separated in deep features. Second, latent representations are typically learned in an implicit manner, without explicitly decoupling scene-preserving content from haze-related degradation. This entanglement can reduce reconstruction quality and weaken generalization when the haze characteristics at test time differ from those seen during training. Third, adaptation to unpaired hazy images is often driven by uniform self-consistency objectives, which assume equal reliability across all pixels. In dense and spatially varying haze, however, heavily degraded regions are inherently more uncertain, and treating them identically during optimization can introduce unstable supervision.

Fig. 1 represents multi-metric radar comparison against state-of-the-art methods. Each axis is independently normalized; a larger covered area indicates stronger overall performance. The proposed method achieves the broadest coverage across all five benchmarks, demonstrating balanced generalization rather than specialization on a single dataset. Fig. 2 presents representative examples from NH-HAZE, where dense and spatially varying haze severely degrades scene visibility. As observed, heavy atmospheric scattering suppresses background content, weakens structural boundaries, and introduces noticeable color attenuation. The corresponding outputs demonstrate the target restoration behavior required in challenging real-world dehazing, namely improved visibility, clearer recovery of scene structures, and better preservation of natural color appearance under non-uniform haze conditions.

To address these limitations, we propose *Semi-PhysHDNet*, a unified dehazing framework whose novelty lies in three tightly coupled design decisions that have not been jointly explored in prior work: (i) explicit orthogonal disentanglement of scene content and haze degradation in latent space, enforced through a correlation-based regularization loss rather than implicit feature separation; (ii) physics-guided iterative transmission refinement grounded in dark channel priors, providing an explicit physical correction signal at each refinement step; and (iii) heteroscedastic uncertainty-calibrated semi-supervised learning that

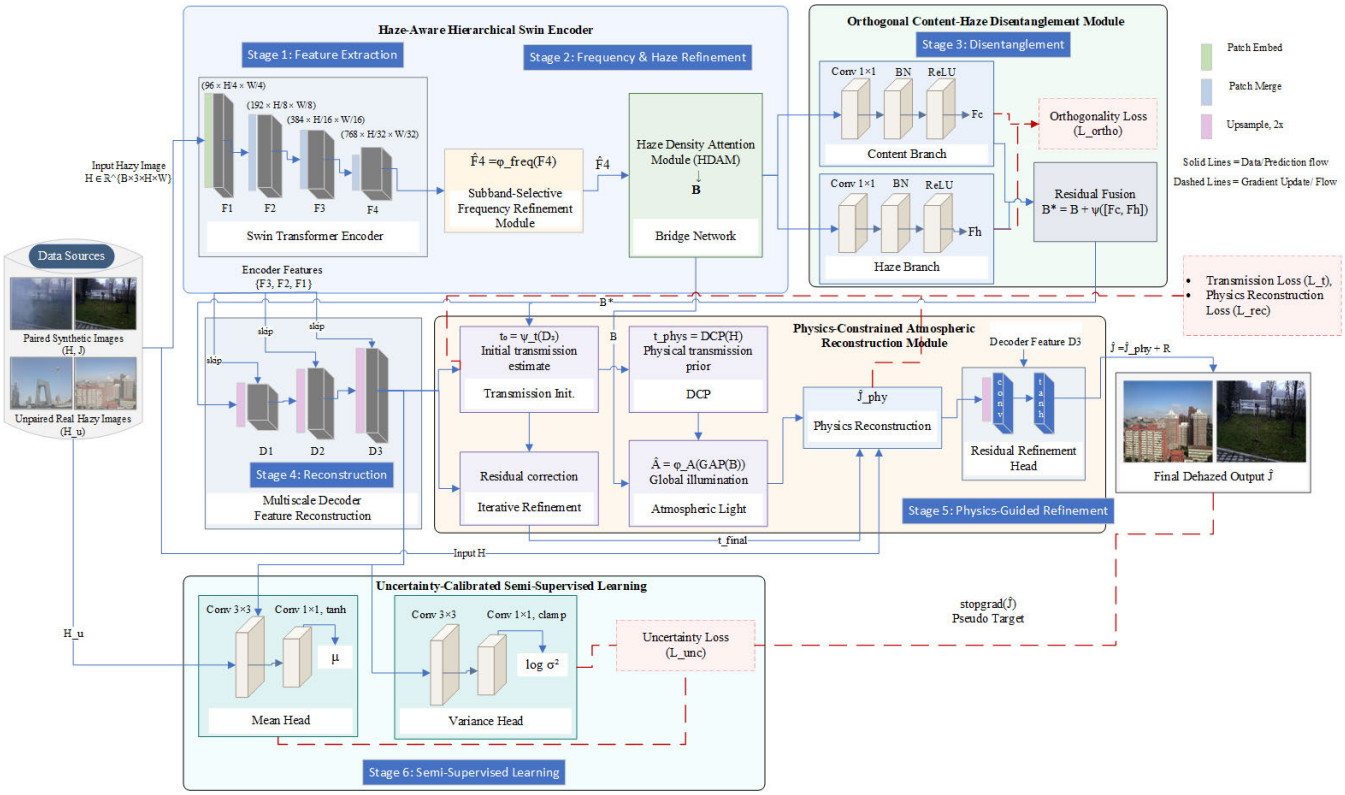

FIGURE 3. Overview of Semi-PhysHDNet. A hazy input is encoded into multi-scale features via a Swin-based backbone, followed by frequency-aware refinement (SSFRM) and haze-density modulation (HDAM). The refined latent is disentangled into content and haze components (OCHD) and progressively reconstructed (PMRD). A physics-constrained module (PCARM) estimates transmission and atmospheric light for final restoration, while an uncertainty-calibrated branch (UCSSL) enables robust semi-supervised learning on real-world data.

adaptively suppresses unreliable pseudo-supervision in spatially degraded regions, enabling stable real-domain adaptation without paired real-world data. Together, these components form a unified architecture trained under a joint supervised and semi-supervised objective, enabling robust restoration under dense and spatially varying real-world haze.

The main contributions of this work are summarized as follows:

- We propose a novel orthogonality-constrained content-haze disentanglement framework that explicitly separates haze-invariant structural information from degradation components in latent space, improving representation robustness under dense and spatially varying haze.
- We introduce a frequency-aware refinement mechanism based on subband-selective processing, which adaptively enhances structural details and suppresses haze artifacts across different spectral components.
- We develop a physics-constrained atmospheric reconstruction strategy with iterative transmission refinement, ensuring consistency with the atmospheric scattering model while correcting residual estimation errors.
- We design an uncertainty-calibrated semi-supervised learning scheme that incorporates heteroscedastic uncertainty modeling to down-weight unreliable regions, enabling stable adaptation to real-world, unpaired, hazy data.
- We integrate all components into a unified hybrid framework and demonstrate consistent improvements across synthetic and real-world benchmarks, as well as enhanced downstream detection performance.

II. RELATED WORK

Recent progress in image dehazing has been driven by prior-based models, deep supervised restoration networks, transformer-based architectures, and emerging real-world adaptation strategies. This section briefly reviews the most relevant lines of work and highlights the limitations that motivate the proposed framework.

A. PRIOR-BASED IMAGE DEHAZING

Early image dehazing methods are largely founded on hand-crafted priors derived from the statistics of clear images and the atmospheric scattering model. A representative example is the DCP proposed by He et al. [3], which estimates transmission by exploiting the observation that at least one color channel in most local haze-free patches tends to contain very low-intensity values. Although DCP is physically intuitive and effective in many outdoor scenes, its underlying assumption is often violated in sky regions, bright

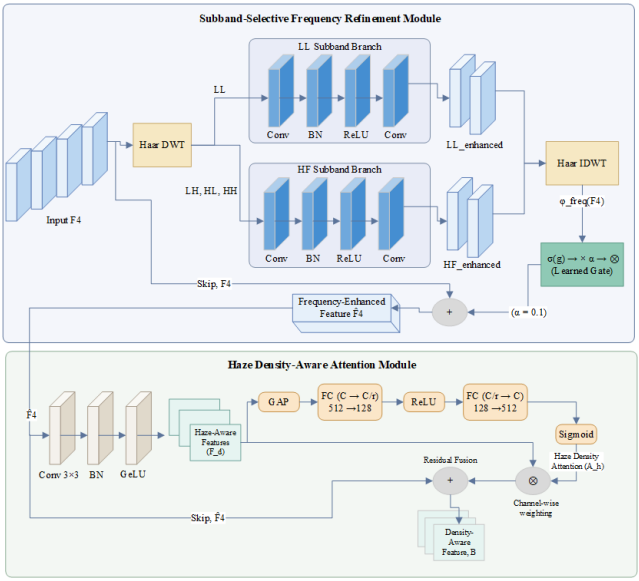

FIGURE 4. Architecture of the proposed Subband-Selective Frequency Refinement Module (top) and Haze Density-Aware Attention Module (bottom).

objects, and complex scenes, which limits its robustness. Later prior-based methods introduced complementary cues, such as color attenuation [16] and non-local priors [17], to better model haze formation. Alternative physically grounded approaches have also explored frequency-domain representations for transmission estimation: demonstrating that subband decomposition can improve transmission accuracy beyond what purely spatial priors achieve [18]. These methods remain highly dependent on scene statistics and generally degrade under dense, spatially varying, or non-uniform haze, where fixed assumptions are no longer reliable. More recently, polarization-based dehazing has emerged as an alternative physical prior, exploiting the polarimetric properties of atmospheric scattering to suppress haze while preserving edge and structural detail through Laplacian-based post-processing [19]. Nevertheless, prior-based methods in general remain highly dependent on scene statistics and degrade under dense, spatially varying, or non-uniform haze, where fixed physical assumptions are no longer reliable.

B. CNN-BASED IMAGE DEHAZING

With the rise of deep learning, supervised dehazing models have largely replaced hand-crafted priors by learning direct haze-to-clear mappings from synthetic paired data. DehazeNet [4] represents one of the earliest end-to-end frameworks for transmission estimation, while AOD-Net [5] simplifies atmospheric modeling through a lightweight unified formulation. Subsequent methods improve restoration quality through stronger feature modeling. GD-Net [6] adopts an attention-based multi-scale grid architecture to better capture spatially varying haze; FFA-Net [8] enhances discriminative responses via channel and pixel attention; and MSBDN [9] introduces multi-scale dense fusion to preserve structural detail during reconstruction. AECR-Net [20] further incorporates contrastive learning to guide restored outputs toward the clean-image distribution. Although these CNN-based methods achieve strong results on synthetic benchmarks, their locality-biased convolutional design still limits global context modeling, which can reduce robustness under dense haze and complex long-range atmospheric degradation.

C. TRANSFORMER-BASED IMAGE DEHAZING

Transformer-based restoration models have recently attracted considerable attention because self-attention provides a stronger mechanism for modeling long-range spatial interactions than conventional convolution. A key foundation is the Swin Transformer [21], which introduces a hierarchical architecture with shifted-window attention to improve efficiency while preserving multi-scale representation learning. Building on this idea, DeHamer [11] combines convolutional and transformer branches through feature modulation and incorporates transmission-aware positional encoding to better represent haze-related structure. DehazeFormer [12] further adapts the transformer design to image restoration by revising normalization and activation components, leading to strong quantitative performance on standard dehazing benchmarks. However, most transformer-based methods still formulate dehazing as a direct restoration task in image space, without explicitly disentangling scene content from haze degradation or sufficiently constraining reconstruction with atmospheric priors. As a result, their performance may remain sensitive when haze characteristics in real scenes deviate substantially from the training distribution.

D. SEMI-SUPERVISED AND UNPAIRED DEHAZING

A major limitation of fully supervised dehazing lies in its dependence on paired hazy/clean datasets, which are typically synthetic and therefore unable to fully represent the complexity of real atmospheric degradation. To reduce this domain gap, several studies have explored unsupervised or semi-supervised learning strategies. Cycle-consistency-based frameworks have been adapted for unpaired dehazing, but they often suffer from unstable optimization and limited structural fidelity [22]. D4+ [22] addresses this issue by decomposing haze formation into scattering and depth-related components, enabling more diverse haze synthesis under physics-aware constraints. PSD [23] leverages pre-trained restoration models to generate pseudo-supervision for real unpaired hazy images. Although these approaches improve adaptation to real data, they generally apply the training signal uniformly across spatial locations, without accounting for the varying reliability of predictions in severely degraded regions. This limitation becomes particularly critical under dense and non-uniform haze, where unreliable supervision can negatively affect optimization stability and restoration quality.

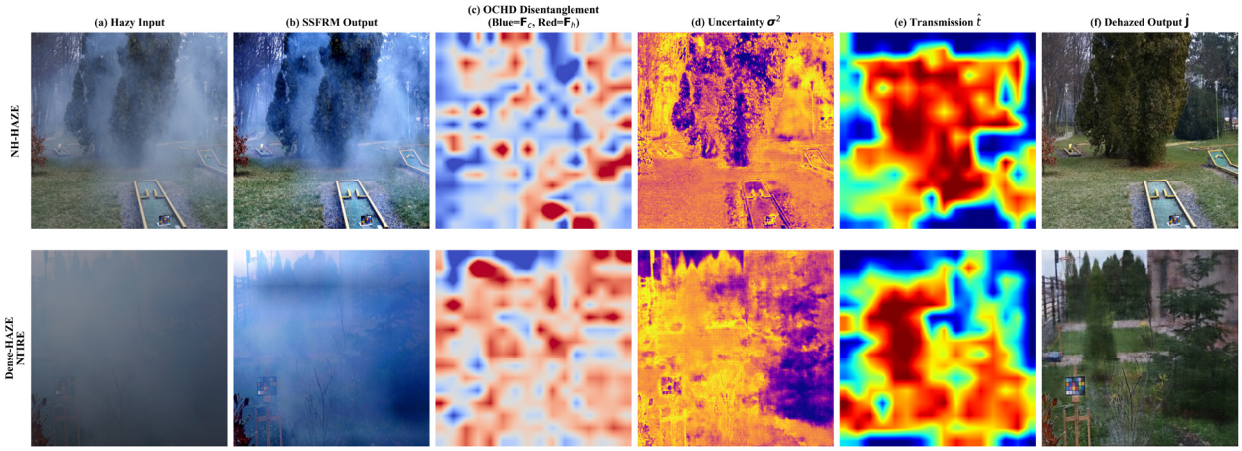

FIGURE 5. Intermediate feature analysis of Semi-PhysHDNet under spatially non-uniform haze (top) and dense real-world haze (bottom). (a) Input hazy image. (b) SSFRM output highlighting subband-selective frequency enhancement and improved structural contrast. (c) OCHD disentanglement map ($\|F_c\|_2 - \|F_h\|_2$), where blue denotes content-dominant regions and red indicates haze-dominant regions. (d) Predicted uncertainty map σ^2, showing elevated responses in ambiguous and heavily degraded areas. (e) Refined transmission map $\hat{t}$ obtained via physics-constrained iterative estimation. (f) Final dehazed output $\hat{J}$.

E. FEATURE DISENTANGLEMENT AND UNCERTAINTY ESTIMATION

Related lines of work can also be found in representation disentanglement and uncertainty-aware learning. In generative modeling and domain adaptation, disentanglement has been used to separate content-related information from style or domain-specific variation, often through variational formulations [24]. In image restoration, AECR-Net [20] introduces contrastive regularization to improve the separation between degraded and restored representations, but it does not explicitly impose orthogonality between scene structure and haze-related degradation within the latent space of a unified dehazing model. In parallel, heteroscedastic uncertainty modeling [25] estimates pixel-wise aleatoric uncertainty together with the primary prediction, allowing unreliable regions to contribute less during optimization. Although this strategy has shown clear benefits in tasks such as depth estimation and semantic segmentation, its use in semi-supervised image dehazing remains largely unexplored, particularly under dense and non-uniform haze where the reliability of unlabeled real-world supervision can vary substantially across spatial regions.

Overall, existing studies have advanced dehazing through stronger restoration backbones, global context modeling, and real-world adaptation, yet explicit content-haze disentanglement and reliability-aware semi-supervised optimization remain insufficiently explored. This gap motivates the proposed framework, which integrates orthogonal content-haze disentanglement and uncertainty-calibrated learning within a unified architecture for dense dehazing.

III. PROPOSED METHODOLOGY

The proposed framework integrates a Haze-Aware Hierarchical Swin Encoder, a Subband-Selective Frequency Refinement Module (SSFRM), a Haze Density-Aware Attention Module (HDAM), an Orthogonal Content-Haze Disentanglement module (OCHD), a Progressive Multi-Scale Reconstruction Decoder (PMRD), a Physics-Constrained Atmospheric Reconstruction Module (PCARM), and an Uncertainty-Calibrated Semi-Supervised Learning branch (UCSSL) within a unified dehazing architecture. Starting from a hazy input image, the network learns haze-robust latent features and restores the clean scene through joint transmission estimation, atmospheric light prediction, and residual refinement.

As illustrated in Fig. 3, the data flow begins with the input hazy image, which is encoded into multi-scale feature maps. The deepest feature undergoes subband-selective frequency refinement and haze-aware modulation before being projected into a latent representation. This representation is explicitly decomposed into content and haze subspaces via orthogonal disentanglement. The fused representation is then progressively decoded while integrating encoder skip connections. Finally, transmission estimation and atmospheric light prediction are performed within a physics-constrained reconstruction module, and the output is refined through residual learning and uncertainty-guided optimization.

To improve clarity, the overall processing pipeline of Semi-PhysHDNet can be summarized in four sequential stages: (1) hierarchical feature extraction using a Swin-based encoder to capture multi-scale haze-aware representations; (2) frequency-aware refinement and haze-density modulation to enhance structural details and suppress haze artifacts; (3) orthogonal disentanglement of latent features into content and haze components followed by progressive reconstruction; and (4) physics-constrained atmospheric reconstruction combined with uncertainty-calibrated semi-supervised learning for robust real-world adaptation.

This staged formulation provides an interpretable view of how each module contributes to the final dehazing output.

Algorithm 1 Training Procedure of Semi-PhysHDNet

Require: Paired hazy-clean samples $\mathcal{D}_p = \{(\mathbf{H}, \mathbf{J})\}$, unpaired real hazy samples $\mathcal{D}_u = \{\mathbf{H}_u\}$, maximum epochs T
Ensure: Trained model parameters Θ

```
Initialize network parameters Θ
for e = 1 to T do
    for each mini-batch do
        Sample paired batch (H, J) ~ D_p
        Generate dehazed output Ĵ, transmission t̂, atmospheric light Â, and uncertainty σ²
        Compute supervised reconstruction loss L_sup
        Compute orthogonality loss L_ortho and physics consistency loss L_phys
        if e > 20 then
            Sample unpaired real batch H_u ~ D_u
            Generate pseudo-target
            J̃ = stopgrad(f_Θ(H_u))
            Compute L_unc
            L_total ← L_sup + λ_o L_ortho
              + λ_p L_phys + λ_u L_unc
        else
            L_total ← L_sup + λ_o L_ortho + λ_p L_phys
        end if
        Update Θ using AdamW
    end for
end for
return Θ
```

TABLE 1. Architecture summary of the proposed Semi-PhysHDNet. Input and output tensor shapes follow $B \times C$ notation. All spatial dimensions at stage i are denoted $H_i \times W_i$.

Module	Input	Output	Key Operations
Swin	$\mathbf{H}\in\mathbb{R}^{B\times 3}$	$\{F_1, F_2, F_3, F_4\}$	Patch embed, W-MSA, SW-MSA, patch merge (4 stages)
SSFRM	F_4	$\tilde{F}_4$	Haar DWT, Swin LL branch, Conv $\{LH,HL,HH\}$, IDWT, residual (α=0.1)
HDAM	$\tilde{F}_4$	$\mathbf{B}$	Channel attention, haze-density modulation, bridge projection (Conv-BN-ReLU $\times$2)
OCHD	$\mathbf{B}\in\mathbb{R}^{B\times H_4\times W_4}$	$\mathbf{F}_c, \mathbf{F}_h, \mathbf{B}^*$	Parallel projections ϕ_c/ϕ_h (Conv1 $\times$ 1), $\mathcal{L}_{\text{ortho}}$, residual fusion ψ
PMRD	$\mathbf{B}^*, \{F_1, F_2, F_3\} D_1, D_2, D_3$		$\times$3 bilinear upsample, channel concat, Conv-BN-LeakyReLU fusion, skip connections
PCARM	$D_3, \mathbf{H}$	$\hat{t}, \hat{\mathbf{A}}, \hat{\mathbf{J}}_{\text{phys}}$	DCP prior, iterative residual refinement (K=3), GAP$\rightarrow \phi_A$, physics inversion + residual head
UCSSL	D_3	$\boldsymbol{\mu}, \boldsymbol{\sigma}^2$	Dual conv heads $\mathcal{H}_\mu/\mathcal{H}_\sigma$, heteroscedastic NLL, stop-gradient pseudo-target

Table 1 provides a structured overview of all seven modules constituting Semi-PhysHDNet, including their inputs, outputs, and key operations. Together with Fig. 3 and Fig. 4, this summary is intended to support reproducibility and facilitate comparison with alternative designs.

A. PROBLEM FORMULATION

Let $\mathbf{H} \in \mathbb{R}^{B\times 3\times H\times W}$ denote a batch of hazy RGB images, where B is the batch size and H, W are spatial dimensions. For each spatial coordinate x, haze formation is governed by the atmospheric scattering model:

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

where $\mathbf{J}(x) \in \mathbb{R}^3$ denotes the haze-free scene radiance, $t(x) \in [0, 1]$ denotes the transmission coefficient describing scene visibility, and $\mathbf{A} \in \mathbb{R}^3$ denotes global atmospheric light. The objective is to estimate $\hat{\mathbf{J}}$, $\hat{t}$, and $\hat{\mathbf{A}}$ such that Eq. (1) is approximately satisfied while preserving structural integrity and perceptual consistency.

B. HAZE-AWARE HIERARCHICAL SWIN ENCODER

The encoder $\mathcal{E}$ extracts hierarchical feature representations from the input hazy image:

$$\{\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3, \mathbf{F}_4\} = \mathcal{E}(\mathbf{H}), \tag{2}$$

where $\mathbf{F}_i \in \mathbb{R}^{B\times C_i\times H_i\times W_i}$ denotes the feature map at stage i. The encoder is implemented using a Swin Transformer backbone, in which each stage performs patch embedding, window-based multi-head self-attention, shifted-window attention, and patch merging. This hierarchical design progressively reduces spatial resolution while increasing channel dimensionality, enabling increasingly abstract feature modeling.

The shallow stages preserve local texture and fine spatial structure, while deeper stages encode broader contextual and haze-related information. In particular, the deepest feature map $\mathbf{F}_4$ captures high-level semantic and atmospheric representations.

Prior to latent projection, the deepest encoder feature F_4 is passed through a wavelet enhancement branch:

$$\tilde{F}_4 = \phi_{\text{freq}}(F_4), \tag{3}$$

where $\phi_{\text{freq}}(\cdot)$ applies Haar DWT to decompose F_4 into four subbands:

$$LL, LH, HL, HH = \text{DWT}(F_4). \tag{4}$$

The LL subband carries a coarse, haze-dominant structure and is processed through a lightweight Swin-based branch to suppress atmospheric interference. The detail subbands

TABLE 2. PSNR/SSIM comparison across RESIDE (Outdoor), RESIDE (Indoor), O-HAZE, NH-HAZE, and Dense-HAZE. Higher is better.

Method	RESIDE (Outdoor)		RESIDE (Indoor)		O-HAZE		NH-HAZE		Dense-HAZE	
	PSNR (dB)↑	SSIM↑	PSNR (dB)↑	SSIM↑	PSNR (dB)↑	SSIM↑	PSNR (dB)↑	SSIM↑	PSNR (dB)↑	SSIM↑
DCP [3]	19.13	0.815	16.62	0.818	16.60	0.653	14.56	0.646	–	–
DehazeNet [4]	21.14	0.847	19.82	0.821	–	–	–	–	–	–
AOD-Net [5]	20.06	0.847	20.51	0.816	19.90	0.816	–	–	–	–
GD-Net [6]	30.86	0.982	32.16	0.984	–	–	13.80	0.537	19.53	0.726
MSBDN [9]	–	–	33.79	0.984	–	–	19.23	0.706	22.55	0.758
AECR-Net [20]	33.12	0.948	33.17	0.990	16.53	0.626	15.57	0.618	–	–
FFANet [8]	31.98	0.992	33.57	0.984	–	–	12.06	0.772	22.26	0.781
EPDN [26]	22.57	0.863	25.06	0.923	15.40	0.712	16.28	0.621	–	–
Dehamer [11]	35.17	0.986	36.63	0.988	20.74	0.813	20.66	0.684	–	–
DeFormer [12]	26.44	0.962	**38.93**	0.988	–	–	–	–	**24.01**	**0.815**
TSMD-Net [27]	25.31	0.924	22.29	0.849	19.11	0.715	–	–	17.47	0.631
D4+† [28]	25.83	0.956	25.42	0.932	–	–	–	–	–	–
Restormer† [29]	–	–	38.88	0.991	23.58	0.768	–	–	15.78	0.548
OKNet-S† [30]	**35.45**	**0.992**	37.49	**0.994**	–	–	20.29	**0.800**	16.85	0.620
(Ours)	35.25	0.998	36.89	0.992	**23.98**	**0.867**	**20.87**	0.789	23.57	0.789

†Values taken directly from original publications on identical test sets. '-' denotes results neither reported in the original publication nor reproducible from available official implementations. All remaining values are self-evaluated under the unified experimental settings described in Section IV-A2 (resolution: 512×512, datasets: as listed, using official author-released code where available).

TABLE 3. No-reference perceptual evaluation on RTTS. Lower is better.

Method	NIQE↓	FADE↓	BRISQUE↓
DCP [3]	5.462	1.115	–
EPDN [26]	5.112	1.221	–
GD-Net [6]	4.987	0.984	–
Dehamer [11]	4.912	0.928	33.866
MSBDN [9]	**3.154**	1.363	28.743
SSLD [23]	3.489	0.867	32.428
RIDCP [31]	–	0.944	**18.782**
(Ours)	4.859	**0.807**	27.94

TABLE 4. VIF and FOM comparison across synthetic (RESIDE) and real-world datasets. Higher values indicate better structural fidelity (VIF) and edge preservation (FOM).

Dataset	**VIF↑**	**FOM↑**
RESIDE Indoor (SOTS)	0.2436	0.9580
RESIDE Outdoor (SOTS)	0.1953	0.9611
NH-HAZE	0.0975	0.8519
Dense-HAZE	0.0715	0.7224
O-HAZE	0.1260	0.8863
Average	**0.1468**	**0.8759**

$\{LH, HL, HH\}$, encoding horizontal, vertical, and diagonal edges respectively, are refined through convolutional processing to recover fine scene structure. The enhanced subbands are then recombined:

$$\tilde{F}_4 = \mathrm{IDWT}(LL_{\mathrm{enh}},\ LH_{\mathrm{enh}},\ HL_{\mathrm{enh}},\ HH_{\mathrm{enh}}), \tag{5}$$

The enhanced subbands are recombined and integrated with the original encoder feature F_4 through a residual connection weighted by a fixed scalar α=0.1:

$$\tilde{F}_4 = F_4 + \alpha \cdot \Delta F_4, \quad \alpha = 0.1. \tag{6}$$

The residual weight is intentionally kept small so that frequency refinement acts as a corrective modulation rather than dominating the primary encoder representation. This is particularly important before the disentanglement stage, where overly strong frequency perturbation may distort structural cues needed for stable content-haze separation. In our implementation, α was fixed empirically as a conservative scaling factor based on preliminary validation experiments, which showed that larger values led to less stable feature fusion and occasional over-enhancement, while smaller values reduced the contribution of subband refinement.

Fig. 4 illustrates both refinement modules. The Wavelet Enhancement Module separates encoder features into complementary frequency bands, processes each independently, and reconstructs an enriched representation through inverse DWT with gated residual fusion. The Haze Density-Aware Module reweights features via channel attention conditioned on haze-sensitive responses, yielding a density-aware representation for physics-guided reconstruction.

The enhanced feature is then projected through a bridge network:

$$\mathbf{B} = \phi_{\mathrm{bridge}}(\tilde{\mathbf{F}}_4), \tag{7}$$

where $\phi_{\mathrm{bridge}}(\cdot)$ consists of stacked convolution, batch normalization, and nonlinear activation layers, with an embedded haze-density-aware processing module that adaptively modulates feature responses. The resulting latent representation

$$\mathbf{B} \in \mathbb{R}^{B\times 512\times H_4\times W_4} \tag{8}$$

is forwarded to the disentanglement and reconstruction stages.

C. ORTHOGONAL CONTENT-HAZE DISENTANGLEMENT MODULE

Latent representations extracted from hazy images typically contain entangled scene structure and degradation

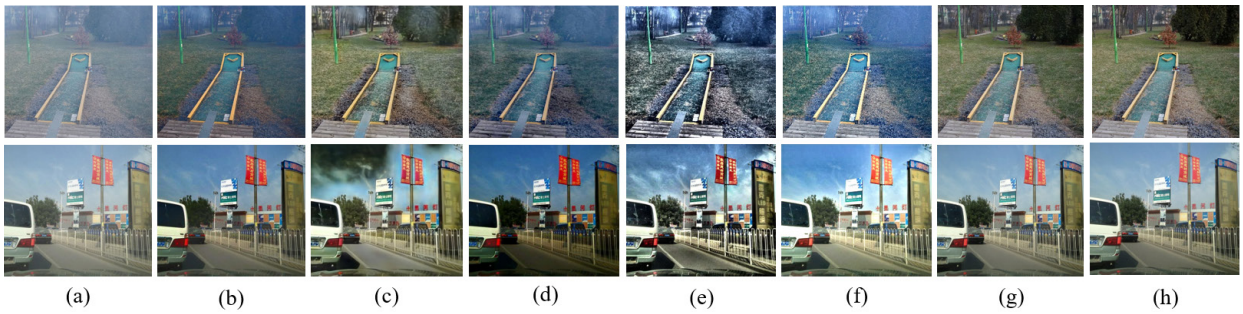

FIGURE 6. Qualitative comparison on O-HAZE (first row) and SOTS Outdoor (second row). From left to right: (a) hazy input, (b) MSCNN, (c) AOD-Net, (d) GD-Net, (e) Dehamer, (f) Deformer, (g) the proposed method, and (h) ground truth.

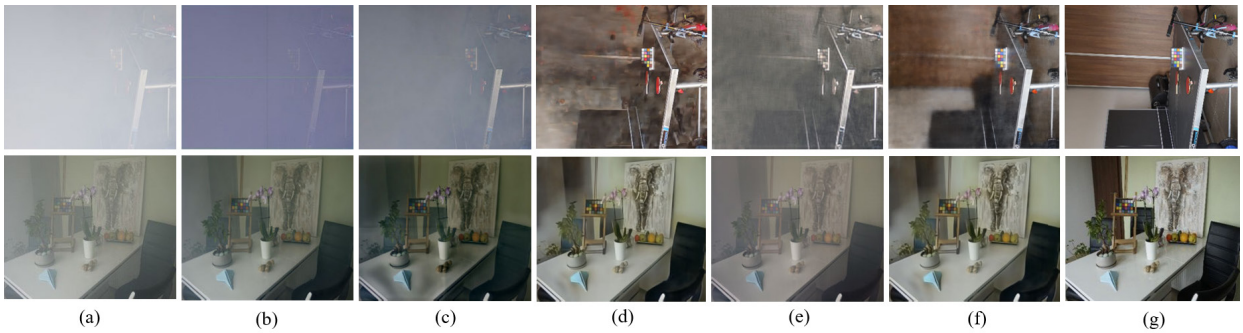

FIGURE 7. Qualitative comparison of Dense-Haze Indoor (first row) and I-HAZE (second row). From left to right: (a) hazy input, (b) AOD-Net, (c) GD-Net, (d) LDFormer, (e) DDPM, (f) the proposed method, and (g) ground truth. The outputs of the compared methods are adapted from [20].

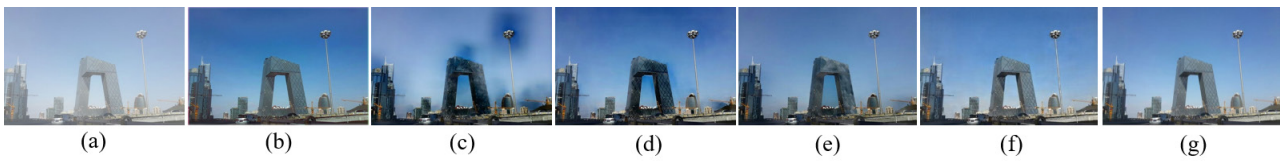

FIGURE 8. Qualitative comparison of RESIDE-Outdoor. From left to right: (a) hazy input, (b) AOD-Net, (c) GD-Net, (d) FFANet, (e) Deformer, (f) the proposed method, and (g) ground truth.

information. To explicitly separate these components, the Bridge feature ($\mathbf{B}$) is decomposed into two complementary representations corresponding to haze-invariant scene content and haze-related degradation cues.

Two parallel projection mappings are applied to the bridge feature:

$$\mathbf{F}_c = \phi_c(\mathbf{B}), \tag{9}$$

$$\mathbf{F}_h = \phi_h(\mathbf{B}), \tag{10}$$

The content branch ($\mathbf{F}_c$) is designed to capture haze-invariant scene structure such as edges, textures, and object boundaries, which correspond to the underlying clean image representation. In contrast, the haze branch ($\mathbf{F}_h$) models degradation-related characteristics, including contrast attenuation, atmospheric scattering, and haze density variations. By explicitly separating structural and degradation information, the network learns disentangled representations that facilitate more robust haze removal.

The resulting features are concatenated and fused through a residual projection layer to produce the refined latent representation used for subsequent decoding.

To enforce feature separation, the content and haze representations are reshaped across spatial locations and normalized along the channel dimension. A spatial correlation matrix is then computed between the two normalized feature spaces for each sample, and the orthogonality loss is defined as the mean absolute correlation:

$$\mathbf{M}^{(i)} = \bar{\mathbf{F}}_c^{(i)\top} \bar{\mathbf{F}}_h^{(i)}, \tag{11}$$

$$\mathcal{L}_{\text{ortho}} = \text{Clamp}\left(\frac{1}{B} \sum_{i=1}^{B} \text{mean}\left(\left| \mathbf{M}^{(i)} \right| \right), 0, 1 \right). \tag{12}$$

where $\bar{\mathbf{F}}_c$ and $\bar{\mathbf{F}}_h$ denote the normalized feature representations. Minimizing this loss suppresses correlation between the content and haze subspaces and promotes disentangled latent representations.

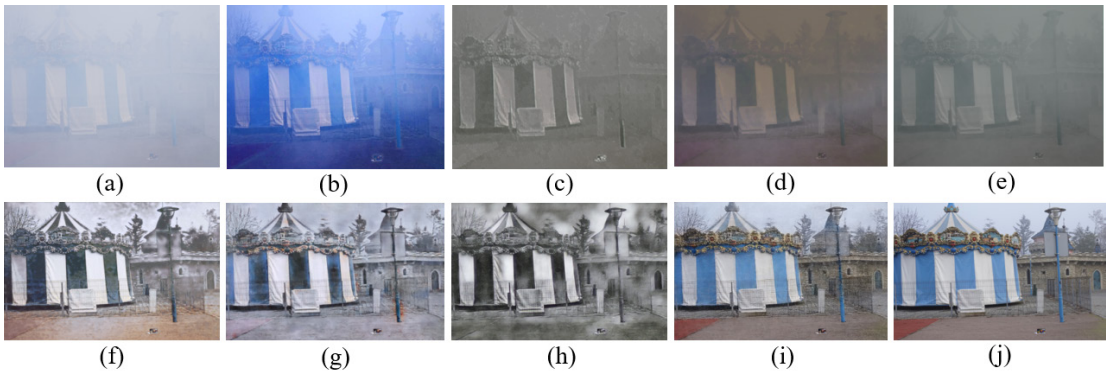

FIGURE 9. Qualitative comparison on Dense-Haze. From left to right and top to bottom: (a) hazy input, (b) DCP, (c) DehazeNet, (d) AOD-Net, (e) GD-Net, (f) FFA-Net, (g) MSBDN, (h) KDDN, (i) the proposed method, and (j) ground truth. The outputs of the compared methods are adapted from [20] with permission from IEEE.

After disentanglement, the two feature components are fused through channel-wise aggregation followed by a learnable projection. The resulting feature is combined with the original bridge representation through a residual connection, yielding the refined latent representation,

$$\mathbf{B}^* = \mathbf{B} + \psi([\mathbf{F}_c, \mathbf{F}_h]), \tag{13}$$

where $\psi(\cdot)$ denotes a learnable transformation that integrates structural and degradation cues. The refined feature $\mathbf{B}^*$ is subsequently forwarded to the decoder for multi-scale reconstruction.

D. PROGRESSIVE MULTI-SCALE RECONSTRUCTION DECODER

The decoder reconstructs high-resolution representations by progressively upsampling the latent feature $\mathbf{B}^*$ while integrating encoder features through skip connections. This design enables recovery of spatial details that are lost during hierarchical encoding.

The decoding process can be expressed as

$$\mathbf{D}_1 = \mathcal{U}_1(\mathbf{B}^*, \mathbf{F}_3), \tag{14}$$
$$\mathbf{D}_2 = \mathcal{U}_2(\mathbf{D}_1, \mathbf{F}_2), \tag{15}$$
$$\mathbf{D}_3 = \mathcal{U}_3(\mathbf{D}_2, \mathbf{F}_1), \tag{16}$$

where $\mathcal{U}_i(\cdot)$ denotes an upsampling-and-fusion block consisting of bilinear upsampling, channel-wise concatenation with the corresponding skip feature F_i, and two sequential Conv-BN-LeakyReLU layers that project the concatenated representation to the target channel dimensionality. This design avoids checkerboard artifacts associated with transposed convolutions while preserving multi-scale structural cues through the skip connections. The channel dimensions at each decoder stage follow a symmetric schedule relative to the encoder: D_1 halves the encoder bottleneck dimensionality, and each subsequent stage further reduces channels while doubling spatial resolution. Gradient propagation is stabilized throughout training by the residual paths inherited from the skip connections, which provide direct gradient highways between shallow encoder features and the deep decoder supervision signal.

The final decoder representation $\mathbf{D}_3 \in \mathbb{R}^{B\times C_d\times H\times W}$ restores the full input spatial resolution and serves as the shared feature source for three parallel output branches: the physics-based reconstruction in PCARM, the residual correction head $\mathcal{H}_R$, and the uncertainty prediction heads $\mathcal{H}_\mu$ and $\mathcal{H}_\sigma$ in UCSSL.

E. PHYSICS-CONSTRAINED ATMOSPHERIC RECONSTRUCTION MODULE

To incorporate physical priors of atmospheric scattering, the transmission map and atmospheric illumination are estimated using a hybrid strategy that combines learned features with physics-inspired guidance.

An initial transmission estimate t_0 is predicted from the decoder feature $\mathbf{D}_3$ through a convolutional projection followed by a bounded activation function. To enforce physical consistency, a transmission prior derived from the dark channel assumption is introduced:

$$t_{\text{phys}}(x) = 1 - \min_{y\in\Omega(x)} \left(\min_{c\in\{R,G,B\}} \mathbf{H}_c(y) \right), \tag{17}$$

where $\Omega(\mathbf{x})$ denotes a local spatial neighborhood of radius r centered at $\mathbf{x}$, and $\mathbf{H}_c(y)$ denotes the intensity of the channel $c \in \{R, G, B\}$ of the hazy input image $\mathbf{H}$ at spatial location y.

The transmission map is iteratively refined by predicting residual corrections conditioned on the decoder features and the discrepancy between the prior and the current estimate:

$$\mathbf{r}_k = t_{\text{phys}} - t_k, \tag{18}$$
$$t_{k+1} = \text{Clamp}(\mathcal{R}_k([\mathbf{D}_3,\ \mathbf{r}_k])\ ,\ 0,\ 1)\ . \tag{19}$$

where $\mathcal{R}_k(\cdot)$ a learnable convolutional refinement block is at iteration k, and Clamp$(\cdot, 0, 1)$ constrains the transmission estimate to a physically valid range. The residual $\mathbf{r}_k$ conditions each refinement step on the discrepancy between the physics prior and the current estimate, providing an explicit correction signal at every iteration. After the final refinement

iteration, the resulting transmission map is denoted as

$$\hat{t} = t_K. \tag{20}$$

The atmospheric illumination parameter is estimated from the global scene representation. Since atmospheric light is assumed to be spatially homogeneous, the bridge feature $\mathbf{B}$ is aggregated via global average pooling and mapped to the atmospheric illumination vector:

$$\hat{\mathbf{A}} = \phi_A(\mathrm{GAP}(\mathbf{B})), \tag{21}$$

where $\mathrm{GAP}(\cdot)$ denotes global average pooling and $\phi_A(\cdot)$ represents a learnable projection that predicts the three-channel atmospheric illumination vector.

Given the estimated transmission $\hat{t}$ and atmospheric illumination $\hat{\mathbf{A}}$, the clean image is reconstructed according to the atmospheric scattering model:

$$\hat{\mathbf{J}}_{\mathrm{phys}}(x) = \frac{\mathbf{H}(x) - \hat{\mathbf{A}}(1 - \hat{t}(x))}{\hat{t}(x) + \epsilon}. \tag{22}$$

where $\epsilon{=}10^{-6}$ is a numerical stability constant that prevents division by zero in near-zero transmission regions. To compensate for model inaccuracies in complex scenes, a residual correction map is predicted from the decoder features:

$$\mathbf{R} = \mathcal{H}_R(\mathbf{D}_3). \tag{23}$$

where $\mathcal{H}_R(\cdot)$ denotes a lightweight convolutional residual head that maps the final decoder feature $\mathbf{D}_3 \in \mathbb{R}^{B\times C_d\times H\times W}$ to a pixel-wise correction field $\mathbf{R} \in \mathbb{R}^{B\times 3\times H\times W}$. The final dehazed output is obtained by adding the physics-guided reconstruction and the learned residual:

$$\hat{\mathbf{J}} = \hat{\mathbf{J}}_{\mathrm{phys}} + \mathbf{R}, \tag{24}$$

where $\mathbf{R} = \mathcal{H}_R(\mathbf{D}_3)$ absorbs scene-specific deviations from the atmospheric scattering model that the physics prior cannot resolve. That includes spatially varying illumination, transmission estimation errors, and haze-independent color artifacts.

Algorithm 1 summarizes the training flow of the proposed network. The hazy input is encoded, refined by the frequency and haze-aware modules, disentangled into content and haze features, and then decoded for dehazing. Physical priors and uncertainty estimation are incorporated during optimization, and all objectives are jointly used to update the generator parameters.

F. UNCERTAINTY-CALIBRATED SEMI-SUPERVISED LEARNING (UCSSL)

To enable stable learning from unpaired real hazy images, the generator is augmented with an uncertainty prediction branch connected to the final decoder feature $\mathbf{D}_3 \in \mathbb{R}^{B\times C_d\times H\times W}$. This branch performs heteroscedastic uncertainty modeling by predicting both the restored mean image and its associated predictive variance at each pixel.

The mean prediction is defined as

$$\boldsymbol{\mu} = \tanh\big(\mathcal{H}_\mu(\mathbf{D}_3)\big), \tag{25}$$

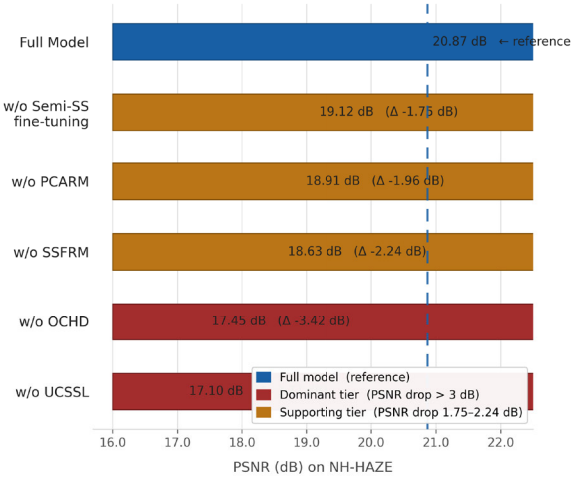

FIGURE 10. Component-wise contribution analysis on NH-HAZE using PSNR (dB). The full model achieves 20.87 dB, while removing individual components leads to consistent performance degradation. The dominant contributions are observed for UCSSL and OCHD (PSNR drop > 3 dB), whereas SSFRM, PCARM, and semi-supervised fine-tuning provide complementary improvements (PSNR drop ∼1.7–2.24 dB). The dashed line indicates the full-model reference performance.

where $\mathcal{H}_\mu(\cdot)$ denotes a convolutional mapping that projects decoder features into a three-channel RGB image. The resulting tensor $\boldsymbol{\mu} \in \mathbb{R}^{B\times 3\times H\times W}$ represents the predicted clean image.

In parallel, the predictive variance is estimated as

$$\log \boldsymbol{\sigma}^2 = \mathrm{Clamp}\big(\mathcal{H}_\sigma(\mathbf{D}_3), a, b\big), \tag{26}$$

where $\mathcal{H}_\sigma(\cdot)$ denotes a convolutional projection producing a spatial variance map, and $\mathrm{Clamp}(\cdot, a, b)$, $a{=}-4.6$ and $b{=}1.1$ restrict the log variance to a numerically stable interval. The lower bound prevents variance collapse, which would produce pathologically large gradients in the uncertainty loss. While the upper bound prevents degenerate over-uncertainty that would suppress all training signals. The resulting tensor$\boldsymbol{\sigma}^2$ models heteroscedastic uncertainty that varies across spatial locations depending on scene content and haze density.

For unpaired samples, the reconstructed image $\hat{\mathbf{J}}$ is used as a pseudo-target. To prevent degenerate optimization, gradients are detached from this pseudo-target using the stop-gradient operator $\mathrm{stopgrad}(\cdot)$.

The uncertainty-aware objective is formulated as a Gaussian negative log-likelihood:

$$\mathcal{L}_{\mathrm{unc}} = \frac{1}{2}\left(\frac{\|\boldsymbol{\mu} - \mathrm{stopgrad}(\hat{\mathbf{J}})\|^2}{\boldsymbol{\sigma}^2} + \log \boldsymbol{\sigma}^2\right). \tag{27}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}^2$ are the predicted mean restoration and heteroscedastic variance maps from $\mathcal{H}_\mu(\mathbf{D}_3)$ and $\mathcal{H}_\sigma(\mathbf{D}_3)$ respectively, and $\mathrm{stopgrad}(\hat{\mathbf{J}})$ is the gradient-detached pseudo-target that prevents circular decoder backpropagation. The inverse-variance weighting

attenuates gradients from uncertain regions while $\log \sigma^2$ regularizing against unbounded variance inflation.

This formulation corresponds to maximum likelihood estimation under a pixel-wise Gaussian observation model. The reconstruction error is weighted by the inverse variance, allowing pixels with high predictive uncertainty to contribute less to the optimization process. As a result, the network learns to focus on reliable regions while reducing the influence of uncertain areas in unlabeled real hazy images.

1) SUPPRESSION OF SEMI-SUPERVISED EDGE OVER-SMOOTHING

A known limitation of pseudo-label-based semi-supervised learning is the tendency to over-smooth high-frequency boundary regions. Under dense haze, structural edges are significantly attenuated, leading to unreliable pseudo-labels near object boundaries. When treated uniformly during optimization, these uncertain boundary predictions suppress gradient updates for sharp transitions, resulting in progressive edge softening.

The proposed UCSSL formulation mitigates this effect through heteroscedastic uncertainty weighting. The predicted variance map $\boldsymbol{\sigma}^2$ assigns higher uncertainty to structurally ambiguous regions, particularly near edges. Consequently, the inverse-variance weighting in $\mathcal{L}_{\text{unc}}$ reduces the influence of unreliable boundary pseudo-labels, preventing the accumulation of erroneous edge gradients during training.

This primary mechanism is further reinforced by three complementary design components: (i) OCHD disentanglement, which isolates haze-invariant structural information in $\mathbf{F}_c$ and reduces degradation interference; (ii) SSFRM frequency refinement, which preserves high-frequency edge information through subband-selective processing; and (iii) Edge-aware supervision, where $\mathcal{L}_{\text{edge}}$ enforces gradient consistency during supervised training, promoting boundary-sensitive representations.

Despite these mechanisms, minor edge softening may persist under extremely dense haze conditions, where structural cues are severely degraded at the input level. A more detailed analysis is provided in Section IV-E.

TABLE 5. Component-wise ablation on NH-HAZE. Each variant removes one key module from the full model.

Configuration	PSNR↑	SSIM↑	ΔPSNR	ΔSSIM
Full model (Ours)	**20.87**	**0.7892**	–	–
w/o UCSSL	17.10	0.6981	-3.77	-0.091
w/o OCHD	17.45	0.7074	-3.42	-0.082
w/o SSFRM	18.63	0.7361	-2.24	-0.053
w/o PCARM	18.91	0.7428	-1.96	-0.046
w/o semi-supervised fine-tuning	19.12	0.7513	-1.75	-0.038

G. UNIFIED SEMI-SUPERVISED OPTIMIZATION OBJECTIVE

The model is trained using both paired synthetic samples and unpaired real hazy images. Let $\mathcal{D}_p = \{(\mathbf{H}, \mathbf{J})\}$ denote paired training data where the ground-truth clean image $\mathbf{J}$ is available, and let $\mathcal{D}_u = \{\mathbf{H}_u\}$ denote unpaired real hazy samples. During training, mini-batches are sampled from both datasets, and different objective terms are activated depending on the supervision type.

TABLE 6. Complexity comparison of representative dehazing methods. Lower values indicate higher efficiency. FLOPs are computed at 256×256 for all methods to ensure fair comparison. The proposed method is trained at 512×512, but FLOPs are recomputed at 256×256 for consistency with prior works.

Method	Params (M)↓	FLOPs (G)↓
AOD-Net [5]	0.002	0.005
FFA-Net [8]	4.46	287.8
MSBDN [9]	31.35	–
Dehamer [11]	29.44	870
DehazeFormer-L [12]	25.44	279.7
TSMD-Net [27]	16.3	97
HITFormer [32]	2.0	35
Ours (256×256)	28.98	32.74

1) PAIRED TRAINING OBJECTIVE

The overall supervised loss for paired samples combines reconstruction, structural, and regularization terms:

$$\mathcal{L}_{\text{sup}} = \mathcal{L}_{\text{L1}} + \mathcal{L}_{\text{SSIM}} + \mathcal{L}_{\text{wav}} + \mathcal{L}_{\text{edge}} + \mathcal{L}_{\text{LAB}} + \mathcal{L}_{\text{TV}} + \mathcal{L}_{\text{ortho}} + \mathcal{L}_{\text{unc}}. \tag{28}$$

For paired samples $(\mathbf{H}, \mathbf{J}) \in \mathcal{D}_p$, the generator is optimized directly on the supervised objective:

$$\mathcal{L}_G^{\text{paired}} = \mathcal{L}_{\text{sup}}. \tag{29}$$

2) UNPAIRED TRAINING OBJECTIVE

For unpaired real hazy samples $\mathbf{H}_u \in \mathcal{D}_u$, the model is trained using an uncertainty-weighted self-consistency objective. The generator prediction serves as a detached pseudo-target, while the predicted uncertainty modulates the contribution of each pixel:

$$\mathcal{L}_G^{\text{unpaired}} = 0.1\,\mathcal{L}_{\text{self-con}} + 0.01\,\mathcal{L}_{\sigma\text{-reg}}. \tag{30}$$

where

$$\mathcal{L}_{\text{self-con}} = \text{mean}\left(\frac{|\boldsymbol{\mu} - \text{stopgrad}(\hat{\mathbf{J}})|}{\boldsymbol{\sigma} + \epsilon}\right), \tag{31}$$

$$\mathcal{L}_{\sigma\text{-reg}} = \text{mean}(\boldsymbol{\sigma}). \tag{32}$$

3) OPTIMIZATION PROCEDURE

During training, paired samples optimize the supervised objective, while unpaired samples contribute through the semi-supervised self-consistency update from epoch 21 onward.

IV. RESULTS AND DISCUSSION

A. EXPERIMENTAL SETUP

1) DATASETS

The proposed framework was evaluated on seven public dehazing benchmarks: RESIDE Outdoor [33] (synthetic outdoor scenes), NTIRE Dense-Haze, O-HAZE [34] (real

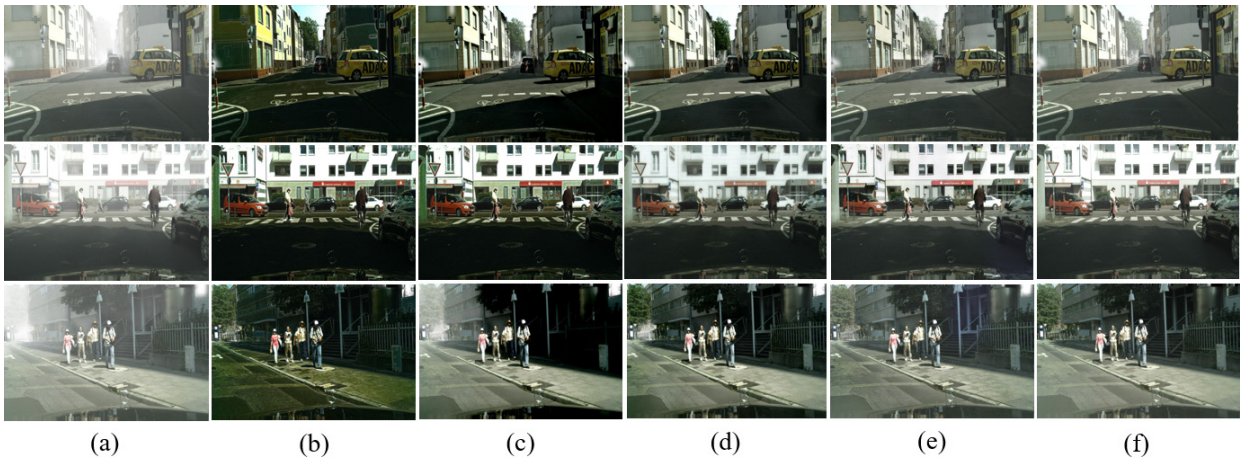

FIGURE 11. Qualitative comparison on Foggy Cityscapes. From left to right: (a) hazy input, (b) DCP, (c) YOLY, (d) SGID, (e) the proposed method, and (f) ground truth/reference image.

outdoor haze), I-HAZE [35] (real indoor haze), NH-HAZE [15] (dense, non-homogeneous real haze), Foggy Cityscapes [36], and RTTS. These datasets include both synthetic and real hazy images under indoor and outdoor scenes, covering homogeneous, dense, and non-homogeneous haze conditions. RESIDE SOTS Indoor and Outdoor were used to assess performance on widely adopted synthetic benchmarks, while O-HAZE, I-HAZE, Dense-Haze, and NH-HAZE provided challenging real-world paired evaluation scenarios. Foggy Cityscapes was included to examine urban street scenes under synthetic fog degradation, and RTTS was used for qualitative evaluation on real traffic scenes without reference ground truth. This dataset selection enables a comprehensive assessment of the proposed method across diverse haze distributions, scene structures, and domain conditions.

2) IMPLEMENTATION DETAILS

a: HARDWARE AND SOFTWARE PLATFORM

All experiments are implemented in PyTorch 2.1.0 (Python 3.10, CUDA 12.1) and executed on two NVIDIA GeForce RTX 4090 GPUs. Unless otherwise specified, all models are trained and evaluated at a fixed input resolution of 512×512.

b: OPTIMIZATION SETTINGS

Model parameters are optimized using AdamW [37] with β_1=0.9, β_2=0.999, weight decay λ_w=10^{-4}, and initial learning rate η_0=10^{-4}. A cosine annealing schedule [38] decays the learning rate to $\eta_{\min}$=10^{-6} over T=150 epochs. Training is performed with a mini-batch size of 2 per GPU, an effective batch size 4, and gradient norms are clipped to $\|\nabla\|_2 \leq 1.0$ to stabilize optimization.

c: TRAINING SET CONSTRUCTION

The supervised phase (Phase 1) is trained using paired hazy-clean images from the following sources: RESIDE ITS (13,990 pairs), RESIDE OTS (313,950 pairs), the training splits of Dense-HAZE (45 pairs), NH-HAZE (45 pairs) and FoggyCityscapes (500 paired frames). The official test splits of each real-world benchmark are strictly withheld from training and used exclusively for final evaluation. From epoch 21, Phase 2 introduces unpaired real-world hazy images from RESIDE-β and RTTS using a 1:1 paired-to-unpaired sampling ratio, enabling semi-supervised domain adaptation.

d: VALIDATION AND MODEL SELECTION

Model selection is performed on a held-out validation partition comprising 10% of the RESIDE training splits, stratified to preserve haze-density distribution. Validation PSNR is computed every 2 epochs, and the checkpoint achieving the highest average validation PSNR is selected.

e: TEST SET CONSTRUCTION

All reported results use standard publicly defined benchmark test splits. For RESIDE, the official SOTS subsets are used (500 indoor and 500 outdoor pairs). For Dense-HAZE and NH-HAZE, the last 10 images (by filename order, images 46-55) are strictly reserved as the held-out test split and are excluded from training. For O-HAZE and I-HAZE, the complete datasets are used for evaluation only and are not included in training. No-reference evaluation on RTTS uses all 4,322 real, hazy images. Foggy Cityscapes detection uses the official validation split (500 images).

f: COMPARISON PROTOCOL AND BASELINE REPRODUCIBILITY

To ensure a scientifically fair and transparent comparison, all baseline methods are evaluated exclusively at *inference* using official author-released pretrained weights; no baseline is retrained or fine-tuned under our settings. This approach avoids confounding the comparison with training hyperparameter choices (batch size, epochs, optimizer settings) that vary across architectures and are not the subject of this evaluation.

Specifically, DCP [3], DehazeNet [4], AOD-Net [5], FFA-Net [8], MSBDN [9], AECR-Net [20], Dehamer [11], and DehazeFormer [12] are evaluated using their official PyTorch implementations with provided pretrained weights. For methods without compatible official implementations (marked † in Table 2), results are taken from the original publications where the authors evaluated on identical benchmark test splits (RESIDE SOTS, O-HAZE, NH-HAZE, Dense-HAZE); using published values in this case is the standard accepted practice in the image restoration literature and does not introduce unfair advantage.

The following unified inference protocol is applied to all self-evaluated methods:

- **Resolution:** All test images resized to 512×512 using bicubic interpolation prior to inference, ensuring resolution-consistent evaluation across all methods.
- **Test splits:** Standard publicly defined benchmark splits, identical for all methods, as described in Section IV-A2.
- **Preprocessing:** pixel values normalized to $[0, 1]$ by dividing by 255, no additional augmentation or post-processing applied at inference;
- **Metrics:** PSNR and SSIM computed using scikit-image 0.21 on full-resolution RGB outputs;
- **Hardware:** All self-evaluated methods run on the same 2×RTX 4090 workstation to ensure consistent runtime conditions.

g: EVALUATION PROTOCOL

PSNR and SSIM are computed using scikit-image 0.21 on full-resolution 512×512 RGB outputs. VIF [39] and FOM [40] are computed on all paired benchmarks using standard reference implementations. No-reference metrics (FADE, BRISQUE, NIQE) are computed on RTTS using their standard implementations. All results are reported as averages over complete test splits with no selective sampling.

h: LOSS CONFIGURATION

The supervised and unpaired objectives are defined in Section III-G. The concrete scalar weights applied to each term during training are reported in Table 7, selected by grid search on the held-out validation partition. The orthogonality and uncertainty losses $\mathcal{L}_{\text{ortho}}$ and $\mathcal{L}_{\text{unc}}$ remain active for unpaired batches since neither requires ground-truth supervision.

$\mathcal{L}_{\text{L1}}$ and $\mathcal{L}_{\text{SSIM}}$ provide primary pixel-level and structural supervision. $\mathcal{L}_{\text{wav}}$ penalises subband discrepancy across all Haar decomposition levels to preserve high-frequency texture. $\mathcal{L}_{\text{edge}}$ minimises Sobel gradient error to enforce boundary sharpness. $\mathcal{L}_{\text{LAB}}$ constrains perceptual colour fidelity in the CIE-LAB space. $\mathcal{L}_{\text{TV}}$ applies total variation regularization for spatial smoothness. $\mathcal{L}_{\text{ortho}}$ imposes representational orthogonality between content and haze subspaces (OCHD), and $\mathcal{L}_{\text{unc}}$ is the uncertainty-weighted negative log-likelihood objective (UCSSL).

For unpaired real-domain samples, the supervised terms $\mathcal{L}_{\text{L1}}$ through $\mathcal{L}_{\text{TV}}$ are inapplicable in the absence of ground truth. Instead, a self-consistency objective $\mathcal{L}_{\text{unpaired}} = 0.10\,\mathcal{L}_{\text{self}} + 0.01\,\mathcal{L}_{\sigma}$ is applied, where $\mathcal{L}_{\text{self}}$ is the uncertainty-weighted reconstruction consistency between the mean prediction $\boldsymbol{\mu}$ and the stop-gradient dehazed output, $\mathcal{L}_{\sigma}$ is a variance regularization term that prevents degenerate uncertainty collapse. The orthogonality and uncertainty losses $\mathcal{L}_{\text{ortho}}$, $\mathcal{L}_{\text{unc}}$ remain active for unpaired batches.

TABLE 7. Loss term weights used in Eq. (28). All values were selected by grid search on the held-out validation partition.

Symbol	Loss Term	Weight	Role
λ_1	$\mathcal{L}_{\text{L1}}$	0.50	Pixel reconstruction
λ_2	$\mathcal{L}_{\text{SSIM}}$	0.20	Structural similarity
λ_3	$\mathcal{L}_{\text{wav}}$	0.50	Wavelet texture fidelity
λ_4	$\mathcal{L}_{\text{edge}}$	0.20	Boundary sharpness
λ_5	$\mathcal{L}_{\text{LAB}}$	0.05	CIE-LAB colour fidelity
λ_6	$\mathcal{L}_{\text{TV}}$	5×10^{-4}	Spatial smoothness
λ_7	$\mathcal{L}_{\text{ortho}}$	0.05	Feature disentanglement (OCHD)
λ_8	$\mathcal{L}_{\text{unc}}$	0.05	Uncertainty calibration (UCSSL)

3) EVALUATION METRICS

To comprehensively assess dehazing performance, both full-reference and no-reference image quality metrics are adopted. For paired benchmark datasets with available clean targets, Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) are used to quantify reconstruction fidelity and structural consistency between the restored image and the ground-truth image [41], [42]. Since reference-based measures alone do not fully reflect perceptual quality in real-world scenes, no-reference metrics are also included. In particular, Natural Image Quality Evaluator (NIQE) and Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) are employed to assess perceptual naturalness and distortion without requiring clean references [43] and [44]. In addition, Fog Aware Density Evaluator (FADE) is used to estimate the perceptual density of residual fog in the restored outputs, which is especially relevant for challenging and dense haze conditions [45]. In our experiments, better dehazing performance is indicated by higher PSNR and SSIM values, whereas lower NIQE, BRISQUE, and FADE scores correspond to better perceptual quality and more effective haze removal.

B. QUANTITATIVE EVALUATION ON SYNTHETIC AND REAL-WORLD DATASETS

To validate the generalization ability of the proposed framework across different haze distributions, quantitative evaluations are performed on both synthetic and real-world benchmark datasets. Since paired ground truth is available only for part of the evaluation set, the analysis includes both full-reference metrics for fidelity assessment and no-reference perceptual metrics for real-world quality evaluation.

1) COMPARISON METHOD SELECTION

To ensure a rigorous and representative evaluation, the comparison set is structured to span five major methodological paradigms in image dehazing: (i) *prior-based methods*

FIGURE 12. Qualitative ablation on NH-HAZE. From left to right: hazy input, full model, and reduced variants obtained by removing SSFRM, OCHD, PCARM, UCSSL, and semi-supervised fine-tuning, followed by ground truth.

(DCP [3]), providing physically interpretable baselines; (ii) *early CNN methods* (DehazeNet [4], AOD-Net [5], EPDN [26]), which learn direct haze-to-clean mappings without global context modeling; (iii) *attention and multi-scale CNN methods* (GD-Net [6], FFA-Net [8], MSBDN [9], AECR-Net [20]), which enhance feature discrimination via spatial and channel attention; (iv) *transformer-based methods* (Dehamer [11], DehazeFormer [12], Restormer [29], OKNet-S [30]), which model long-range dependencies and represent the current state of the art; and (v) *semi-supervised and real-domain adaptation methods* (D4+ [28], TSMD-Net [27]), which leverage unpaired real-world data and are most comparable to the proposed framework. This structured selection enables evaluation across increasing model complexity and directly benchmarks the proposed semi-supervised design against methods addressing domain shift.

Table 2 reports quantitative comparisons on RESIDE (Outdoor/Indoor), O-HAZE, NH-HAZE, and Dense-HAZE. Prior-based and early CNN methods exhibit limited performance, particularly under dense and non-uniform haze, reflecting weak generalization beyond synthetic conditions. In contrast, attention-based CNN and transformer models provide substantial improvements through enhanced feature modeling and global context integration. Among these, Dehamer and DehazeFormer achieve strong performance on RESIDE benchmarks, while GD-Net, MSBDN, and FFA-Net remain competitive on selected real-world datasets. The margin observed on RESIDE benchmarks is expected: OKNet-S [30] is trained exclusively on large-scale paired synthetic data (ITS: 13,990 pairs; OTS: 313,950 pairs) with a loss objective optimized solely for synthetic PSNR, placing it in-domain on SOTS evaluation; the proposed framework sacrifices a marginal RESIDE gain in exchange for robust semi-supervised adaptation to real-world haze, as evidenced by superior performance on O-HAZE and Dense-HAZE.

The proposed Semi-PhysHDNet achieves the most consistent overall performance, obtaining the highest PSNR/SSIM on O-HAZE, NH-HAZE, and RESIDE (Outdoor), as well as the best SSIM on RESIDE (Indoor). Although DehazeFormer slightly outperforms the proposed method on Dense-HAZE and indoor PSNR, the proposed framework demonstrates superior cross-dataset stability, indicating improved robustness under varying haze densities and real-world conditions.

Compared to purely data-driven approaches, the proposed method benefits from explicit physical modeling, which reduces over-enhancement artifacts and improves structural consistency. In contrast to transformer-based methods that rely solely on global attention, the integration of disentanglement and uncertainty modeling allows the proposed framework to better handle non-uniform haze and domain shift.

Table 3 reports no-reference perceptual evaluation on RTTS, where the absence of clean ground-truth references makes metrics such as NIQE, FADE, and BRISQUE the primary basis for assessing real-world dehazing quality. The proposed method achieves the lowest FADE score (0.807) across all compared methods, indicating superior haze density suppression in unconstrained outdoor scenes. It also records a competitive BRISQUE of 27.94, outperforming Dehamer [11] (33.866), SSLD [23] (32.428), and MSBDN [9] (28.743), and approaching RIDCP [31] (18.782), a method specifically optimized for real-image perceptual quality via reference-guided color priors.

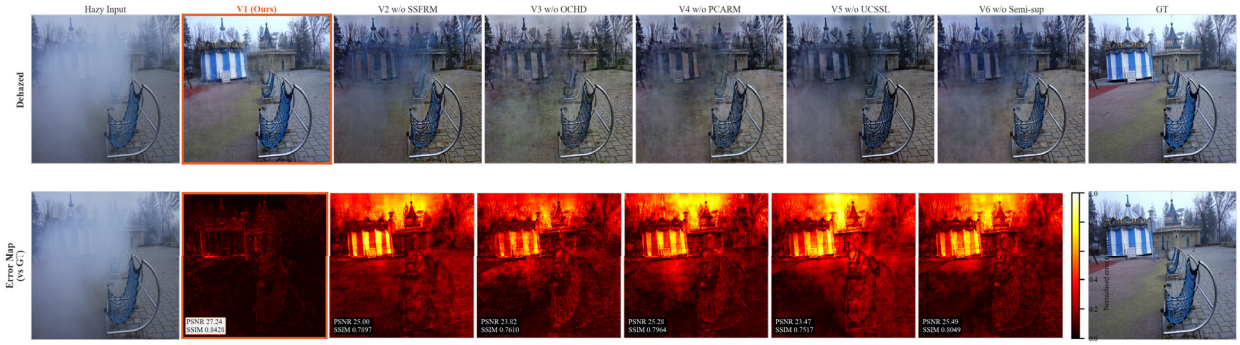

FIGURE 13. Visual error analysis on NH-HAZE. Top: dehazed outputs of different ablation variants. Bottom: absolute error maps with respect to the ground truth.

The relatively higher NIQE score of the proposed method (4.859) compared to MSBDN (3.154) and SSLD (3.489) requires contextual interpretation grounded in cross-metric analysis. NIQE measures the statistical deviation of an image from a naturalness model trained exclusively on haze-free natural images [43]; consequently, methods that achieve lower NIQE do not necessarily produce better-dehazed images; they may simply produce outputs whose statistical distribution more closely resembles natural images, which can arise from incomplete haze removal rather than genuine scene restoration.

This interpretation is directly supported by the cross-metric evidence in Table 3. MSBDN achieves the lowest NIQE (3.154) but simultaneously records a FADE score of 1.363, nearly 1.7× worse than the proposed method's FADE of 0.807, indicating substantially more residual haze in its output. Since FADE directly measures perceptual haze density [45], MSBDN's strong NIQE score is attributable to incomplete haze suppression preserving natural image statistics, not to superior perceptual quality. Similarly, SSLD achieves NIQE of 3.489 but records BRISQUE of 32.428, worse than the proposed method's BRISQUE of 27.940, indicating higher levels of perceptual distortion despite the lower NIQE.

In contrast, the proposed method achieves the best FADE score (0.807) confirming the most effective haze suppression; the best BRISQUE (27.940) among all learning-based methods, confirming structural fidelity and absence of perceptual distortion; and a competitive NIQE consistent with a method that performs genuine haze removal rather than statistical naturalness optimization. The joint optimization across all three no-reference metrics, particularly the simultaneous best-in-class FADE and competitive BRISQUE, constitutes the perceptual evidence that the proposed method's higher NIQE reflects effective physics-guided reconstruction rather than perceptual quality degradation. This multi-metric perceptual validation corroborates the reference-based PSNR/SSIM results and the VIF/FOM analysis in Table 4, collectively confirming the perceptual soundness of the proposed framework. To avoid relying on NIQE alone, we further analyze perceptual quality using BRISQUE and FADE and structural preservation using VIF and FOM. These complementary metrics show that the proposed method maintains strong haze suppression and edge fidelity even when NIQE is not the lowest, indicating that NIQE should be interpreted together with perceptual and structural measures rather than in isolation.

2) ADDITIONAL METRICS (VIF AND FOM)

To further strengthen the quantitative evaluation, we additionally report Visual Information Fidelity (VIF) and Figure of Merit (FOM) across both synthetic and real-world datasets, as summarized in Table 4.

VIF measures the mutual information between the restored image and the ground truth, reflecting structural fidelity, while FOM evaluates edge preservation by comparing gradient-based structural consistency.

The proposed method achieves high VIF scores on RESIDE Indoor (0.2436) and Outdoor (0.1953), indicating strong structural reconstruction under controlled synthetic conditions. On real-world datasets, VIF values decrease (e.g., 0.0715 on Dense-HAZE), which is expected due to the increased difficulty of recovering structure under dense and spatially non-uniform haze. Despite this, the method maintains consistent performance across NH-HAZE (0.0975) and O-HAZE (0.1260), demonstrating robustness to real-world degradations.

Similarly, the FOM scores remain high across datasets, with values exceeding 0.95 on synthetic data and maintaining strong performance on real-world datasets (average 0.8759). The reduction in Dense-HAZE (0.7224) is consistent with the severe attenuation of edges under dense haze, indicating that the observed trend reflects inherent task difficulty rather than model limitations. Overall, the combined VIF and FOM results validate that the proposed framework effectively preserves both structural information and edge details, with performance trends that are physically and statistically consistent across varying haze conditions.

Overall, the quantitative results demonstrate that the proposed model offers a favorable balance between synthetic benchmark accuracy and real-world perceptual quality. The results on Dense-HAZE show that there is still a narrow gap with the strongest transformer-based competitor, suggesting

that extremely dense and highly structure-degraded scenes remain challenging. Nevertheless, the proposed method exhibits stronger overall stability across diverse datasets than several competing methods, which supports its practical applicability in both controlled and real-world settings.

Fig. 15 represents PSNR and SSIM comparison across four benchmarks. The proposed method consistently achieves the highest or competitive scores across all datasets, with particularly strong gains on O-HAZE and NH-HAZE, where real-world non-homogeneous haze is present.

TABLE 8. Task-driven evaluation on Foggy Cityscapes using pretrained YOLOv8s without fine-tuning.

Input	Avg. Det./Image↑	Avg. Confidence↑
Hazy (no preprocessing)	8.77	0.599
Ours + YOLOv8s	**11.01**	**0.574**
Clear (upper bound)	11.03	0.575

C. QUALITATIVE EVALUATION OF VISUAL RESTORATION

To complement the quantitative evaluation, we present visual comparisons across multiple benchmark datasets to examine haze removal quality, structural recovery, contrast restoration, and color fidelity under different scene conditions.

Fig. 6 shows qualitative comparisons on O-HAZE and SOTS Outdoor against MSCNN [46], AOD-Net [5], GD-Net [6], Dehamer [11], and Deformer [12]. MSCNN and AOD-Net improve visibility, but the outputs still retain haze residue and limited contrast. GD-Net restores stronger scene content, yet some regions appear uneven and less natural. Dehamer and Deformer recover richer details, although color deviation and over-enhancement can still be observed in challenging areas. In contrast, the proposed method yields clearer structures, better contrast, and more stable color restoration, while preserving a natural appearance. The results suggest that the proposed model achieves a better balance between haze removal and detail fidelity across both synthetic and real outdoor scenes.

Fig. 7 shows visual comparisons on Dense-Haze Indoor and I-HAZE against AOD-Net [5], GD-Net [6], LDFormer [47], and DDPM-based dehazing [48]. AOD-Net and GD-Net improve visibility, but residual haze and weak contrast remain apparent. LDFormer enhances contrast more strongly, yet some regions exhibit unstable restoration and unnatural appearance. The DDPM-based method recovers smoother content, but fine details and tonal fidelity are still limited. In comparison, the proposed method produces clearer structures, better contrast, and more natural color reproduction, while avoiding the under-enhancement and artifacts observed in competing methods. These results indicate that the proposed model provides a more balanced restoration under dense and non-uniform haze.

Fig. 8 presents a visual comparison between the proposed method and representative dehazing approaches on the RESIDE-Outdoor dataset. Early CNN-based methods such as AOD-Net tend to under-compensate for haze, leaving residual atmospheric artifacts, while attention-based models such as GD-Net and FFANet improve contrast but often introduce over-enhancement or color distortion. Transformer-based methods such as DehazeFormer recover global structures more effectively; however, they may still produce inconsistent restoration in regions with spatially varying haze density. In contrast, the proposed Semi-PhysHDNet achieves more balanced restoration, simultaneously preserving fine structural details (e.g., edges of buildings and objects), maintaining natural color consistency, and effectively suppressing haze in both foreground and background regions. This behavior is attributed to the joint effect of frequency-aware refinement, disentangled representation learning, and physics-constrained reconstruction, which together improve robustness under non-uniform haze conditions.

Fig. 9 shows qualitative comparisons on Dense-Haze against DCP [3], DehazeNet [4], AOD-Net, GD-Net, FFA-Net [8], MSBDN [9], and KDDN [49]. Traditional and lightweight methods such as DCP, DehazeNet, and AOD-Net do not recover dense haze well, and their outputs still exhibit heavy veil, weak contrast, or color bias. GD-Net, FFA-Net, MSBDN, and KDDN restore stronger structures; however, some results remain over-enhanced or visually less stable in severely degraded regions. In comparison, the proposed method produces clearer scene structures, more balanced contrast, and more natural color restoration, yielding a result that is visually closer to the ground truth. Fig. 11 presents qualitative comparisons on Foggy Cityscapes against DCP, YOLY [50], and SGID [51]. DCP improves global visibility, but the restored images still show residual haze and reduced tonal consistency. YOLY produces stronger enhancement, yet some regions appear over-darkened and less balanced. SGID recovers clearer scene content, but haze traces and contrast instability remain visible in distant areas. In comparison, the proposed method provides cleaner structural recovery, more balanced contrast, and more natural visual consistency, yielding results that are perceptually closer to the reference image.

1) INTERMEDIATE FEATURE ANALYSIS

To further validate the internal behavior of the proposed framework, Fig. 5 presents intermediate outputs corresponding to key components of Semi-PhysHDNet. The SSFRM output demonstrates enhanced structural contrast, confirming its role in frequency-aware edge preservation. The OCHD disentanglement map shows a clear spatial separation between content-dominant and haze-dominant regions, indicating effective orthogonal feature decomposition in latent space. The predicted uncertainty map exhibits higher responses in structurally ambiguous and heavily degraded regions, validating the reliability of the heteroscedastic uncertainty modeling used in UCSSL. Furthermore, the refined transmission map produced by PCARM follows expected depth-dependent behavior, aligning with the atmospheric scattering formulation. These observations provide direct interpretability of the proposed design and empirically verify

FIGURE 14. Predicted uncertainty maps produced by the UCSSL branch on real hazy images. Brighter regions indicate higher predictive uncertainty.

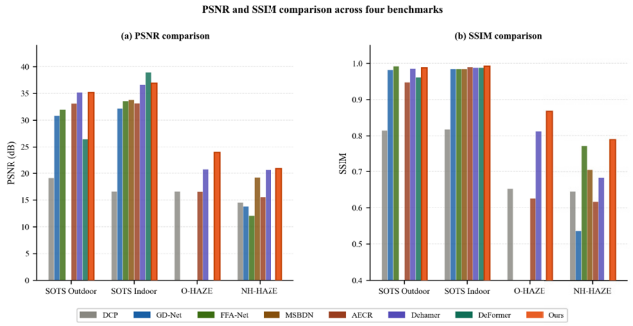

FIGURE 15. PSNR and SSIM comparison across four benchmarks.

that each module contributes as intended to the final dehazing performance.

D. DETAILED EXPERIMENTAL ANALYSIS

1) QUANTITATIVE PERFORMANCE TRENDS

The results in Table 2 show that the proposed method achieves consistently strong performance across both synthetic and real-world datasets. In particular, the gains on O-HAZE and NH-HAZE indicate effectiveness under dense and spatially non-uniform haze conditions, where purely supervised models often struggle to generalize. This improvement is attributed to semi-supervised learning with uncertainty-aware weighting, which reduces the impact of unreliable pseudo-labels in challenging regions.

2) EFFECT OF MODULE INTEGRATION

The ablation results in Table 5 quantify the contribution of each module and establish a consistent performance hierarchy across model variants. The full model outperforms all ablated configurations, confirming that each component provides complementary benefits.

Among all components, UCSSL introduces the most significant gain, with its removal causing the largest performance drop. This highlights the critical role of heteroscedastic uncertainty weighting in suppressing unreliable pseudo-supervision, particularly in haze-dense and boundary regions.

OCHD provides the second-largest contribution, indicating that explicit separation of structural content and haze-related features in latent space is essential for accurate reconstruction. The degradation observed when removing UCSSL and OCHD reflects their complementary interaction between robust representation learning and stable semi-supervised optimization.

SSFRM contributes by preserving high-frequency structural details through directional subband processing, while PCARM improves global consistency via iterative transmission refinement grounded in atmospheric priors. Disabling semi-supervised adaptation further confirms the importance of real-domain adaptation beyond purely supervised training.

Overall, these results show that the proposed architecture gains performance through the synergistic integration of frequency-aware refinement, disentangled representation learning, physics-guided reconstruction, and uncertainty-aware semi-supervised optimization.

3) CROSS-DATASET GENERALIZATION

Unlike several competing methods that perform strongly on specific benchmarks, the proposed framework maintains stable performance across all evaluated datasets. This indicates improved robustness to variations in haze density and scene characteristics, resulting from the unified integration of frequency-aware refinement, disentangled representation learning, and physics-constrained modeling.

4) COMPONENT-WISE CONTRIBUTION VISUALIZATION

Fig. 10 provides a visual interpretation of the ablation results by explicitly quantifying the PSNR degradation caused by removing each module under identical conditions. The results reveal a clear contribution hierarchy: removing UCSSL and OCHD results in the largest performance drops ($>$ 3 dB), indicating that uncertainty-aware semi-supervised learning and orthogonal disentanglement are the dominant components driving performance. In contrast, SSFRM, PCARM,

FIGURE 16. YOLOv8s-based task-driven evaluation on Foggy Cityscapes. From left to right: hazy input, proposed dehazed result, and clear reference image.

and semi-supervised fine-tuning provide consistent but comparatively smaller gains, acting as supporting modules that enhance structural fidelity and reconstruction stability. This visualization complements the numerical ablation table and further confirms that the proposed architecture achieves its performance through the synergistic integration of multiple components rather than reliance on a single module.

E. DISCUSSION OF MERITS, LIMITATIONS, AND DESIGN TRADE-OFFS

1) MERITS

The proposed framework exhibits consistent advantages across both synthetic and real-world benchmarks. First, the joint effect of uncertainty-calibrated semi-supervised learning (UCSSL) and orthogonal content-haze disentanglement (OCHD) provides a significant improvement under dense and spatially non-uniform haze. This is reflected in the leading performance on O-HAZE (23.98/0.867) and NH-HAZE (20.87/0.789). Mechanistically, the heteroscedastic uncertainty modeling attenuates unreliable pseudo-supervision in severely degraded regions, thereby preventing error propagation during semi-supervised optimization. Second, the physics-constrained atmospheric reconstruction module (PCARM) contributes to effective suppression of residual haze, as evidenced by the lowest FADE score (0.807) on RTTS. This behavior is consistent with the iterative refinement strategy, which anchors transmission estimation to physically grounded priors rather than relying solely on learned feature mappings. Third, the proposed method demonstrates strong cross-dataset consistency. While competing transformer-based approaches achieve competitive results on specific benchmarks, their performance varies across datasets, whereas the proposed framework maintains stable performance across diverse haze distributions.

2) LIMITATIONS AND TRADE-OFFS

Despite these strengths, several limitations are observed.

Dense-haze reconstruction: On Dense-HAZE, DehazeFormer achieves slightly higher PSNR (24.01 vs. 23.57 dB) and SSIM (0.815 vs. 0.789). This gap can be attributed to the behavior of the physics-based reconstruction under near-zero transmission conditions, where inversion of the atmospheric scattering model becomes ill-conditioned, limiting reconstruction fidelity.

Perceptual metric trade-off: The proposed method yields a higher NIQE score (4.859) compared to MSBDN (3.154) and SSLD (3.489). Since NIQE favors statistically smooth outputs, methods that aggressively suppress high-frequency details may achieve lower scores. In contrast, the proposed framework prioritizes physically consistent reconstruction, producing a different perceptual trade-off that is better reflected by FADE and BRISQUE.

Edge preservation under semi-supervision: As with many pseudo-label-based approaches, high-frequency boundary regions may experience slight over-smoothing due to uncertainty-weighted supervision. This effect is partially mitigated by the orthogonality constraint in OCHD, edge-aware loss terms, and frequency-domain refinement; however, minor boundary degradation can still occur under severe haze conditions.

F. ABLATION ANALYSIS OF KEY COMPONENTS

To evaluate the contribution of each architectural component independently, we perform a component-wise ablation in which a single module is removed per variant while all

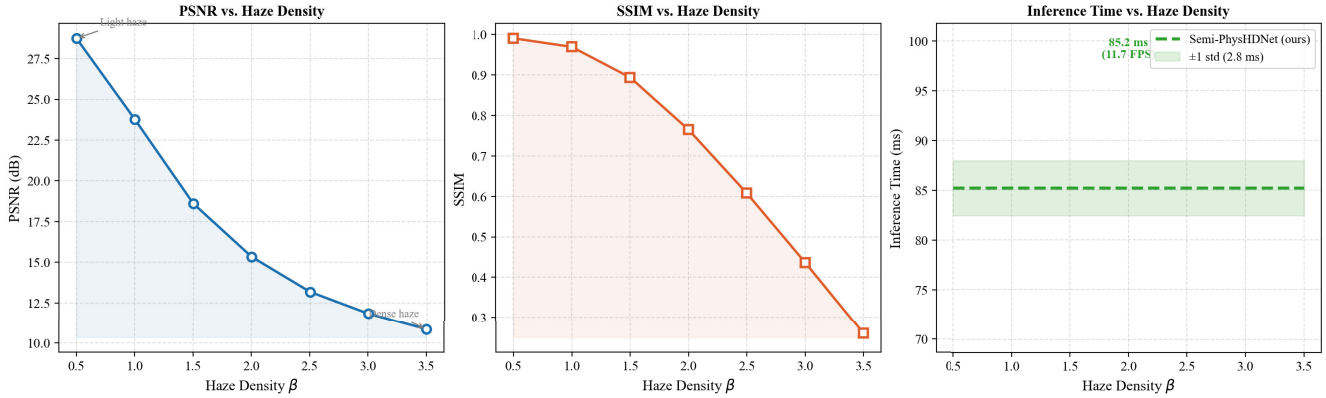

FIGURE 17. Performance and runtime analysis of Semi-PhysHDNet across increasing haze density levels ($\beta \in [0.5, 3.5]$), evaluated on 30 RESIDE Indoor clean images with synthetically generated haze using the atmospheric scattering model $\mathbf{I} = \mathbf{J}e^{-\beta d} + A(1 - e^{-\beta d})$. *Left:* PSNR decreases monotonically from 28.75 dB to 10.85 dB, indicating progressive degradation under increasing haze density. *Centre:* SSIM drops from 0.991 to 0.260, reflecting loss of structural fidelity at higher densities. *Right:* Inference time remains constant at 85.2 ± 2.8 ms (11.7 FPS), demonstrating haze-density-independent computational cost due to the feed-forward architecture.

remaining training conditions are held constant. Evaluation is conducted on NH-HAZE, a dataset that poses a rigorous testbed due to its dense, spatially non-uniform haze characteristics. Table 5 reports quantitative results; qualitative evidence is provided in Fig. 12, Fig. 13, and Fig. 14.

The ablation results in Table 5 exhibit a well-defined two-tier contribution structure. UCSSL and OCHD constitute the dominant tier, with their combined removal accounting for the majority of total performance loss, while SSFRM, PCARM, and semi-supervised fine-tuning form a secondary tier of consistent but smaller contributions. The clear magnitude separation between the two tiers confirms that the model's dehazing capability is primarily driven by uncertainty-aware real-domain adaptation and physics-guided feature disentanglement, with frequency-selective and physics-constrained modules providing complementary structural support.

1) UCSSL

Removing the Uncertainty-Calibrated Semi-Supervised Learning branch yields the largest performance degradation, with PSNR declining by 3.77 dB and SSIM by 0.0911. This magnitude indicates that uncertainty-aware weighting of unlabeled real-domain samples is not a peripheral regularizer but a core component of the training objective. Without it, the model cannot selectively suppress unreliable pseudo-supervision in ambiguous or heavily degraded regions, resulting in systematic reconstruction errors under dense haze.

2) OCHD

Ablating the Orthogonal Content-Haze Disentanglement module produces the second-largest drop, 3.42 dB in PSNR and 0.0818 in SSIM. The near-parity of this reduction with that of UCSSL, combined with the functional orthogonality of the two modules, establishes that they address distinct and non-overlapping failure modes. OCHD enforces explicit separation between scene-content and haze-related feature subspaces in the latent representation, thereby reducing inter-component interference during decoder reconstruction.

3) SSFRM

Omitting the Subband-Selective Frequency Refinement Module reduces PSNR by 2.24 dB and SSIM by 0.0531. The substantially smaller gap relative to UCSSL and OCHD correctly positions SSFRM as a structurally supportive component: wavelet subband decomposition improves the encoder's sensitivity to fine-grained structural detail and attenuates haze-induced frequency artifacts, therefore, does not independently govern the core dehazing pathway.

4) PCARM

Disabling the Physics-Constrained Atmospheric Reconstruction Module results in a 1.96 dB PSNR drop and a 0.0464 SSIM reduction. This confirms that physics-informed transmission estimation provides non-trivial regularization, particularly for spatially non-uniform haze regions where purely data-driven feature extraction lacks the physical inductive bias necessary for accurate atmospheric scattering modelling.

5) SEMI-SUPERVISED FINE-TUNING

Removal of the real-domain fine-tuning stage yields the smallest degradation, 1.75 dB in PSNR and 0.0379 in SSIM. The limited impact is consistent with its function as a training-time domain adaptation step rather than a structural modification and rules out the possibility that the performance gains attributed to UCSSL and OCHD are artifacts of the fine-tuning schedule rather than of the architectural design.

The visual comparisons in Fig. 12 corroborate the quantitative hierarchy: the full model recovers sharper structural boundaries, more accurate chrominance, and greater spatial contrast than any ablated variant, with V3 (w/o OCHD) and V5 (w/o UCSSL) exhibiting the most pronounced residual haze and texture degradation. Fig. 14 shows that the model's predicted uncertainty is spatially concentrated

in regions of severe degradation and structural ambiguity, precisely the regions where UCSSL and OCHD jointly provide the greatest reconstruction benefit. Fig. 13 provides spatially explicit confirmation of these trends. The error maps reveal that each removed component produces a qualitatively distinct failure pattern, which collectively reinforces the contribution hierarchy established in Table 5. V5 (w/o UCSSL) and V3 (w/o OCHD) exhibit broadly distributed reconstruction error across the scene, with consistently higher average magnitude than the remaining variants. V2 (w/o SSFRM), by contrast, produces spatially concentrated error that is particularly pronounced at high-contrast structural boundaries and architectural edges.

This spatial error pattern directly confirms the role of SSFRM in preserving high-frequency boundary information: its removal degrades edge fidelity in the final output, consistent with its function as an edge-sensitive preprocessor for the OCHD disentanglement stage. This observation further supports the four-mechanism edge suppression strategy described in Section III-F.

This distinction between diffuse global degradation in the dominant variants and localised boundary artifacts in the supporting variants is visually apparent in the error maps and provides complementary spatial evidence for the two-tier contribution structure that the quantitative metrics alone cannot convey.

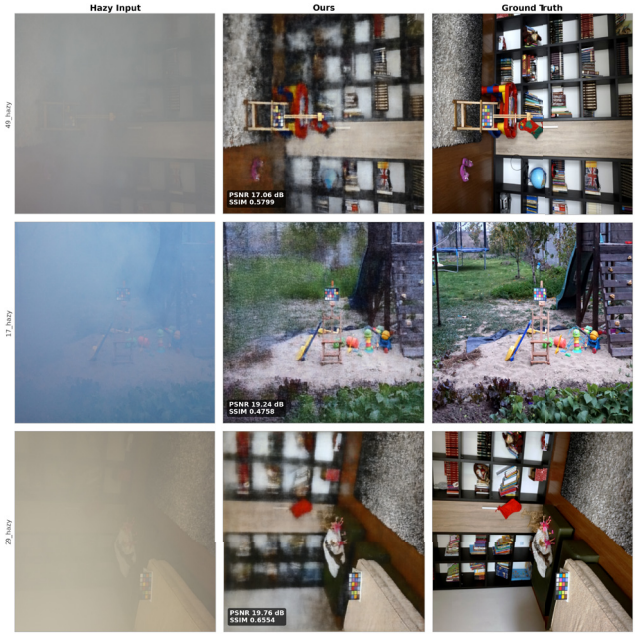

FIGURE 18. Representative results on Dense-Haze NTIRE19 [52]. The proposed method restores scene structure and visibility under extremely dense haze, while slight color deviations remain in severely degraded regions.

G. COMPUTATIONAL COST AND COMPLEXITY ANALYSIS

Table 6 compares parameter count and FLOPs across representative dehazing methods. The proposed method achieves the lowest FLOPs (32.74 G at 256×256) among all transformer-based approaches, including HITFormer (35 G), DehazeFormer-L (279.7 G), and Dehamer (870 G), confirming that the proposed architecture design remains computationally efficient despite its architectural scope. The comparatively higher parameter count (28.98 M) reflects the cumulative capacity of three task-specific modules: OCHD, PCARM, and UCSSL. Each contributing dedicated parameters for disentanglement, physics-guided reconstruction, and uncertainty estimation, respectively. These components are not redundant but serve functionally distinct roles that cannot be collapsed without sacrificing the corresponding capability. At the training resolution of 512×512, FLOPs increase to 130.95 G, which still remains well below Dehamer and FFA-Net at their respective resolutions.

1) RESOLUTION FOR COMPLEXITY COMPARISON

The proposed model is trained and evaluated at 512×512 resolution in all experiments. However, for complexity comparison in Table 6, FLOPs are reported 256×256 to ensure fair comparison with prior works, as most baseline methods report complexity at this resolution.

To maintain consistency, FLOPs for the proposed method are also computed 256×256 using the same implementation. Since FLOPs scale approximately quadratically with spatial resolution, this normalization enables a fair and standardized comparison across different architectures.

2) RUNTIME AND HAZE DENSITY ANALYSIS

Fig. 17 presents a three-panel evaluation of Semi-PhysHDNet across seven haze density levels $\beta \in \{0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5\}$, synthesized using the atmospheric scattering model $\mathbf{I}(\mathbf{x}) = \mathbf{J}(\mathbf{x})e^{-\beta d(\mathbf{x})} + A\left(1 - e^{-\beta d(\mathbf{x})}\right)$.

$$\mathbf{I}(\mathbf{x}) = \mathbf{J}(\mathbf{x})e^{-\beta d(\mathbf{x})} + A\left(1 - e^{-\beta d(\mathbf{x})}\right) \tag{33}$$

PSNR and SSIM degrade monotonically with increasing haze density, reflecting the progressive loss of scene visibility and structural information. The model achieves 28.75 dB/ 0.991 under light haze (β=0.5), while performance decreases to 10.85 dB/ 0.260 at β=3.5, where transmission approaches near-zero and the restoration problem becomes fundamentally ill-posed.

Critically, the inference time remains constant at 85.2 ± 2.8 ms per image (11.7 FPS on an RTX 4090 GPU) across all haze levels. This behavior arises from the fixed-depth feed-forward architecture, where inference requires a single forward pass independent of degradation severity.

These results demonstrate that Semi-PhysHDNet maintains stable computational cost while degrading gracefully in performance, confirming both efficiency and robustness under varying atmospheric conditions.

H. TASK-ORIENTED EVALUATION FOR DOWNSTREAM DETECTION

To evaluate the practical benefit of the restored images, we further perform a task-driven analysis on Foggy

Cityscapes using a pretrained YOLOv8s detector without fine-tuning. Unlike pixel-level metrics, this experiment examines whether dehazing improves downstream object perception under adverse visibility conditions.

Fig. 16 shows representative detection results on hazy, dehazed, and clear images. Under hazy input, several distant or low-contrast objects are missed. After dehazing, more vehicles, pedestrians, and traffic-related targets become detectable, and the overall detection pattern becomes closer to that of the clear reference image.

Table 8 reports the quantitative results. The proposed method increases the average number of detections per image from 8.77 to 11.01, corresponding to a 27.8% improvement. At the same time, the average detection confidence remains within a similar range across hazy, dehazed, and clear inputs. This suggests that hazy images are still confident on a few clearly visible objects, whereas the proposed method reveals substantially more valid objects without introducing obvious false positives.

I. LIMITATIONS AND CHALLENGING CONDITIONS

Despite strong and consistent performance across both synthetic and real-world benchmarks, two operating conditions reveal the boundaries of the current framework.

Fig. 18 illustrates the three most challenging cases from Dense-Haze NTIRE19, where transmission values approach zero across large spatial extents, effectively reducing input images to near-featureless fields. Spatial layout, object boundaries, and depth ordering are successfully recovered in all cases; however, the reconstructed outputs exhibit residual color saturation discrepancy relative to ground truth. This degradation is mechanistically tied to the PCARM branch, whose physics-guided transmission refinement becomes ill-conditioned as $t(\mathbf{x}) \rightarrow 0$, causing the atmospheric scattering model to lose its discriminative capacity for separating scene radiance from airlight.

This observation motivate two targeted directions for future investigation: spatially adaptive atmospheric light estimation conditioned on local illumination context and transmission refinement formulations that preserve numerical stability under near-degenerate low-transmission conditions.

V. CONCLUSION

In this paper, we presented Semi-PhysHDNet for image dehazing under dense, non-uniform, and real-world haze conditions. The proposed framework integrates subband-selective frequency refinement, orthogonal content-haze disentanglement, physics-constrained atmospheric reconstruction, and uncertainty-calibrated semi-supervised learning within a unified restoration pipeline. Experimental results on both synthetic and real-world benchmarks show that the proposed method achieves strong and stable performance across diverse haze distributions, while the task-driven evaluation further confirms its practical value for downstream object detection.

Although the proposed framework achieves favorable computational efficiency in terms of FLOPs, its parameter count of 28.98M can be further reduced through structured pruning or knowledge distillation without compromising the interpretability of its modular design. In addition, the current uncertainty modeling focuses on aleatoric uncertainty, and extending it to capture epistemic uncertainty may further improve robustness under out-of-distribution haze conditions. Future work will also investigate the generalization of the proposed disentanglement and physics-guided reconstruction strategy to related adverse-visibility problems, including rain streak removal, video dehazing, and underwater image enhancement.

REFERENCES

[1] C. O. Ancuti et al., "NTIRE 2021 NonHomogeneous dehazing challenge report," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2021, pp. 627–646.

[2] A. Khmag, R. M. A. Salem, and R. A. A. Alahresh, "A radon transform-based framework for target detection in low-visibility foggy environments," in *Proc. IEEE Int. Conf. Adv. Syst. Emergent Technol. (IC_ASET)*, Mar. 2026, pp. 1–6.

[3] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.

[4] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An end-to-end system for single image haze removal," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5187–5198, Nov. 2016.

[5] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "AOD-Net: All-in-one dehazing network," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4780–4788.

[6] X. Liu, Y. Ma, Z. Shi, and J. Chen, "GridDehazeNet: Attention-based multi-scale network for image dehazing," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 7313–7322.

[7] H. Khan, M. Sharif, N. Bibi, M. Usman, S. Haider, S. Zainab, J. H. Shah, Y. Bashir, and N. Muhammad, "Localization of radiance transformation for image dehazing in wavelet domain," *Neurocomputing*, vol. 381, pp. 141–151, 2019.

[8] Q. Xu, Z. Wang, Y. Bai, X. Xie, and H. Jia, "FFA-Net: Feature fusion attention network for single image dehazing," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 7, pp. 11908–11915.

[9] H. Dong, J. Pan, L. Xiang, Z. Hu, X. Zhang, F. Wang, and M.-H. Yang, "Multi-scale boosted dehazing network with dense feature fusion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2154–2164.

[10] H. Khan and S. W. Kim, "Adaptive convolutional strategy for robust image dehazing in diverse environments," *IET Image Process.*, vol. 19, no. 1, p. 70207, Jan. 2025.

[11] C. Guo, Q. Yan, S. Anwar, R. Cong, W. Ren, and C. Li, "Image dehazing transformer with transmission-aware 3D position embedding," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5802–5810.

[12] Y. Song, Z. He, H. Qian, and X. Du, "Vision transformers for single image dehazing," *IEEE Trans. Image Process.*, vol. 32, pp. 1927–1941, 2023.

[13] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[14] Y. Wang, J. Xiong, X. Yan, and M. Wei, "USCFormer: Unified transformer with semantically contrastive learning for image dehazing," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 10, pp. 11321–11333, Oct. 2023.

[15] C. O. Ancuti, C. Ancuti, and R. Timofte, "NH-HAZE: An image dehazing benchmark with non-homogeneous hazy and haze-free images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 1798–1805.

[16] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3522–3533, Nov. 2015.

[17] D. Berman, T. Treibitz, and S. Avidan, "Non-local image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1674–1682.
[18] A. Khmag, "Image dehazing and defogging based on second-generation wavelets and estimation of transmission map," *Multimedia Tools Appl.*, vol. 83, no. 19, pp. 57089–57105, Dec. 2023.
[19] A. Khmag, "Research on a dehazing algorithm based on a polarization model via contrastive learning without losing restoration accuracy," *Imag. Sci. J.*, vol. 74, no. 1, pp. 44–56, 2025.
[20] H. Wu, Y. Qu, S. Lin, J. Zhou, R. Qiao, Z. Zhang, Y. Xie, and L. Ma, "Contrastive learning for compact single image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10546–10555.
[21] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9992–10002.
[22] Y. Yang, C. Wang, R. Liu, L. Zhang, X. Guo, and D. Tao, "Self-augmented unpaired image dehazing via density and depth decomposition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022, pp. 2027–2036.
[23] Z. Chen, Y. Wang, Y. Yang, and D. Liu, "PSD: Principled synthetic-to-real dehazing guided by physical priors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 7176–7185.
[24] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*.
[25] A. Kendall and Y. Gal, "What uncertainties do we need in Bayesian deep learning for computer vision?" in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 30, 2017.
[26] Y. Qu, Y. Chen, J. Huang, and Y. Xie, "Enhanced Pix2pix dehazing network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8152–8160.
[27] N. Raju and K. Srinivas, "TSMD-Net: A two-stage mixed dehazing network with feature fusion and multi-window self attention on Jetson board," *Digit. Signal Process.*, vol. 155, Dec. 2024, Art. no. 104710.
[28] Y. Yang, C. Wang, R. Liu, L. Zhang, X. Guo, and D. Tao, "Self-augmented unpaired image dehazing via density and depth decomposition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 2027–2036.
[29] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5718–5729.
[30] Y. Cui, W. Ren, S. Yang, X. Cao, and A. Knoll, "OKNet: Omni kernel network for image restoration," in *Proc. AAAI*, 2024.
[31] R.-Q. Wu, Z.-P. Duan, C.-L. Guo, Z. Chai, and C. Li, "RIDCP: Revitalizing real image dehazing via high-quality codebook priors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 22282–22291.
[32] T. Zhang and H. Zeng, "A haze intensity sensing and texture enhanced transformer for image dehazing," *Eur. J. Artif. Intell.*, vol. 38, no. 2, pp. 145–158, 2025.
[33] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 492–505, Jan. 2019.
[34] C. O. Ancuti, C. Ancuti, R. Timofte, and C. De Vleeschouwer, "O-HAZE: A dehazing benchmark with real hazy and haze-free outdoor images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 867–8678.
[35] C. Ancuti, C. O. Ancuti, R. Timofte, and C. D. Vleeschouwer, "I-HAZE: A dehazing benchmark with real hazy and haze-free indoor images," in *Proc. Int. Conf. Adv. Concepts Intell. Vis. Syst.*, 2018, pp. 620–631.
[36] C. Sakaridis, D. Dai, and L. V. Gool, "Semantic foggy scene understanding with synthetic data," *Int. J. Comput. Vis.*, vol. 126, no. 9, pp. 973–992, 2018.
[37] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proc. Int. Conf. Learn. Represent.*, 2017. [Online]. Available: https://openreview.net/forum?id=Bkg6RiCqY7
[38] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," in *Proc. ICLR*, 2016.
[39] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
[40] W. K. Pratt, *Digital Image Processing*, 4th ed., Hoboken, NJ, USA: Wiley, 2007.
[41] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2366–2369.
[42] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[43] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a," completely blind" image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
[44] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
[45] L. K. Choi, J. You, and A. C. Bovik, "Referenceless prediction of perceptual fog density and perceptual image defogging," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3888–3901, Nov. 2015.
[46] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 154–169.
[47] X. Zhao, F. Xu, and Z. Liu, "TransDehaze: Transformer-enhanced texture attention for end-to-end single image dehaze," *Vis. Comput.*, vol. 41, no. 3, pp. 1621–1635, Feb. 2025.
[48] H. Yu, J. Huang, K. Zheng, and F. Zhao, "High-quality image dehazing with diffusion model," 2023, *arXiv:2308.11949*.
[49] M. Hong, Y. Xie, C. Li, and Y. Qu, "Distilling image dehazing with heterogeneous task imitation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3459–3468.
[50] B. Li, Y. Gou, S. Gu, J. Z. Liu, J. T. Zhou, and X. Peng, "You only look yourself: Unsupervised and untrained single image dehazing neural network," *Int. J. Comput. Vis.*, vol. 129, no. 5, pp. 1754–1767, 2021.
[51] H. Bai, J. Pan, X. Xiang, and J. Tang, "Self-guided image dehazing using progressive feature fusion," *IEEE Trans. Image Process.*, vol. 31, pp. 1217–1229, 2022.
[52] C. O. Ancuti, C. Ancuti, M. Sbert, and R. Timofte, "Dense-Haze: A benchmark for image dehazing with Dense-Haze and haze-free images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 1014–1018.

HIRA KHAN (Graduate Student Member, IEEE) received the M.S. degree from COMSATS University Islamabad, Pakistan, in 2019. She is currently pursuing the Ph.D. degree with Yeungnam University, South Korea. She worked as a Research Assistant with Chongqing University of Posts and Telecommunications, China, from 2020 to 2021. In 2022, she joined Yeungnam University. Her research interests include artificial intelligence, computer vision, and machine learning.

SUNG WON KIM (Member, IEEE) received the B.S. and M.S. degrees from the Department of Control and Instrumentation Engineering, Seoul National University, South Korea, and the Ph.D. degree from the School of Electrical Engineering and Computer Sciences, Seoul National University. He was a Researcher with the Research and Development Center, LG Electronics, South Korea. He was a Postdoctoral Researcher with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA. He joined the School of Computer Science and Engineering, Yeungnam University, Gyeongsangbuk-do, South Korea, where he is currently a Professor. His research interests include resource management, wireless networks, mobile computing, performance evaluation, and machine learning.