PASDiff: Physics-Aware Semantic Guidance for Joint Real-World Low-Light Face Enhancement and Restoration

Yilin Ni[1,⋆], Wenjie Li[2,⋆], Zhengxue Wang[3], Juncheng Li[4], Guangwei Gao[3,†], and Jian Yang[3]

[1] College of Automation, Nanjing University of Posts and Telecommunications, Nanjing, China
[2] School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing, China
[3] PCA Lab, School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China
[4] School of Computer Science and Technology, East China Normal University, Shanghai, China
{nixiaolin26,lewj2408}@gmail.com, gwgao@njust.edu.cn

Abstract. Face images captured in real-world low light suffer multiple degradations—low illumination, blur, noise, and low visibility, etc. Existing cascaded solutions often suffer from severe error accumulation, while generic joint models lack explicit facial priors and struggle to resolve clear face structures. In this paper, we propose PASDiff, a Physics-Aware Semantic Diffusion in a training-free manner. To achieve a plausible illumination and color distribution, we leverage inverse intensity weighting and Retinex theory to introduce photometric constraints, thereby reliably recovering visibility and natural chromaticity. To faithfully reconstruct facial details, our Style-Agnostic Structural Injection (SASI) extracts structures from an off-the-shelf facial prior while filtering out its intrinsic photometric biases, seamlessly harmonizing identity features with physical constraints. Furthermore, we construct WildDark-Face, a real-world benchmark of 700 low-light facial images with complex degradations. Extensive experiments demonstrate that PASDiff significantly outperforms existing methods, achieving a superior balance among natural illumination, color recovery, and identity consistency. Code and dataset will be available at https://github.com/IVIPLab/PASDiff.

Keywords: Low-Light Enhancement · Real-world Face Restoration · Diffusion Models · Training-Free Manner

1 Introduction

Capturing high-quality facial images in real-world low-light scenarios, such as surveillance and handheld photography, remains a formidable challenge. To com-

⋆Equal contribution.
†Corresponding author: Guangwei Gao (gwgao@njust.edu.cn)

Fig. 1: Visual comparisons on degraded synthetic and real-world data. (a) Input. Cascaded paradigms (b) L-Diff [20]→DiffBIR [35] and (c) DiffBIR [35]→L-Diff [20] suffer from severe noise amplification and structural collapse, respectively. Generic joint models like (d) DarkIR [12] and (e) FDN [40] struggle with complex degradations, leading to residual blur and detail loss. (f) Our PASDiff achieves superior perceptual quality with crisp facial details and natural illumination, effectively suppressing artifacts. (g) Ground Truth. (The real-world dataset lacks GT references.)

pensate for insufficient illumination, imaging systems are forced to utilize high ISO and prolonged exposure. This physical trade-off inevitably leads to a compound degradation, where low visibility, severe noise, and blur coexist. Such coupled impairments not only deteriorate visual quality but also severely hinder downstream applications like face recognition, which demand the recovery of faithful facial identities and expressions.

Previous strategies typically decompose this task into two processing tasks, i.e., Low-Light Image Enhancement (LLIE) [1,2,47,50] and Blind Face Restoration (BFR) [28,35,41,43]. However, these methods rely on independent assumptions tailored to their specific sub-problems. Consequently, a naive cascading of these independent modules cannot solve the joint degradation. Specifically, as shown in Fig. 1(b) and (c), enhancing visibility first blindly amplifies latent noise, which downstream BFR models misinterpret as facial textures, leading to unnatural hallucinations. Conversely, restoring before low-light enhancing deprives BFR models of essential structural cues hidden in the darkness, resulting in irreversible over-smoothing. Alternatively, existing end-to-end approaches [12,37,40,49,58,59] attempt to address joint degradation holistically. While achieving promising results on synthetic datasets constructed with simple noise models and uniform blur kernels, they struggle to generalize to real-world nighttime scenes. The non-linear response of sensors in extreme darkness, coupled with the intricate geometry of human faces, renders real-world degradation patterns far more complex than synthetic simulations. As observed in Fig. 1(d) and (e), when applied to real-world captures, these generic models often fail to recover fine-grained facial components or suffer from residual blur, highlighting the urgent need for a robust, face-specific solution.

To effectively contend with complex real-world degradations and circumvent the inherent domain gap of synthetic data-driven methods, Diffusion Proba-

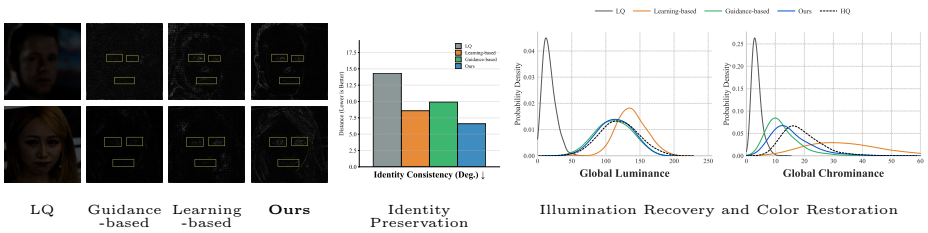

Fig. 2: From the phase reconstructions and statistical metrics, it can be seen that end-to-end learning strategies [35] inherently adhere to the degraded intensity distribution, suffering from compromised facial identities and severe color shifts. Existing guidance approaches [36] correct global illumination and chromaticity, but lack semantic guidance for fine geometries. In contrast, PASDiff elegantly integrates physical and structural guidance to restore crisp textures, natural illumination, and a color distribution better aligned with the high-quality reference, effectively preserving facial identity.

bilistic Models (DPMs) offer a promising avenue. Their robust generative priors provide the possibility of restoring high-fidelity details from severely degraded inputs. A straightforward solution for Joint Low-light Enhancement and Blind Face Restoration (Joint LL-BFR) is to directly fine-tune an existing face restoration diffusion [35] on paired data. However, our empirical pilot study reveals a fundamental bottleneck in this end-to-end paradigm. Under extreme low-light conditions, the network is overwhelmed by the "dual burden" of simultaneously rectifying drastic photometric shifts and hallucinating missing facial geometries. As shown in Fig. 2, the model inherently adheres to the degraded intensity distribution, either failing to reconstruct fine-grained facial structures or suffering from uncontrollable color shifts and catastrophic identity loss. In contrast, existing guidance approaches [36], which rely primarily on global physical constraints to steer the sampling, suffer from limited generative capability. Despite leveraging explicit physical priors, they lack the semantic guidance required to synthesize intricate facial geometries from noise, often producing structurally ambiguous outputs that successfully recover the low-frequency information of the image but fail to resolve fine-grained high-frequency features.

Motivated by these observations, we propose **PASDiff**, a training-free Physics-Aware Semantic Diffusion (PASDiff) that reformulates the joint task as a physically and structurally constrained generative process. We design a Multi-Objective Energy-Based Guidance strategy to explicitly steer the diffusion sampling trajectory. Specifically, our core idea is to decompose the complex degradation into distinct physical and structural components, orchestrating a synergy between the pre-trained diffusion and domain-specific priors. On the physical dimension, we incorporate spatially varying exposure constraints and Retinex-based reflectance priors, which regulate the optimization path, ensuring that both the recovered illumination and chromaticity align with natural scene statistics. On the structural dimension, to address the ill-posed nature of blind face restoration, we devise a Style-Agnostic Structural Injection (SASI) mechanism. We

leverage an off-the-shelf restore to provide structural cues, but we identify that these priors carry incorrect illumination and color estimates, which can introduce incorrect global style biases into the generative process. To solve this, we propose a Statistic-Aligned Guidance Loss. By dynamically aligning the first and second-order statistics (mean and variance) of the guidance signal with the current diffusion state, this mechanism strictly distills high-frequency structural gradients from the prior while statistically filtering out its low-frequency biases in both luminance and chromaticity. This facilitates an effective decoupling of texture recovery from illumination and color enhancement, enabling our model to synthesize plausible details that are both structurally faithful and visually harmonious. Through our design, as shown in Fig. 2, restoration results exhibit faithful identity preservation and natural illumination distribution.

In summary, our main contributions are as follows:

- We propose PASDiff, a training-free framework for joint low-light enhancement and face restoration. By reformulating the task as a dually constrained generation, we harness priors of diffusion without training and paired data.
- We devise a SASI strategy via a Statistic-Aligned Guidance Loss, which elegantly decouples texture recovery from global photometry, enabling the precise distillation of structural semantics from off-the-shelf priors while explicitly filtering out their intrinsic lighting and color biases.
- We construct WildDark-Face, a real-world benchmark comprising 700 facial images with complex compound degradations. Extensive experiments demonstrate that PASDiff surpasses existing solutions, delivering superior perceptual quality, natural illumination, and identity preservation.

2 Related Work

2.1 Joint Restoration of Compound Degradation

Restoring images captured in unconstrained environments is highly challenging. We briefly review methods addressing individual degradations and recent attempts toward joint restoration.

Low-Light Image Enhancement. Early approaches relied on Histogram Equalization or Retinex theory [16] to decompose reflectance and illumination. Deep learning methods [47, 48] integrated these physical models into neural networks, while unsupervised approaches like Zero-DCE [15] eliminated the need for paired data. Recently, generative methods such as LLFlow [46] and GDP [11] have employed normalizing flows and diffusion priors to model complex light distributions. Furthermore, DLFN [56] integrated diffusion models with Laplacian decomposition to balance noise suppression and detail preservation. To improve efficiency, RetinexMamba [1] leveraged State Space Models for long-range dependency modeling, and HVI [50] introduced a hue-value-insensitive color space to reduce distortion. However, these methods operate on global statistics and lack specific awareness of facial semantics, often leading to over-smoothing or color artifacts on human faces.

Blind Face Restoration. Blind Face Restoration (BFR) [30] focuses on recovering facial details from degraded inputs. While early works used geometric priors [3], GAN-based methods like GFPGAN [43] and CodeFormer [57] have become mainstream by leveraging pre-trained StyleGAN latent spaces or VQ-codebooks. More recently, diffusion-based models such as DiffBIR [35] have established strong baselines by decoupling degradation removal and generation. To accelerate inference, OSDFace [41] utilized a vector-quantized dictionary to achieve one-step restoration. Nevertheless, these models assume "normal" lighting; severe noise and uneven illumination in low-light conditions disrupt their feature extraction, causing structural deformation.

Restoration of Coupled Degradations. Addressing multiple degradations simultaneously is critical but difficult. Simple cascaded strategies often lead to error accumulation. Pioneering this direction, LEDNet [58] proposed a specific encoder-decoder architecture with a re-weighting mechanism to jointly handle low-light and blur. Unified networks like AirNet [26] and PromptIR [38] further attempted to handle various degradations within a single framework. Most recently, InstructIR [6] introduced instruction tuning to guide the model in removing specific degradation types. Despite these efforts, methods like DarkIR [12] and FDN [40], which utilized spatial-frequency attention or Fourier decoupling, usually operate discriminatively. They struggle to hallucinate high-frequency details needed for face restoration and often rely on specific blur assumptions, lacking the generalization ability of generative priors.

2.2 Diffusion Models for Inverse Problems

Leveraging pre-trained diffusion priors for inverse problems has gained significant traction in low-level vision tasks. One type of work, like DPS [5], DDRM [21], and MCS [32], approximated the posterior distribution using measurement errors or singular value decomposition, while DDNM [45] introduced null-space decomposition for linear consistency. For unknown degradations, BlindDPS [4] jointly modeled the image and degradation operator. FreeDoM [53] employed multiple energy functions to guide sampling without retraining, DDPG [14] proposed an iteratively preconditioned strategy to robustly solve non-linear inverse problems, and SSDiff [31] constructed pseudo-labels to assist the guidance. Recent advancements like DAPS [54] and FlowDPS [23] further improved sampling stability and generative quality.

In the realm of illumination and visibility enhancement, methods adapt these priors using unpaired data [24] or by learning specific degradation representations [42]. Notably, Reti-Diff [17] integrated Retinex theory with latent diffusion to explicitly tackle illumination degradation. Crucially, however, these guidance strategies primarily focus on handling single degradation types independently. They lack the mechanism to seamlessly incorporate the composite physical constraints required for handling complex coupled degradations (e.g., simultaneous low-light enhancement and identity-preserving face restoration). In this work, we bridge this gap by designing a multi-objective energy function to guide the diffusion trajectory toward a high-fidelity, well-lit facial manifold.

3 Methodology

3.1 Preliminaries

Denoising Diffusion Probabilistic Models (DDPM). DDPM [19] defines a forward diffusion process that gradually adds Gaussian noise to a data sample $x_0 \sim q(x_0)$ over T steps. The forward transition $q(x_t|x_{t-1})$ is parameterized as a Markov chain:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t\mathbf{I}), \tag{1}$$

where β_t is a pre-defined variance schedule. Let $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$. A remarkable property of this process is that we can sample x_t at any arbitrary timestep t directly from x_0:

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, \quad \epsilon \sim \mathcal{N}(0, \mathbf{I}). \tag{2}$$

The reverse process learns to denoise x_t to recover x_0. It is modeled by a neural network $\epsilon_\theta(x_t, t)$ which predicts the noise component added in the forward process. The reverse transition $p_\theta(x_{t-1}|x_t)$ is defined as $p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$, where the mean $\mu_\theta(x_t, t)$ is derived as:

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t)\right). \tag{3}$$

In practice, we can directly estimate x_0 from the predicted noise $\epsilon_\theta(x_t, t)$ via:

$$\hat{x}_0 = \frac{x_t - \sqrt{1-\bar{\alpha}t}\epsilon\theta(x_t, t)}{\sqrt{\bar{\alpha}_t}}. \tag{4}$$

Classifier Guidance. To introduce semantic control into the generation process, Dhariwal et al. [8] proposed classifier guidance. Instead of training a conditional diffusion model from scratch, this approach modifies the sampling trajectory of a pre-trained unconditional model using gradients from an auxiliary classifier $c_\phi(y|x_t)$. The perturbed reverse transition probability can be approximated as a Gaussian distribution with a shifted mean:

$$p_{\theta,\phi}(x_{t-1}|x_t, y) \approx \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t) + \Sigma_\theta(x_t, t)\nabla_{x_t} \log c_\phi(y|x_t), \Sigma_\theta(x_t, t)), \tag{5}$$

where s is the guidance scale. The gradient term $g = \nabla_{x_t} \log c_\phi(y|x_t)$ acts as a guidance signal that biases the sampling distribution toward the target semantics defined by y, effectively steering the generative process without altering the pre-trained weights.

3.2 Physics-Aware Semantic Guidance Framework

As analyzed in Sec. 1, existing solutions for complex degraded low-light face images face a fundamental conflict between *structural fidelity* and *photometric*

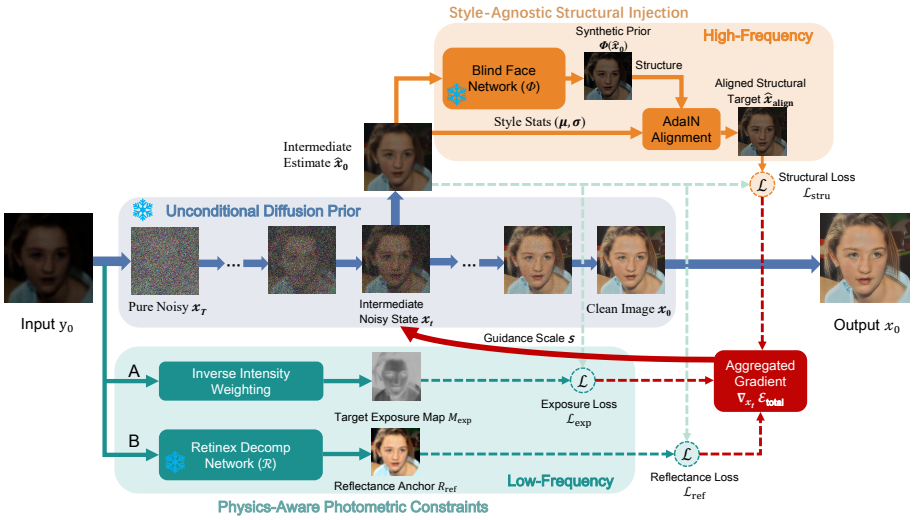

Fig. 3: Overall framework of the proposed training-free PASDiff. Our method reformulates joint restoration via dual-dimensional diffusion guidance. Physically, Retinex-based photometric constraints steer the sampling trajectory to recover natural illumination and chromaticity. Structurally, our Style-Agnostic Structural Injection (SASI) and Statistic-Aligned Guidance Loss extract high-frequency facial semantics from priors while explicitly filtering out their intrinsic lighting biases. This synergy seamlessly harmonizes texture realism with physical reliability.

restoration. To address this conflict, we propose PASDiff, a train-free framework that decouples this task into two orthogonal objectives: *photometric correction* and *structural refinement*, unifying them within the latent trajectory of an unconditional DDPM. As illustrated in Fig. 3 and Algorithm 1, given a degraded input y_0, our goal is to sample x_0 from a posterior distribution $p(x_0 \mid y_0)$. Since the exact posterior is intractable, we approximate it by steering the reverse process of an unconditional DDPM using a composite energy function $\mathcal{E}_{total}$. Specifically, at each timestep t, we estimate the clean image $\hat{x}_0$ from the current noisy state x_t. Then, we apply physical constraints to correct the illumination and color of $\hat{x}_0$, while leveraging a Style-Agnostic Structural Injection (SASI) to extract high-frequency facial semantics without introducing incorrect global style biases. The aggregated gradient $\nabla_{x_t}\mathcal{E}_{total}$ ultimately guides the trajectory toward a high-fidelity manifold. In the following, we elaborate on its two key components: Physics-Aware Photometric Constraints and Style-Agnostic Structural Injection.

Physics-Aware Photometric Constraints. While generative models provide powerful priors, they lack explicit knowledge of the scene's physical lighting conditions. In low-light scenarios, unconstrained generation often leads to inconsistent chromatic shifts, as the model attempts to synthesize chromatic features from noise without a reliable reference. To ground the generative process in phys-

Algorithm 1 Inference via Physics-Aware Semantic Guidance

Require: Pre-trained diffusion model $(\mu_\theta(x_t,t), \Sigma_\theta(x_t,t))$, Restoration Model Φ, Retinex Net $\mathcal{R}$, guidance scale s, gradient steps N.

Input: A low-light face image y_0 with complex degradation.
Output: High-quality normal-light face image $\hat{x}_0$.
$M_{exp} \leftarrow \text{GetExposure}(y_0)$; $R_{ref} \leftarrow \mathcal{R}(y_0)$
$x_T \leftarrow \sqrt{\bar{\alpha}_T} y_0 + \sqrt{1-\bar{\alpha}_T}\epsilon$, $\epsilon \sim \mathcal{N}(0, \mathbf{I})$
for $t = T$ **to** 1 **do**
  $\mu_t, \Sigma_t \leftarrow \mu_\theta(x_t,t), \Sigma_\theta(x_t,t)$
  $\hat{x}_0 \leftarrow \frac{1}{\sqrt{\bar{\alpha}_t}} x_t - \frac{\sqrt{1-\bar{\alpha}_t}}{\sqrt{\bar{\alpha}_t}} \epsilon_\theta(x_t,t)$
  repeat
    $\mathcal{L}_{phy} = \lambda_{exp}\|\text{Mean}_c(\hat{x}_0) - M_{exp}\|^2 + \lambda_{ref}\|\mathcal{R}(\hat{x}_0) - R_{ref}\|^2$; $\mathcal{L}_{stru} = \lambda_{stru}\|\hat{x}_0 - \text{AdaIN}(\Phi(\hat{x}_0), \hat{x}_0)\|^2$
    $g \leftarrow \nabla_{\hat{x}_0}(\mathcal{L}_{phy} + \mathcal{L}_{stru})$
    $x_t \sim \mathcal{N}(\mu_t - s\Sigma_t g, \Sigma_t)$
    $\hat{x}_0 \leftarrow \frac{1}{\sqrt{\bar{\alpha}_t}} x_t - \frac{\sqrt{1-\bar{\alpha}_t}}{\sqrt{\bar{\alpha}_t}} \epsilon_\theta(x_t,t)$
  until $N-1$ times
  $x_{t-1} \sim \mathcal{N}(\mu_t - s\Sigma_t g, \Sigma_t)$
end for
return $\hat{x}_0$

ical reality, we leverage Retinex theory [25] as the theoretical foundation for our dual constraints. Retinex theory decomposes an image I into an illumination component L (governing visibility) and a reflectance component R (governing intrinsic color), modeled as $I = R \circ L$. Based on this physical decoupling, we guide the diffusion trajectory via two complementary terms: one regulating the illumination L to ensure adequate visibility, and the other constraining the reflectance R to promote natural chromaticity.

(a) Spatially-Varying Exposure Guidance. A critical challenge in low-light enhancement is the extreme dynamic range: blindly boosting global brightness often renders dark regions visible at the cost of blowing out originally bright areas (e.g., street lamps or reflections). Uniform exposure adjustments fail to balance these conflicting demands. To address this, we formulate a target exposure map M_{exp} based on an inverse intensity weighting strategy [27]. Specifically, to decouple lightness from chromaticity, we transform the input y_0 into the HSI color space and extract the Intensity component I_{in} (calculated as the pixel-wise average of RGB channels). To construct a spatially adaptive guide, we compute the deviation of local intensity from the global average intensity $\bar{I}_{in}$. The target exposure map M_{exp} (Fig. 4 Left) is then synthesized as:

$$M_{exp} = \alpha + \beta \cdot \text{Norm}(\bar{I}_{in} - I_{in}), \tag{6}$$

where $\text{Norm}(\cdot)$ denotes min-max normalization. Based on statistical observations of low-light distributions, the hyperparameters α (base exposure) and β (adjustment amplitude) are empirically set to 0.46 and 0.25, respectively. This map acts as a pixel-wise attention mechanism: assigning higher target exposures

to underexposed regions while restricting gains in brighter areas. Consequently, the exposure loss is defined as:

$$\mathcal{L}_{exp} = \|\text{Mean}_c(\hat{x}_0) - M_{exp}\|_2^2, \tag{7}$$

where $\text{Mean}_c(\cdot)$ computes the average along the channel dimension. By minimizing $\mathcal{L}_{exp}$, we enforce a physically balanced global illumination distribution on the estimated $\hat{x}_0$.

(b) Retinex-based Reflectance Prior. Recovering precise chromaticity from extreme darkness is an intrinsically ill-posed problem. The lack of sufficient photon counts inevitably leads to color undersaturation and unpredictable chromatic shifts. To constrain the solution space within a plausible color manifold, we exploit the illumination-invariant property of reflectance. According to the Retinex theory, the reflectance component represents the intrinsic chromatic properties of objects, and this chromatic consistency provides robust cues even under extremely low-light conditions. Therefore, we utilize the reflectance map (Fig. 4 Left) extracted from the input y_0 as a robust "chromatic anchor" to constrain the solution space of the restoration process. We employ a pre-trained Retinex decomposition network [13], denoted as $\mathcal{R}(\cdot)$, to extract the reference reflectance $R_{ref} = \mathcal{R}(y_0)$. We premise that a valid high-quality restoration $\hat{x}_0$ should maintain chromatic consistency with the intrinsic reflectance of the input. Thus, we formulate the reflectance loss $\mathcal{L}_{ref}$ as:

$$\mathcal{L}_{ref} = \|\mathcal{R}(\hat{x}_0) - R_{ref}\|_2^2. \tag{8}$$

By minimizing this term, we effectively restrict the generative diversity toward a visually plausible color manifold. This ensures that the diffusion model [29] focuses its capacity on synthesizing high-frequency textures while aligning closely with the scene's available chromatic cues. The total physical guidance energy $\mathcal{L}_{phy}$ is then formulated as a weighted sum:

$$\mathcal{L}_{phy} = \lambda_{exp}\mathcal{L}_{exp} + \lambda_{ref}\mathcal{L}_{ref}. \tag{9}$$

Style-Agnostic Structural Injection. While the aforementioned constraints ensure photometric plausibility, they operate primarily in the low-frequency domain. Recovering intricate high-frequency facial details (e.g., pores and eyelashes) from severe degradations remains a formidable challenge that physics alone cannot solve. To bridge this gap, we introduce a semantic prior from a potent blind face restoration network [41], denoted as the external restoration prior Φ. However, integrating this prior into our framework introduces a critical domain conflict: predictions from off-the-shelf models ($\hat{x}_{prior} = \Phi(\hat{x}_0)$) are typically biased toward canonical laboratory lighting and synthetic color distributions. Directly minimizing $\|\hat{x}_0 - \hat{x}_{prior}\|_2^2$ would force the generative process to mimic these synthetic styles, which contradicts our physically grounded constraints and degrades visual naturalness. To resolve this dilemma, we propose a Style-Agnostic Structural Injection strategy. Our core insight is that the *identity* and *structure* of a face reside in high-frequency spatial variations, while the

Table 1: Face recognition accuracy comparisons.

Method	Accuracy ↑
L-Diff [20] → TSD-SR [10]	60.06%
L-Diff [20] → DiffBIR [35]	63.96%
TSD-SR [10] → L-Diff [20]	60.71%
DiffBIR [35] → L-Diff [20]	49.68%
DarkIR [12]	58.44%
LIEDNet [37]	57.14%
LEDNet [58]	57.14%
VQCNIR [59]	49.68%
URWKV [49]	51.95%
FDN [40]	61.04%
PASDiff (Ours)	**71.43%**

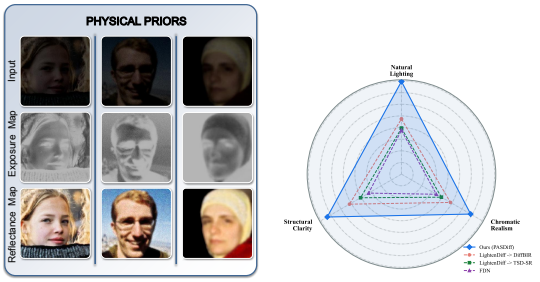

Fig. 4: Left: Visualization of physical priors (Exposure and Reflectance maps). **Right:** Subjective preferences from the user study.

photometric *style* is dominated by low-frequency global statistics. Therefore, our goal is to extract the structural gradients strictly from Φ while statistically stripping away its photometric biases. We achieve this by aligning the feature statistics (mean μ and standard deviation σ) of the prior prediction with the current intermediate state $\hat{x}_0$ via Adaptive Instance Normalization (AdaIN). The aligned structural target $\hat{x}_{align}$ is formulated as:

$$\hat{x}_{align} = \sigma(\hat{x}_0) \left(\frac{\Phi(\hat{x}_0) - \mu(\Phi(\hat{x}_0))}{\sigma(\Phi(\hat{x}_0)) + \epsilon} \right) + \mu(\hat{x}_0), \tag{10}$$

where ϵ is a small constant for numerical stability. This transformation effectively normalizes the prior's output to a zero-mean, unit-variance space, stripping away its original illumination and color biases, and then re-projects it onto the intensity distribution of $\hat{x}_0$. Consequently, $\hat{x}_{align}$ inherits the high-fidelity structural details from Φ but strictly adheres to the illumination and color atmosphere of $\hat{x}_0$ (which is governed by our physical constraints). Finally, we define the structural guidance loss $\mathcal{L}_{stru}$ as the distance between the current estimate and this aligned target:

$$\mathcal{L}_{stru} = \|\hat{x}_0 - \hat{x}_{align}\|_2^2. \tag{11}$$

By combining this with the physical gradients, the total guidance gradient g_{total} is derived by aggregating the weighted physical and structural objectives:

$$g_{total} = \nabla_{\hat{x}_0}(\mathcal{L}_{phy} + \mathcal{L}_{stru}) = \nabla_{\hat{x}_0}(\lambda_{exp}\mathcal{L}_{exp} + \lambda_{ref}\mathcal{L}_{ref} + \lambda_{stru}\mathcal{L}_{stru}), \tag{12}$$

where, to handle the varying magnitudes of the loss terms, the balancing weights are empirically set to $\lambda_{exp} = 1200$, $\lambda_{ref} = 0.03$, and $\lambda_{stru} = 10000$. This aggregated gradient steers the diffusion process toward a manifold that is both photometrically natural and structurally faithful.

Table 2: Quantitative comparisons with existing methods on the synthetic FFHQ dataset and the real-world WildDark-Face benchmark. 'C' and 'J' denote Cascaded and Joint general restoration approaches, respectively. The best and second-best results are highlighted in red and blue, respectively.

Method	Type	FFHQ					WildDark-Face			
		PSNR↑	LPIPS↓	DISTS↓	Deg.↓	LMD↓	MUSIQ↑	MANIQA↑	HyperIQA↑	FID↓
LightenDiffusion [20] → TSD-SR [10]	C	17.85	0.4126	0.2745	9.0737	2.6448	42.5937	0.2644	0.5817	191.74
LightenDiffusion [20] → DiffBIR [35]	C	18.20	0.3063	0.2284	7.2990	2.1906	49.4678	0.3302	0.6335	121.47
TSD-SR [10] → LightenDiffusion [20]	C	17.52	0.3850	0.2931	9.0773	3.0751	24.3632	0.1855	0.2680	157.44
DiffBIR [35] → LightenDiffusion [20]	C	16.62	0.4591	0.3024	8.9765	2.8251	49.1503	0.2879	0.5748	177.81
DarkIR [12]	J	15.74	0.4487	0.3090	9.1225	2.6302	18.5099	0.1665	0.2131	186.82
LIEDNet [37]	J	15.50	0.4397	0.3084	9.1932	2.6580	18.6445	0.1645	0.2116	183.96
LEDNet [58]	J	16.49	0.4135	0.2956	8.9023	2.8506	21.1880	0.1830	0.2277	173.70
VQCNIR [59]	J	14.35	0.4618	0.3164	9.3756	2.7004	18.8857	0.1870	0.2264	177.35
URWKV [49]	J	13.59	0.4747	0.3321	9.5086	2.7045	18.5134	0.1671	0.2052	176.95
FDN [40]	J	18.56	0.4428	0.3031	9.4073	2.3465	16.2443	0.1451	0.2036	161.70
PASDiff (Ours)	J	22.59	0.2559	0.2274	6.8031	1.6213	53.1825	0.3658	0.7057	127.04

4 Experiments

4.1 Experimental Settings

Implementation Details. Following previous methods [5, 23, 45], our method is built upon a pre-trained unconditional diffusion model (256×256 resolution) trained on ImageNet [8]. To formulate the multi-objective guidance, we adopt a pre-trained Retinex decomposition network [13] to provide photometric constraints and leverage an off-the-shelf blind face restoration network [41] to extract structural semantics. For inference, the standard diffusion process in our method consists of $T = 10$ timesteps to execute noise injection and attribute guidance. All experiments are conducted on a single NVIDIA RTX 3090 GPU.

Datasets and Metrics. For synthetic test datasets, we utilize 1,000 high-quality images from FFHQ, uniformly resized to 256×256. We then apply a physically grounded two-stage degradation pipeline to approximate real-world low-light conditions. Following the protocol in [33, 34], the first stage simulates general face degradation: the ground truth y is sequentially degraded by Gaussian blur ($\sigma \in [0.1, 5]$), downsampling ($r \in [1, 4]$), additive white Gaussian noise ($\delta \in [0, 15]$), and JPEG compression ($q \in [60, 100]$). The second stage simulates low-light physics in the linear domain: we apply an exposure adjustment factor $\alpha = 0.25$ and a hue-preserving Gamma correction sampled from $\gamma \in [1.7, 1.9]$. Formally, the overall degradation process is formulated as:

$$x = \text{Gamma}_{\gamma}\Big(\alpha \cdot \text{JPEG}_{q}\big((y \circledast k_{\sigma}) \downarrow_{r} + \mathbf{n}_{\delta}\big)\Big), \tag{13}$$

where x denotes the synthesized low-light facial image, α is the exposure adjustment factor, k_{σ} is the Gaussian blur kernel, $\downarrow_{r}$ is the downsampling operation, and $\mathbf{n}_{\delta}$ represents the additive white Gaussian noise.

To further assess real-world scenes, we construct WildDark-Face, a real-world benchmark comprising 700 in-the-wild low-light facial images. Specifically, we derive this dataset from the widely recognized DarkFace benchmark [52] by cropping individual faces using the provided bounding box annotations. To ensure

Fig. 5: Visual comparisons on the synthetic FFHQ dataset.

evaluation reliability, we filter out instances with extremely low spatial resolutions and meticulously curate a representative subset. These real-world captures encapsulate a wide spectrum of authentic, unconstrained degradations, encompassing extreme photon starvation, severe sensor noise, unpredictable motion blur, and diverse head poses. For quantitative evaluation on synthetic data, we employ PSNR and LPIPS [55] for signal and perceptual fidelity, and DISTS [9] for texture similarity. For real-world assessment, we utilize no-reference metrics including MUSIQ [22], MANIQA [51], HyperIQA [39], and FID [18]. Furthermore, we report Deg. (degree of identity preservation) [7] and LMD (Landmark Distance) [44] to explicitly measure semantic and geometric identity alignment. More details about the WildDark-Face dataset are provided in the Supplements.

4.2 Comparisons with Existing Methods

Since existing methods typically treat low-light enhancement and face restoration in isolation, there is currently no end-to-end solution specifically tailored for this joint task. To provide an evaluation, we compare PASDiff against two categories. First, we examine Cascaded Approaches by combining low-light enhancers with blind face restorers. Specifically, we employ LightenDiffusion [20] as the enhancer, and select TSD-SR [10] and DiffBIR [35] as restorers, evaluating both Enhancement → Restoration and Restoration → Enhancement cascades. Second, we compare against Joint General Restoration Approaches designed for low-light enhancement and restoration in generic scenes, including DarkIR [12], LIEDNet [37], LEDNet [58], VQCNIR [59], URWKV [49], and FDN [40].

Quantitative Evaluation. Table 2 summarizes quantitative results on both synthetic and real test sets. While the cascaded combination of LightenDiffusion [20] and DiffBIR [35] achieves competitive perceptual scores, it suffers from inherent errors accumulation typical of multi-stage pipelines. Similarly, although generic restoration models such as FDN achieve decent PSNR scores, they fail

Fig. 6: Visual comparisons on the real-world WildDark-Face dataset.

to capture high-frequency facial semantics, resulting in poor performance on identity-aware metrics (Deg. and LMD). In contrast, PASDiff achieves the best performance across almost all evaluation metrics.

Furthermore, we conduct a face recognition accuracy test using the pre-trained InsightFace (`buffalo_l`) model [7]. Specifically, we extract normalized facial embeddings from both restored and ground-truth faces to compute cosine similarity, defining a successful identity match with a threshold of 0.42. As reported in Table 1, our method achieves over 8% higher face recognition accuracy than the second-best method.

Qualitative Evaluation. Visual comparisons on the synthetic FFHQ and real-world WildDark-Face datasets are presented in Fig. 5 and Fig. 6, respectively. As observed, cascaded methods exhibit specific limitations depending on their execution order. In the L-Diff [20]→DiffBIR [35] sequence, the restoration model tends to hallucinate unnatural textures based on noise amplified by the preceding enhancer, resulting in blotchy color artifacts and structural fractures on faces. Conversely, the reverse order (DiffBIR [35]→L-Diff [20]) often suffers from grid-like artifacts and severe detail degradation. Meanwhile, representative joint general methods (e.g., LEDNet, FDN, and DarkIR) struggle with such complex composite degradations; they fail to effectively lift the illumination and frequently produce excessively blurry and under-enhanced results. In contrast, PASDiff generates photometrically natural and identity-consistent facial features. More visual results of other baseline methods are provided in the Supplements.

User Study. To capture human visual preference, we conducted a subjective user study using 10 synthetic and 10 real-world images. Evaluators assessed our method against the top three baselines by ranking the results across three dimensions: natural lighting, chromatic realism, and structural clarity. As illustrated in Fig. 4 (Right), PASDiff consistently receives the highest preference rankings, validating its superiority in generating visually pleasing results.

Table 3: Ablation on physical constraints.

Method	PSNR↑	LPIPS↓	LMD↓
w/o $\mathcal{L}_{exp}$	21.54	**0.2463**	1.6295
w/o $\mathcal{L}_{ref}$	11.15	0.6680	13.3718
Ours	**22.59**	0.2559	**1.6213**

LQ w/o $\mathcal{L}_{exp}$ w/o $\mathcal{L}_{ref}$ **Ours**

Fig. 7: Visual ablation on physical constraints.

Table 4: Ablation on structural injection.

Method	PSNR↑	LPIPS↓	LMD↓
w/o $\mathcal{L}_{stru}$	**23.57**	0.4501	2.1251
Ours	22.59	**0.2559**	**1.6213**

LQ w/o $\mathcal{L}_{stru}$ **Ours**

Fig. 8: Visual ablation on structural injection.

4.3 Ablation Study

To further validate the effectiveness of each component, we report ablation results on our synthetic dataset concerning the physical constraints, structural injection, and style-agnostic design.

Impact of Physical Constraints. The physical constraints, comprising the exposure loss $\mathcal{L}_{exp}$ and the reflectance loss $\mathcal{L}_{ref}$, form the foundation of our restoration. As shown in Table 3 and Fig. 7, when removing $\mathcal{L}_{exp}$, although the basic content remains visible due to the reflectance constraint, the restored images appear overall darker, failing to meet the visibility standards of normal-light scenes. Conversely, removing $\mathcal{L}_{ref}$ deprives the model of a reliable chromatic anchor. The results suffer from a profound loss of chromatic information, failing to recover their intrinsic colors. This confirms that both terms are essential for establishing a physically plausible foundation.

Impact of Structural Injection. Removing the structural guidance loss $\mathcal{L}_{stru}$ forces the model to rely solely on physical constraints. As illustrated in Table 4 and Fig. 8, the resulting images suffer from residual blur, losing critical high-frequency identity characteristics. Although this low-frequency-biased optimization trivially yields a higher PSNR, it severely degrades perceptual fidelity (LPIPS) and structural alignment (LMD), which demonstrates that physics-based guidance alone is insufficient for realistic blind face restoration.

Effectiveness of Style-Agnostic Design. We evaluate the necessity of our SASI by replacing it with an MSE loss between estimated images and the prior's output. Results in Table 5 and Fig. 9 reveal a critical gradient conflict: the naive approach forces diffusion models to mimic the prior's synthetic lighting style, directly contradicting our physically grounded exposure and reflectance constraints. Consequently, this mismatch leads to distinct color shifts and discordant global tones. In contrast, our AdaIN-based SASI successfully distills

Table 5: Ablation on style-agnostic design.

Method	PSNR↑	LPIPS↓	LMD↓
MSE	21.86	0.2571	1.6518
Ours	**22.59**	**0.2559**	**1.6213**

Fig. 9: Visual ablation on style-agnostic design.

structural semantics while statistically filtering out mismatched styles, effectively harmonizing high-fidelity details with the correct physical atmosphere.

5 Limitations and Future Work

Despite achieving SOTA quality, PASDiff presents two main limitations. First, relying on iterative diffusion sampling incurs a slower inference speed compared to end-to-end feed-forward networks. Second, recovering precise chromaticity from extremely low-light inputs remains intrinsically ill-posed. Although our physical constraints significantly shift the overall color distribution toward high-quality references, the model can still exhibit a certain degree of under-saturation in near-pitch-black regions where the original color information is irreversibly destroyed. In future work, we intend to integrate advanced diffusion acceleration techniques to streamline inference and explore explicit generative color priors to further close the chromaticity gap in extreme real degradations.

6 Conclusion

We propose PASDiff, a training-free framework tailored for joint low-light enhancement and blind face restoration. To achieve physically plausible illumination and color distributions, we introduce photometric constraints leveraging inverse intensity weighting and Retinex theory. To faithfully recover high-frequency facial details, we devise a SASI strategy, which distills structural semantics from an off-the-shelf facial prior while explicitly filtering out its intrinsic photometric biases. By seamlessly harmonizing physical and semantic constraints, PASDiff effectively resolves the error accumulation and structural infidelity issues prevalent in existing cascaded and joint paradigms. Extensive evaluations on both synthetic and real-world benchmarks demonstrate the superiority of our approach in yielding naturally illuminated, chromatically plausible, and identity-consistent.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant Nos. U24A20330 and 62361166670.

References

1. Bai, J., Yin, Y., He, Q., Li, Y., Zhang, X.: RetinexMamba: Retinex-based Mamba for low-light image enhancement. In: ICONIP. pp. 427–442. Springer (2024) 2, 4
2. Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., Zhang, Y.: Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In: ICCV. pp. 12504–12513 (2023) 2
3. Chen, Y., Tai, Y., Liu, X., Shen, C., Yang, J.: FSRNet: End-to-end learning face super-resolution with facial priors. In: CVPR. pp. 2492–2501 (2018) 5
4. Chung, H., Kim, J., Kim, S., Ye, J.C.: Parallel diffusion models of operator and image for blind inverse problems. In: CVPR. pp. 6059–6069 (2023) 5
5. Chung, H., Kim, J., Mccann, M.T., Klasky, M.L., Ye, J.C.: Diffusion posterior sampling for general noisy inverse problems. In: ICLR (2023) 5, 11
6. Conde, M.V., Geigle, G., Timofte, R.: InstructIR: High-quality image restoration following human instructions. In: ECCV. pp. 1–21. Springer (2024) 5
7. Deng, J., Guo, J., Xue, N., Zafeiriou, S.: Arcface: Additive angular margin loss for deep face recognition. In: CVPR. pp. 4690–4699 (2019) 12, 13
8. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. NeurIPS **34**, 8780–8794 (2021) 6, 11
9. Ding, K., Ma, K., Wang, S., Simoncelli, E.P.: Image quality assessment: Unifying structure and texture similarity. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(5), 2567–2581 (2020) 12
10. Dong, L., Fan, Q., Guo, Y., Wang, Z., Zhang, Q., Chen, J., Luo, Y., Zou, C.: TSD-SR: One-step diffusion with target score distillation for real-world image super-resolution. In: CVPR. pp. 23174–23184 (2025) 10, 11, 12
11. Fei, B., Lyu, Z., Pan, L., Zhang, J., Yang, W., Luo, T., Zhang, B., Dai, B.: Generative diffusion prior for unified image restoration and enhancement. In: CVPR. pp. 9935–9946 (2023) 4
12. Feijoo, D., Benito, J.C., Garcia, A., Conde, M.V.: DarkIR: Robust low-light image restoration. In: CVPR. pp. 10879–10889 (2025) 2, 5, 10, 11, 12
13. Fu, Z., Yang, Y., Tu, X., Huang, Y., Ding, X., Ma, K.K.: Learning a simple low-light image enhancer from paired low-light instances. In: CVPR. pp. 22252–22261 (2023) 9, 11
14. Garber, T., Tirer, T.: Image restoration by denoising diffusion models with iteratively preconditioned guidance. In: CVPR. pp. 25245–25254 (2024) 5
15. Guo, C., Li, C., Guo, J., Loy, C.C., Hou, J., Kwong, S., Cong, R.: Zero-reference deep curve estimation for low-light image enhancement. In: CVPR. pp. 1780–1789 (2020) 4
16. Guo, X., Li, Y., Ling, H.: LIME: Low-light image enhancement via illumination map estimation. IEEE Transactions on Image Processing **26**(2), 982–993 (2016) 4
17. He, C., Fang, C., Zhang, Y., Li, K., Tang, L., You, C., Xiao, F., Guo, Z., Li, X.: Reti-Diff: Illumination degradation image restoration with Retinex-based Latent Diffusion Model. In: ICLR (2025) 5
18. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. In: NeurIPS. vol. 30 (2017) 12
19. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. NeurIPS **33**, 6840–6851 (2020) 6
20. Jiang, H., Luo, A., Liu, X., Han, S., Liu, S.: Lightendiffusion: Unsupervised low-light image enhancement with latent-retinex diffusion models. In: ECCV. pp. 161–179. Springer (2024) 2, 10, 11, 12, 13

21. Kawar, B., Elad, M., Ermon, S., Song, J.: Denoising diffusion restoration models. NeurIPS **35**, 23593–23606 (2022) 5
22. Ke, J., Wang, Q., Wang, Y., Milanfar, P., Yang, F.: Musiq: Multi-scale image quality transformer. In: ICCV. pp. 5148–5157 (2021) 12
23. Kim, J., Kim, B.S., Ye, J.C.: FlowDPS: Flow-driven posterior sampling for inverse problems. arXiv preprint arXiv:2503.08136 (2025) 5, 11
24. Lan, Y., Cui, Z., Liu, C., Peng, J., Wang, N., Luo, X., Liu, D.: Exploiting diffusion prior for real-world image dehazing with unpaired training. In: AAAI. vol. 39, pp. 4455–4463 (2025) 5
25. Land, E.H.: The retinex theory of color vision. Scientific american **237**(6), 108–129 (1977) 8
26. Li, B., Liu, X., Hu, P., Wu, Z., Lv, J., Peng, X.: All-in-one image restoration for unknown corruption. In: CVPR. pp. 17452–17462 (2022) 5
27. Li, C., Guo, C., Feng, R., Zhou, S., Loy, C.C.: Cudi: Curve distillation for efficient and controllable exposure adjustment. arXiv preprint arXiv:2207.14273 (2022) 8
28. Li, W., Guo, H., Liu, X., Liang, K., Hu, J., Ma, Z., Guo, J.: Efficient face super-resolution via wavelet-based feature enhancement network. In: ACM MM. pp. 4515–4523 (2024) 2
29. Li, W., Shi, J., Han, J., Guo, H., Ma, Z.: Seeing through the rain: Resolving high-frequency conflicts in deraining and super-resolution via diffusion guidance. In: AAAI (2026) 9
30. Li, W., Wang, M., Zhang, K., Li, J., Li, X., Zhang, Y., Gao, G., Ma, Z.: Survey on deep face restoration: From non-blind to blind and beyond. ACM Computing Surveys (2025) 5
31. Li, W., Wang, X., Guo, H., Gao, G., Ma, Z.: Self-supervised selective-guided diffusion model for old-photo face restoration. In: NeurIPS (2025) 5
32. Li, W., Zhang, Y., Gao, G., Guo, H., Ma, Z.: Measurement-constrained sampling for text-prompted blind face restoration (2025) 5
33. Li, X., Chen, C., Zhou, S., Lin, X., Zuo, W., Zhang, L.: Blind face restoration via deep multi-scale component dictionaries. In: ECCV. pp. 399–415. Springer (2020) 11
34. Li, X., Liu, M., Ye, Y., Zuo, W., Lin, L., Yang, R.: Learning warped guidance for blind face restoration. In: ECCV. pp. 272–289 (2018) 11
35. Lin, X., He, J., Chen, Z., Lyu, Z., Dai, B., Yu, F., Qiao, Y., Ouyang, W., Dong, C.: DiffBIR: Toward blind image restoration with generative diffusion prior. In: ECCV. pp. 430–448. Springer (2024) 2, 3, 5, 10, 11, 12, 13
36. Lin, Y., Ye, T., Chen, S., Fu, Z., Wang, Y., Chai, W., Xing, Z., Li, W., Zhu, L., Ding, X.: Aglldiff: Guiding diffusion models towards unsupervised training-free real-world low-light image enhancement. In: AAAI. vol. 39, pp. 5307–5315 (2025) 3
37. Liu, M., Cui, Y., Ren, W., Zhou, J., Knoll, A.C.: Liednet: A lightweight network for low-light enhancement and deblurring. IEEE Transactions on Circuits and Systems for Video Technology **35**(7), 6602–6615 (2025) 2, 10, 11, 12
38. Potlapalli, V., Zamir, S.W., Khan, S., Khan, F.S.: PromptIR: Prompting for all-in-one blind image restoration. In: NeurIPS. vol. 36, pp. 24706–24746 (2023) 5
39. Su, S., Yan, Q., Zhu, Y., Zhang, C., Ge, X., Sun, J., Zhang, Y.: Blindly assess image quality in the wild guided by a self-adaptive hyper network. In: CVPR. pp. 3660–3669 (2020) 12
40. Tu, L., Wu, J., Wang, C., Meng, D., Jin, Z.: Fourier-based decoupling network for joint low-light image enhancement and deblurring. IEEE Transactions on Image Processing **34**, 5184–5199 (2025) 2, 5, 10, 11, 12

41. Wang, J., Gong, J., Zhang, L., Chen, Z., Liu, X., Gu, H., Liu, Y., Zhang, Y., Yang, X.: Osdface: One-step diffusion model for face restoration. In: CVPR. pp. 12626–12636 (2025) 2, 5, 9, 11
42. Wang, T., Zhang, K., Zhang, Y., Luo, W., Stenger, B., Lu, T., Kim, T.K., Liu, W.: LLDiffusion: Learning degradation representations in diffusion models for low-light image enhancement. Pattern Recognition **166**, 111628 (2025) 5
43. Wang, X., Li, Y., Zhang, H., Shan, Y.: Towards real-world blind face restoration with generative facial prior. In: CVPR. pp. 9168–9178 (2021) 2, 5
44. Wang, X., Bo, L., Fuxin, L.: Adaptive wing loss for robust face alignment via heatmap regression. In: ICCV. pp. 6971–6981 (2019) 12
45. Wang, Y., Yu, J., Zhang, J.: Zero-Shot image restoration using Denoising Diffusion Null-Space Model. In: ICLR (2023) 5, 11
46. Wang, Y., Wan, R., Yang, W., Li, H., Chau, L.P., Kot, A.: Low-light image enhancement with normalizing flow. In: AAAI. vol. 36, pp. 2604–2612 (2022) 4
47. Wei, C., Wang, W., Yang, W., Liu, J.: Deep Retinex decomposition for low-light enhancement. In: BMVC (2018) 2, 4
48. Wu, W., Weng, J., Zhang, P., Wang, X., Yang, W., Jiang, J.: URetinex-Net: Retinex-based deep unfolding network for low-light image enhancement. In: CVPR. pp. 5901–5910 (2022) 4
49. Xu, R., Niu, Y., Li, Y., Xu, H., Liu, W., Chen, Y.: Urwkv: Unified rwkv model with multi-state perspective for low-light image restoration. In: CVPR. pp. 21267–21276 (2025) 2, 10, 11, 12
50. Yan, Q., Feng, Y., Zhang, C., Pang, G., Shi, K., Wu, P., Dong, W., Sun, J., Zhang, Y.: HVI: A new color space for low-light image enhancement. In: CVPR. pp. 5678–5687 (2025) 2, 4
51. Yang, S., Wu, T., Shi, S., Lao, S., Gong, Y., Yuan, M., Wang, M., Yang, Y.: Maniqa: Multi-dimension attention network for no-reference image quality assessment. In: CVPR. pp. 1191–1200 (2022) 12
52. Yang, W., Yuan, Y., Ren, W., Liu, J., Scheirer, W.J., Wang, Z., et al.: Advancing image understanding in poor visibility environments: A collective benchmark study. IEEE Transactions on Image Processing **29**, 5737–5752 (2020) 11
53. Yu, J., Wang, Y., Zhao, C., Ghanem, B., Zhang, J.: Freedom: Training-free energy-guided conditional diffusion model. In: ICCV. pp. 23174–23184 (2023) 5
54. Zhang, B., Chu, W., Berner, J., Meng, C., Anandkumar, A., Song, Y.: Improving diffusion inverse problem solving with decoupled noise annealing. In: CVPR. pp. 20895–20905 (2025) 5
55. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: CVPR. pp. 586–595 (2018) 12
56. Zhou, L., Li, W., Li, J., Gao, G., Lin, C.W.: Diffusion-based laplacian frequency-aware network for low-light image enhancement. Pattern Recognition p. 113060 (2026) 4
57. Zhou, S., Chan, K., Li, C., Loy, C.C.: Towards robust blind face restoration with codebook lookup transformer. NeurIPS **35**, 30599–30611 (2022) 5
58. Zhou, S., Li, C., Change Loy, C.: LEDNet: Joint low-light enhancement and deblurring in the dark. In: ECCV. pp. 573–589. Springer (2022) 2, 5, 10, 11, 12
59. Zou, W., Gao, H., Ye, T., Chen, L., Yang, W., Huang, S., Chen, H., Chen, S.: Vqcnir: clearer night image restoration with vector-quantized codebook. In: AAAI. vol. 38, pp. 7873–7881 (2024) 2, 10, 11, 12