Learning to Corrupt for Better Restoration

Joonkyu Park[1], Wooseok Lee[1], Jaeha Kim[1],
Sehoon Kim[3], Bokyeung Lee[3], and Kyoung Mu Lee[1,2]

[1]Dept. of ECE&ASRI, [2]IPAI, Seoul National University, Korea
[3] MX Business, Samsung Co., Ltd.
{jkpark0825, adntjr, hyjkim2, kyoungmu}@snu.ac.kr

Abstract. Recalling that diffusion models map noisy images to clean ones, the choice of noise addition (*i.e.*, corruption) critically influences the denoising process. However, prior diffusion-based image restoration (IR) methods often employ naïve corruption strategies—such as randomly sampling noise, fixing the timestep, or even omitting corruption altogether—which may not accurately reflect the actual degradation of each image. To handle this, we propose Input-Aware Corruption for IR (IAC-IR) framework, which maps each low-quality (LQ) input into an optimal noisy sample that lies on the pretrained diffusion trajectory. Instead of choosing corruption heuristically, we predict the timestep and noise using supervision derived from the properties of the pretrained diffusion model. Specifically, the predicted timestep aligns the corrupted sample with the Gaussian noise corrupted distribution, while the predicted noise preserves the recoverable content of the input. Moreover, we use these input-aware corruption factors to improve conventional score-based distillation. Rather than relying on random corruption, which often produce unreliable target scores and weak gradients, we perform distillation with input-aware corruption, yielding more reliable score estimates and more stable distillation. By modeling input-aware corruption and integrating it into distillation, our method better leverages the pretrained diffusion prior, achieving superior perceptual quality in image restoration.

Keywords: Image Restoration · Diffusion · Distillation

1 Introduction

Image restoration (IR) aims to recover high-quality (HQ) images from their corresponding low-quality (LQ) inputs and plays a crucial role in various computer vision applications, including computational photography [6,31] and perceptual enhancement [12,14,19,25,33,50,56]. To handle this, early approaches [16,27,38,40,58] primarily relied on GAN [7]-based methods, which encourage restored outputs to align with the distribution of real HQ images. More recently, pretrained diffusion models [9,28] have been widely adopted for IR [2,11,20,21,26,36,41,46,54], leveraging their powerful generative priors to progressively reconstruct realistic images through iterative denoising of corrupted samples.

Specifically, leveraging the strong generative prior of pretrained diffusion models [9, 28], previous IR methods commonly adopt them as backbone networks for restoration [4, 20, 21, 45, 55]. Since diffusion models are originally designed to generate realistic images by progressively denoising random Gaussian noise, early approaches condition the model on LQ inputs while still performing a full diffusion process starting from random noise [20, 36, 46, 49, 54]. However, such strategies incur substantial computational cost due to multiple denoising steps and often redundantly regenerate information already present in the LQ input. To improve efficiency, subsequent works [4, 21, 45, 55] adopt one-step diffusion approaches that directly restore images from LQ inputs. However, although diffusion models are inherently trained to recover clean images from Gaussian-corrupted samples, implying that an optimal corruption may exist for each input, prior methods often rely on random noise [42], fixed timesteps [55], or even omit the corruption process [4, 21, 45], thereby failing to fully exploit the pretrained diffusion model's denoising capability.

To address this, we propose a novel Input-Aware Corruption for IR (IAC-IR) with IAC strategy that maps LQ inputs to diffusion-aligned corrupted samples from which a pretrained diffusion model can reliably reconstruct the corresponding HQ images. While a straightforward solution would be to directly supervise the optimal timestep and noise (*i.e.*, the two key factors controlling corruption), no ground-truth timestep or noise exists for each LQ image, making direct supervision infeasible. To overcome this challenge, we instead supervise these variables indirectly by leveraging intrinsic properties of the pretrained diffusion model.

Specifically, as an optimal noisy sample can be viewed as a linear combination of an HQ image and Gaussian noise, where the mixing ratio is determined by the timestep, we search for the optimal timestep that makes the added noise most consistent with the Gaussian distribution assumed during diffusion training; Our predicted timestep is supervised to match this optimal timestep. Moreover, optimal noise should facilitate an easy and reliable denoising process, preserving content while aligning with the diffusion prior. Hence, we supervise the predicted noise by ensuring that, when added to HQ samples, the pretrained diffusion model can denoise it effectively and reconstruct the original content.

In addition, by using the predicted input-aware corruption factors (*i.e.*, timestep and noise), we introduce a novel input-aware distillation strategy [8]. Conventional score-based diffusion distillation methods [43, 52] align the distribution of generated images with that of natural images by matching variational scores; however, these scores are typically computed from randomly corrupted samples, which often yield unreliable targets and unstable gradients [4]. In contrast, our IAC predicts input-aware timestep and noise, enabling us to corrupt samples in a deterministic and aligned manner. The pretrained diffusion model then produces more reliable target scores under this controlled corruption, leading to more stable and effective distillation.

By learning input-aware corruption and incorporating it into distillation, our method achieves highly photorealistic restoration across diverse synthetic [3] and real-world [1, 44, 46, 54, 56] benchmarks. Comprehensive ablation studies further

confirm the robustness and generalization of the predicted timestep and noise. Our contributions can be summarized as follows:

- We propose Input-Aware Corruption for IR (IAC-IR) framework that learns input-specific timestep and noise to align degraded images with the pretrained diffusion trajectory.
- For IAC, we design an indirect supervision scheme for learning optimal corruption by leveraging intrinsic properties of diffusion models.
- We propose an input-aware distillation framework that produces more reliable score supervision and achieves superior restoration performance.

2 Related Works

GAN-based IR. Early image restoration (IR) approaches primarily focused on designing effective network architectures, including deep convolutional layers [12, 14, 19] and attention [34, 37]-based models [57], and were mainly optimized using pixel-wise loss functions. However, such methods often produce overly smooth results that deviate from the real image distribution [14, 40]. To enhance perceptual realism, subsequent works aimed to align the distribution of restored images with that of real HQ images. With the emergence of Generative Adversarial Networks (GANs) [7], a powerful framework for distribution matching, several IR methods adopted GAN-based architectures [17, 27, 39, 40]. These methods typically consist of a generator (*i.e.*, IR network) and a discriminator trained jointly in an adversarial manner: the generator learns to produce realistic outputs that deceive the discriminator, while the discriminator distinguishes restored images from real ones based on their distributions. While GAN-based IR methods significantly improve perceptual quality compared to pixel-wise optimization, they often suffer from training instability, mode collapse, and difficulties in balancing the generator and discriminator. Moreover, with the rise of more powerful generative models (*i.e.*, diffusion models [9, 28]), GANs have gradually become less dominant in recent state-of-the-art image restoration research.

Diffusion-based IR. Leveraging the strong generative capability of diffusion models, which progressively transform simple Gaussian noise into realistic natural images, recent IR methods have adopted pretrained diffusion models as the restoration backbone. Early approaches connect intermediate diffusion steps with LQ inputs through variational modeling [2, 11] or null-space decomposition [41]. Although effective, these methods require solving additional optimization problems at each diffusion step, leading to substantial computational overhead. To improve this, later works [20, 21, 36, 49, 54] fine-tune pretrained diffusion models for image restoration. Rather than updating the full model, which is computationally expensive and may compromise the pretrained generative prior, most methods adopt lightweight adaptation strategies, such as ControlNet [59] or Low-Rank Adaptation (LoRA) [10], updating only a small number of additional layers. Despite these improvements, such approaches still rely on multi-step diffusion processes initialized from randomly sampled noise. This results in high

computational cost and redundant early denoising steps that regenerate information already present in the LQ input.

More recent methods [4, 21, 45, 55] propose single-step diffusion-based IR frameworks that directly feed LQ images into the diffusion model rather than using them as conditions. While these approaches also fine-tune diffusion models to better adapt to LQ inputs, pretrained diffusion models are originally trained on Gaussian-corrupted distributions. Consequently, directly feeding LQ images often leads to a mismatch with the diffusion training distribution, limiting the effective use of the pretrained generative prior. Furthermore, although InvSR [55] predicts noise to corrupt LQ inputs instead of directly feeding them into the diffusion model, it lacks explicit supervision for noise estimation and assumes identical corruption for LQ and HQ images, which may not hold in practice. Moreover, InvSR adopts a fixed timestep, ignoring its critical role in controlling the signal–noise ratio in diffusion. These limitations leave room for more accurate, input-adaptive corruption prediction.

Score-based Distillation. To accelerate diffusion by reducing the number of sampling steps (even a single) step, recent works [24,29] adopt distillation strategies. To this end, these methods train an additional model (*i.e.*, student) to approximate the ordinary differential equation (ODE) sampling trajectory of a pretrained diffusion model (*i.e.*, teacher), effectively compressing the multi-step diffusion process. Later, using the score function predicted by pretrained diffusion models, score-based distillation [43] was first applied in text-to-3D synthesis, where rendered images are optimized to match the text-conditioned image distribution of a diffusion model. This idea was later extended to 2D image synthesis [53] and further improved by integrating GAN-based objectives [51].

In image restoration, recent single-step diffusion methods also adopt score-based distillation to reduce sampling steps by matching diffusion scores [4, 45]. However, this approach relies heavily on the reliability of the teacher's predicted scores. Therefore, when the input distribution severely deviates from the diffusion training distribution, the teacher may produce inaccurate scores, leading to unstable optimization. Although auxiliary target score distillation [4] has been proposed to improve stability, most methods still select noise and timestep heuristically when computing target scores. Since both the noise type and mixing ratio significantly influence score reliability, such heuristic choices often yield suboptimal and unstable distillation signals.

3 Proposed Method

3.1 Preliminary: Diffusion process

Pretrained diffusion models are trained to recover natural images from inputs corrupted by random noise ϵ. The noise is sampled from a Gaussian distribution $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, while the balance between the preserved image signal and the injected noise is controlled by the diffusion timestep, formulated as follows:

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, \tag{1}$$

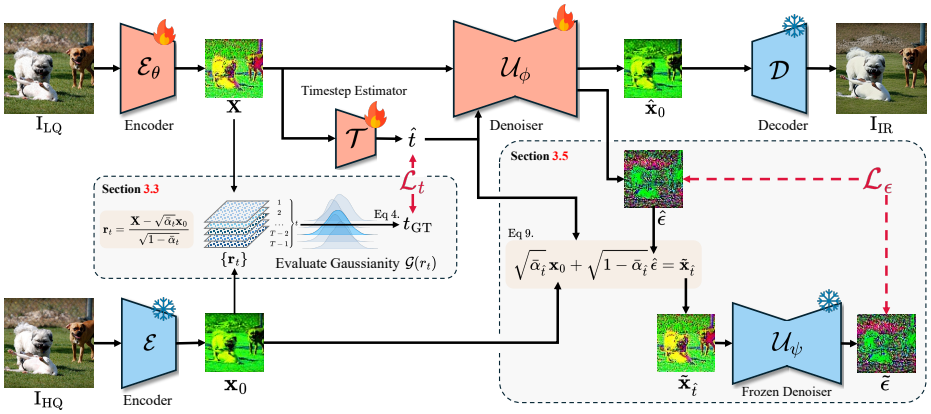


Fig. 1: Overall pipeline of IAC-IR. The LQ image $\mathbf{I}_{\mathrm{LQ}}$ is encoded into a latent modeled as an optimally corrupted sample $\mathbf{X}$. Then, we predict the ground-truth timestep t_{GT} by selecting the one whose residual noise best matches Gaussian statistics. An input-aware timestep $\hat{t}$ is predicted, supervised with t_{GT}. Also, we supervise $\hat{\epsilon}$ by corrupting the clean latent $\mathbf{x}_0$ to obtain $\tilde{\mathbf{x}}_{\hat{t}}$ and enforcing consistency with a frozen denoiser $\mathcal{U}_\psi$.

where $\mathbf{x}_t$ denotes the noisy sample at timestep t, obtained by corrupting the clean image $\mathbf{x}_0$ with random Gaussian noise ϵ. The coefficient $\bar{\alpha} = \{\bar{\alpha}_1, \bar{\alpha}_2, \ldots, \bar{\alpha}_T\}$ represents the cumulative product of the noise schedule, defined as $\bar{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s)$, where β is specified by a predefined variance schedule.

Based on Equation 1, as numerous noisy samples $\mathbf{x}_*$ can be produced under different timesteps t and noise ϵ, the resulting outputs $\hat{\mathbf{x}}_0$ may vary accordingly. Hence, finding the optimal timestep and noise to construct an optimal corrupted sample $\mathbf{x}_*$ is crucial for fully exploiting the power of pretrained diffusion models.

3.2 Goal: Searching Optimal Noisy Input and Timestep

As discussed in Section 3.1, the quality of restored output $\hat{\mathbf{x}}_0$ depends on the specific combination of the timesteps t and the injected noise ϵ. However, in real-world scenarios, the clean image $\mathbf{x}_0$ is unavailable, making it difficult to directly determine the optimal timestep t and noise ϵ. Therefore, our goal is to predict *an optimal noisy input* from the degraded observation I_{LQ}, enabling the diffusion model prior be utilized most effectively.

In the following sections, we first describe how to find the optimal timestep t_{GT} and its prediction $\hat{t}$ by examining whether the candidate injected noise follows a Gaussian distribution. Then, we describe how to optimize our model using the proposed objective function to find the optimal noisy input. This objective includes an indirect supervision scheme based on the core properties of diffusion model. Figure 1 shows the overall pipeline of our proposed IAC-IR framework. All diffusion processes are performed in the latent space using a pretrained VAE [13].

3.3 Finding Optimal Timestep t_{GT}

Our encoder $\mathcal{E}_\theta$ first maps the LQ input $\mathbf{I}_{\mathrm{LQ}}$ to latent $\mathbf{X} \in \mathbb{R}^{h \times w \times c}$, where (h, w) denote the height and width of the latent representation, respectively. From the encoded latent $\mathbf{X}$, which we aim to serve as an optimal noisy input, we follow Equation 1 and model $\mathbf{X}$ as a combination of the predicted high-quality image latent $\hat{\mathbf{x}}_0$ and the predicted Gaussian-like noise $\hat{\epsilon}$. The combination ratio is governed by the predicted timestep $\hat{t}$:

$$\mathbf{X} = \sqrt{\bar{\alpha}_{\hat{t}}}\hat{\mathbf{x}}_0 + \sqrt{1 - \bar{\alpha}_{\hat{t}}}\hat{\epsilon}, \tag{2}$$

where $\mathbf{x}_t$, $\mathbf{x}_0$, t, and ϵ are replaced by the encoded noisy latent $\mathbf{X}$ and their corresponding predicted counterparts $\hat{\mathbf{x}}_0$, $\hat{t}$, and $\hat{\epsilon}$. Here, since the encoded latent $\mathbf{X}$ is not explicitly decomposed into the image latent $\hat{\mathbf{x}}_0$ and noise component $\hat{\epsilon}$, we first estimate the ground-truth timestep to enable this decomposition of the noisy latent into its image and noise constituents.

To this end, we exploit a key property of pretrained diffusion models: they are trained on samples corrupted with Gaussian noise. Based on this, the optimal noise component is expected to follow a Gaussian distribution. Concretely, we construct a set of candidate residuals corresponding to all diffusion timesteps by utilizing the Equation 2. For each timestep t, we compute a residual noise by removing the scaled clean image component $\mathbf{x}_0$ from the latent representation $\mathbf{X}$:

$$\mathbf{r}_t = \frac{\mathbf{X} - \sqrt{\bar{\alpha}_t}\mathbf{x}_0}{\sqrt{1 - \bar{\alpha}_t}}, \forall t \in \{0, \ldots, T-1\}, \tag{3}$$

where $\mathbf{r}_t \in \mathbb{R}^{h \times w \times c}$ denotes the candidate residual noise component assuming timestep t. We then determine the optimal timestep t_{GT} by selecting the residual that most closely follows a standard Gaussian distribution $\mathcal{N}(0, 1)$:

$$t_{\mathrm{GT}} = \arg\min_{t \in \{0, \ldots, T-1\}} \mathcal{G}(\mathbf{r}_t), \tag{4}$$

where $\mathcal{G}(\cdot)$ measures the deviation from a normal distribution. Specifically, we evaluate Gaussianity based on the statistical characteristics of a normal distribution, including mean, variance, skewness, and kurtosis, defined as follows:

$$\mathcal{G}(\mathbf{r}_t) = \|\mu(\mathbf{r}_t)\|_2^2 + \left\|\sigma^2(\mathbf{r}_t) - 1\right\|_2^2 + \|\gamma(\mathbf{r}_t)\|_2^2 + \|\kappa(\mathbf{r}_t) - 3\|_2^2, \tag{5}$$

where $\mu(\mathbf{r}_t)$, $\sigma^2(\mathbf{r}_t)$, $\gamma(\mathbf{r}_t)$, and $\kappa(\mathbf{r}_t)$ correspond to the sample mean, variance, skewness, and kurtosis of $\mathbf{r}_t$, respectively.

3.4 Predicting Timestep $\hat{t}$

Given the encoded latent representation $\mathbf{X}$ from Section 3.3, we predict the timestep $\hat{t}$ using a lightweight timestep estimator module $\mathcal{T}$. Specifically, the timestep estimator $\mathcal{T}$ consists of 12 stacked Vision Transformer (ViT) layers [5]. The input tokens $\in \mathbb{R}^{(1+hw) \times c}$ are formed by concatenating a learnable {class

token$\} \in \mathbb{R}^{1\times c}$ with the $\{$flattened latent representation $\mathbf{X}\} \in \mathbb{R}^{hw\times c}$. The Transformer processes these tokens and outputs an updated class token representation. We then apply a linear classifier to the output class token to produce logits over T candidate timesteps. The predicted timestep $\hat{t}$ is then obtained by selecting the index of the most activated channel as:

$$\hat{t} = \arg \max_{t\in\{0,\dots,T-1\}} \mathcal{T}(\mathbf{X})_t. \tag{6}$$

During training, the predicted timestep $\hat{t}$ is supervised by the ground-truth timestep t_{GT} (defined in Section 3.3) using timestep loss $\mathcal{L}_t$ as:

$$\mathcal{L}_t = \mathrm{CE}\big(\mathcal{T}(\mathbf{X}),\, t_{\mathrm{GT}}\big), \tag{7}$$

where $\mathrm{CE}(\cdot,\cdot)$ denotes the standard cross-entropy loss between the predicted logits and the ground-truth timestep label.

3.5 Predicting Noise ($\hat{\epsilon}$)

Given the noisy latent representation $\mathbf{X}$ and the predicted timestep $\hat{t}$, we feed them into the pretrained UNet denoiser $\mathcal{U}_\phi$, whose parameters are adapted using a LoRA strategy [10]. Since the denoiser $\mathcal{U}_\phi$ predicts the noise component conditioned on the timestep, we decompose the output as:

$$\begin{aligned} \hat{\epsilon} &= \mathcal{U}_\phi\big(\mathbf{X},\, \hat{t}\big), \\ \hat{\mathbf{x}}_0 &= \big(\mathbf{X} - \sqrt{1-\bar{\alpha}_{\hat{t}}}\hat{\epsilon}\big)/\sqrt{\bar{\alpha}_{\hat{t}}}, \end{aligned} \tag{8}$$

where $\hat{\epsilon}$ denotes the predicted noise and $\hat{\mathbf{x}}_0$ denotes the predicted clean image latent. Here, we use a fixed text prompt[1] as input to the UNet. The predicted clean latent $\hat{\mathbf{x}}_0$ can be directly supervised using the corresponding ground-truth latent $\mathbf{x}_0$, whereas supervising the predicted noise $\hat{\epsilon}$ is more challenging. While InvSR [55] implicitly learns noise without explicit supervision, such a strategy may cause the predicted noise to deviate from a Gaussian distribution, which contradicts our assumption that the pretrained diffusion model is trained to denoise Gaussian-like noise. To address this, we explicitly supervise the predicted noise $\hat{\epsilon}$. Specifically, we first corrupt the ground-truth latent $\mathbf{x}_0$ using the predicted timestep $\hat{t}$ and noise $\hat{\epsilon}$ as Equation 1:

$$\tilde{\mathbf{x}}_{\hat{t}} = \sqrt{\bar{\alpha}_{\hat{t}}}\,\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_{\hat{t}}}\,\hat{\epsilon}. \tag{9}$$

We then feed the corrupted latent $\tilde{\mathbf{x}}_{\hat{t}}$ into the *frozen* pretrained UNet denoiser $\mathcal{U}_\psi$ to obtain the reconstructed noise:

$$\tilde{\epsilon} = \mathcal{U}_\psi\big(\tilde{\mathbf{x}}_{\hat{t}}, \hat{t}\big). \tag{10}$$

[1] 'Ultra-high-definition 4K image, sharpness, and lifelike detail. Superior high fidelity, true-to-life colors, immaculate textures, and flawless, authentic, cinematic quality. Crisp, refined, High-end professional photograph.'

Finally, in a self-supervised manner, we enforce consistency between the predicted noise and the denoised output with noise loss $\mathcal{L}_\epsilon$ as:

$$\mathcal{L}_\epsilon = \|\tilde{\epsilon} - \hat{\epsilon}\|_2^2 . \tag{11}$$

This encourages the predicted noise $\hat{\epsilon}$ to be the exact noise component that the pretrained diffusion model can reliably denoise, ensuring alignment with the Gaussian prior and stabilizing timestep prediction.

3.6 Objective Functions

In addition to the loss functions for the predicted timestep $\hat{t}$ and predicted noise $\hat{\epsilon}$, defined in Equations 7 and 11, respectively, we also include supervision for the predicted clean latent $\hat{\mathbf{x}}_0$ following conventional image restoration objectives. Specifically, we first decode $\hat{\mathbf{x}}_0$ into the restored image $\mathbf{I}_{\text{IR}}$ using the frozen pretrained decoder $\mathcal{D}$. Then, we optimize the restored image using a combination of L1 loss, VGG perceptual loss [30], and adversarial loss [7] with HQ image $\mathbf{I}_{\text{HQ}}$.

Furthermore, we incorporate a score-based distillation loss to enhance fine-detail recovery, even under single-step inference. Here, conventional score-based distillation [43, 51, 52] obtains the target score by randomly corrupting inputs with randomly sampled timestep and noise, which often yields suboptimal score estimates. In contrast, our IAC predicts input-adaptive timestep $\hat{t}$ and noise $\hat{\epsilon}$ to perform diffusion-aligned corruption, enabling these factors to be used when computing the scores. Since the corrupted latent sample in our IAC-IR is $\mathbf{X}$, the corresponding score can be represented by $\hat{\mathbf{x}}_0$, as defined in Equation 8. Likewise, the target score can be obtained using the same formulation in the equation, but by replacing the learnable denoiser $\mathcal{U}_\phi$ with the frozen denoiser $\mathcal{U}_\psi$, resulting in $\hat{\mathbf{x}}_0^\psi$. The corresponding score-based distillation loss $\mathcal{L}_{\text{distill}}$ are defined as:

$$\mathcal{L}_{\text{distill}}(\theta) = \mathbb{E}_{\mathbf{X}}\Big[\big\|\hat{\mathbf{x}}_0 - \hat{\mathbf{x}}_0^\psi\big\|_2^2\Big], \tag{12}$$

where distillation loss only updates the encoder $\mathcal{E}_\theta$. Our distillation loss based on the input-aware corruption produces more reliable score supervision, leading to more stable and effective distillation.

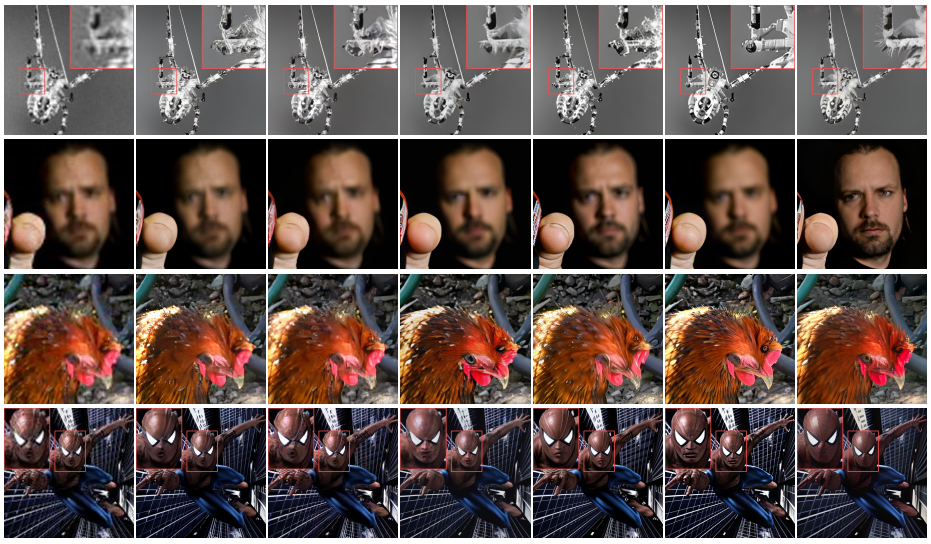
4 Experiments

4.1 Configurations

Datasets. For training our IAC-IR, we use the DF2K [32] and LSDIR [15] datasets. For evaluation, we consider various benchmarks, including synthetically degraded low-quality (LQ) images generated from ImageNet [3]. Specifically, we follow the protocol of InvSR to select and corrupt ImageNet images. In addition, we evaluate on real-world degraded datasets, including DRealSR [44], RealLR200 [46], RealSR [1], RealPhoto [54], and RealSet80 [56]. Among these, RealPhoto and RealSet80 do not provide corresponding ground-truth images,

Table 1: Quantitative comparison on various LQ benchmark datasets. The **best** and **second-best** results are presented in **red** and **blue**, respectively.

Method	Dataset	PSNR↑	SSIM↑	LIQE↑	MUSIQ↑	CLIPIQA↑	MANIQA↑	Dataset	LIQE↑	MUSIQ↑	CLIPIQA↑	MANIQA↑
DiffIR	DRealSR	27.037	0.789	2.334	49.217	0.379	0.301	RealLR200	3.274	61.245	0.521	0.337
SinSR		25.821	0.712	3.080	55.395	0.639	0.388		3.203	61.148	0.660	0.409
OSEDiff		25.852	0.754	3.938	64.683	0.695	0.465		4.055	**69.607**	0.675	0.438
InvSR		23.641	0.679	**4.055**	**65.995**	**0.713**	**0.467**		4.078	68.905	0.683	0.437
HYPIR		24.089	0.695	3.766	61.468	0.627	0.443		**4.146**	69.495	**0.707**	**0.496**
IAC-IR (Ours)		24.605	0.724	**4.303**	**65.676**	**0.673**	**0.547**		**4.457**	**71.942**	**0.758**	**0.578**
DiffIR	RealSR	25.430	0.752	2.699	55.992	0.372	0.325	RealPhoto	3.973	66.545	0.626	0.409
SinSR		24.523	0.708	3.193	60.609	0.625	0.403		2.746	54.515	0.708	0.443
OSEDiff		23.589	0.707	**4.068**	**69.091**	0.668	0.472		4.625	71.328	0.749	0.483
InvSR		22.559	0.685	4.039	68.537	**0.679**	0.455		1.863	30.171	0.418	0.298
HYPIR		21.408	0.656	3.970	66.247	0.638	**0.476**		**4.686**	**73.048**	**0.778**	**0.581**
IAC-IR (Ours)		22.509	0.672	**4.668**	**69.853**	**0.709**	**0.589**		**4.829**	**74.089**	**0.805**	**0.635**
DiffIR	ImageNet	25.516	0.704	3.807	65.386	0.511	0.353	RealSet80	3.502	63.092	0.564	0.352
SinSR		24.925	0.703	3.939	67.652	0.662	0.444		1.863	63.644	0.710	0.426
OSEDiff		23.056	0.666	4.561	71.751	0.678	0.459		4.196	70.034	0.713	0.462
InvSR		22.018	0.648	4.560	72.382	**0.711**	0.469		**4.291**	**69.797**	0.727	0.466
HYPIR		21.116	0.609	**4.606**	**72.693**	0.710	**0.561**		4.247	69.085	**0.731**	**0.504**
IAC-IR (Ours)		22.187	0.610	**4.721**	**73.546**	**0.755**	**0.617**		**4.623**	**71.762**	**0.778**	**0.584**



(a) LQ Input **(b)** DiffIR **(c)** SinSR **(d)** OSEDiff **(e)** InvSR **(f)** HYPIR **(g)** **IAC-IR**

Fig. 2: Visual comparison on various LQ benchmark dataset. (Zoom-in for best view)

whereas DRealSR and RealSR include paired ground truth captured using specialized camera settings.

Metrics. To comprehensively evaluate image restoration performance, we employ both reference-based and no-reference metrics. Specifically, for datasets that do not provide corresponding high-quality ground-truth images, we report only no-reference metrics, including LIQE [60], CLIP-IQA [35], MUSIQ [18], and MANIQA [48]. For datasets with available ground-truth images, we additionally report reference-based metrics, including PSNR and SSIM. All metrics are computed in the standard RGB color space, except for SSIM, which is evaluated on the luminance (Y) channel.

Experimental setups. Starting from the pretrained weights of Stable Diffusion 2.1 (SD 2.1), we fully train the encoder $\mathcal{E}_\theta$ and apply LoRA [10] to the UNet $\mathcal{U}_\phi$. During training, we use paired LQ and HQ image patches of size $512 \times 512 \times 3$, which matches the resolution used to train SD 2.1. The LQ images are generated by randomly degrading the HQ images following the ResShift protocol [56]. We train the model on four H100 GPUs with a batch size of 4 per GPU using the AdamW optimizer [22] for 300k iterations. The initial learning rate is set to 1×10^{-4} and is reduced by a factor of 0.5 every 100k iterations.

4.2 Comparison with State-of-the-Art Methods

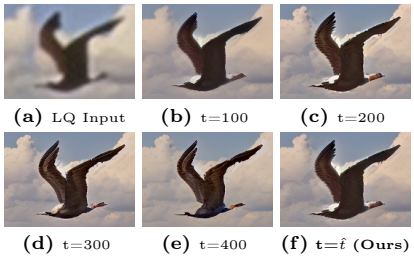
Table 1 presents a comprehensive quantitative comparison between IAC-IR and previous image restoration (IR) methods. Here, we include recent single-step diffusion-based approaches, covering both super-resolution methods [42, 45, 55] and general IR methods [21, 47]. All compared methods (including our IAC-IR) adopt a single-step inference strategy for efficiency. Following the standard evaluation protocol, all LQ inputs—except those from RealPhoto—are first 4× upsampled by a factor of four using bicubic interpolation before restoration, while RealPhoto inputs remain at their original resolution. For datasets that provide ground-truth HQ images, we report reference-based metrics (*i.e.*, PSNR and SSIM); for datasets without ground truth, we report non-reference perceptual metrics (*i.e.*, LIQE, MUSIQ, CLIPIQA, and MANIQA).

As shown in the table, IAC-IR consistently outperforms prior methods by a large margin in perceptual metrics across all datasets, including both real-world [1, 44, 46, 54, 56] and synthetic degradations [3]. Although PSNR and SSIM are not always the highest, IAC-IR achieves competitive distortion performance and consistently surpasses HYPIR. Depending on the metric, OSEDiff [45], InvSR [55], or HYPIR [21] occasionally rank second; however, IAC-IR consistently achieves the best overall perceptual performance across evaluation criteria. Notably, HYPIR exhibits relatively poor performance on real LQ datasets captured with focal adjustments (*e.g.*, DRealSR and RealSR), and InvSR shows a significant performance drop on RealPhoto, where inputs are not bicubic-upsampled. In contrast, IAC-IR demonstrates stable and strong performance across all scenarios. These consistent improvements validate the effectiveness of input-aware corruption in aligning degraded inputs with the pretrained diffusion trajectory, resulting in more reliable and perceptually faithful restoration.

Furthermore, Figure 2 presents visual comparisons between IAC-IR and prior methods. As shown, other approaches struggle to restore the spider's legs—often even producing metallic artifacts (first row of Figure 2f)—whereas our method accurately reconstructs the legs with fine hair details (first row of Figure 2g). Similarly, for the defocused man and Spider-Man examples, our approach better preserves the LQ input and restores clear facial structures, while other methods retain blur or generate unrealistic facial features. Moreover, Figure 6 further provides visual comparisons on historical real-world images. Additional diverse comparisons can be found in the supplementary material.

Table 2: Quantitative comparison of our IAC-IR on RealSR dataset with various timesteps. For the ablation variants, we adopt fixed timesteps (*i.e.*, 100, 200, and 300), whereas our final framework predicts the timestep $\hat{t}$ directly from the LQ input $\mathbf{I}_{\text{LQ}}$.

Timestep	PSNR↑	SSIM↑	LIQE↑	MUSIQ↑	CLIPIQA↑	MANIQA↑
100	**23.064**	0.660	3.816	66.122	0.694	0.559
200	22.077	0.669	4.448	69.756	0.684	0.559
300	20.707	0.628	4.464	69.312	0.701	0.578
400	20.492	0.615	4.571	**69.927**	**0.709**	0.588
$\hat{t}$ (Ours)	22.509	**0.672**	**4.668**	69.853	**0.709**	**0.589**
t_{GT}	22.698	0.690	4.656	70.466	0.717	0.606



(a) LQ Input **(b)** t=100 **(c)** t=200 **(d)** t=300 **(e)** t=400 **(f)** t=$\hat{t}$ **(Ours)**

Fig. 3: Visual comparison under different timestep. The captions below each image note the timestep used. ($\hat{t} = 156$ at this sample).

4.3 Effect of Input-Aware Corruption (IAC)

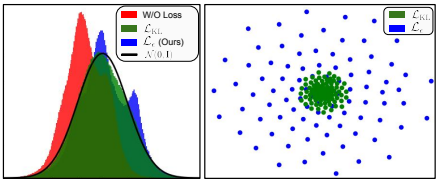
Ablation study on $\hat{t}$. Table 2 compares our predicted timestep strategy with fixed timestep settings. For a fair comparison, we retrain the model under each fixed timestep configuration and use the corresponding fixed noise level. Fixing the timestep to a larger value—thereby allowing greater generative freedom—improves perceptual metrics, but severely degrades distortion-based metrics. Conversely, fixing the timestep to a smaller value—thus preserving more of the original input—yields better distortion-based performance, while significantly compromising perceptual quality. This highlights the inherent trade-off between perceptual quality and distortion when using a fixed timestep. Notably, our predicted timestep strategy consistently achieves significantly better perceptual performance than any fixed-timestep setting, while maintaining competitive distortion metrics. Moreover, the last row shows the performance using the ground-truth timestep t_{GT} derived from the corresponding HQ images. Using t_{GT} consistently improves both distortion- and perceptual-based metrics, highlighting the importance of accurately predicting the timestep for each sample.

Figure 3 further supports this observation. Larger fixed timesteps (*e.g.*, Figures 4d and 4e) enhance perceptual realism but tend to distort original content. Conversely, smaller fixed timesteps (*e.g.*, Figures 4b and 4c) better preserve input structure but produce less perceptually pleasing results. In contrast, using the predicted timestep $\hat{t}$ (*e.g.*, Figure 4f) effectively balances perceptual quality and content preservation, achieving superior overall restoration performance.

Ablation study on $\hat{\epsilon}$. While noise is predicted in our IAC framework, it is not directly used during inference (as the predicted noise is composed with the image latent to form $\mathbf{X}$). Instead, we analyze its effect by varying the noise supervision term $\mathcal{L}_\epsilon$. Specifically, Table 3 compares variants where no supervision is applied to the predicted noise (the first row), and the proposed $\mathcal{L}_\epsilon$ is replaced with a KL divergence loss $\mathcal{L}_{\text{KL}}$ (the second row) that explicitly enforces the predicted noise to follow a Gaussian distribution. When no noise supervision is applied (similar to InvSR [55]), performance degrades noticeably. Moreover, although $\mathcal{L}_{\text{KL}}$ constrains the predicted noise to match a Gaussian distribution, it

Table 3: Quantitative comparison of IAC-IR on the RealSR dataset under different noise supervision strategies. For the ablation variants, we either omit the noise loss or replace it with a KL-divergence-based loss term, $\mathcal{L}_{\text{KL}}$, whereas our final model employs the proposed $\mathcal{L}_\epsilon$.

Loss for noise	LIQE↑	MUSIQ↑	CLIPIQA↑	MANIQA↑
W/O loss	4.004	67.256	0.626	0.483
$\mathcal{L}_{\text{KL}}$	4.218	69.363	0.659	0.523
$\mathcal{L}_\epsilon$ (Ours)	**4.668**	**69.853**	**0.709**	**0.589**



(a) Distribution of $\hat{\epsilon}$ **(b)** t-sne of $\hat{\epsilon}$

Fig. 4: Distribution histogram and t-SNE projection of predicted noise $\hat{\epsilon}$. For the t-sne, we reshape each predicted noise $\hat{\epsilon}$ into a single vector before projection; thus, each point in (b) represents the noise corresponding to one input sample.

Table 4: Quantitative comparison of IAC-IR with and without $\mathcal{L}_t$ and $\mathcal{L}_\epsilon$. The markers ✓and ✗indicate whether the corresponding loss is applied or not, respectively.

$\mathcal{L}_t$	$\mathcal{L}_\epsilon$	LIQE↑	MUSIQ↑	CLIPIQA↑	MANIQA↑
✗	✗	3.906	66.482	0.610	0.457
✓	✗	4.004	67.256	0.626	0.483
✗	✓	4.448	69.756	0.684	0.559
✓	✓	**4.668**	**69.853**	**0.709**	**0.589**

Table 5: Quantitative comparison of IAC-IR using different variants of $\mathcal{L}_{\text{distill}}$. 'Rand' denotes random sampling.

$\mathcal{L}_{\text{distill}}$ Timestep	Noise	LIQE↑	MUSIQ↑	CLIPIQA↑	MANIQA↑
Rand	Rand	3.990	68.120	0.636	0.479
$\hat{t}$	Rand	4.174	69.234	0.676	0.506
Rand	$\hat{\epsilon}$	4.045	68.519	0.653	0.499
$\hat{t}$	$\hat{\epsilon}$	**4.668**	**69.853**	**0.709**	**0.589**

also results in inferior performance compared to our proposed $\mathcal{L}_\epsilon$. These results demonstrate that simply enforcing Gaussianity by applying KL divergence is insufficient; input-aware noise supervision is critical.

Figure 4 provides further visual analysis. As shown in Figure 5a, although the predicted timestep $\hat{t}$ is selected to align the corrupted sample with a Gaussian distribution, this alone does not guarantee that the predicted noise follows the true Gaussian distribution. Without explicit supervision, the predicted noise deviates from Gaussian characteristics (see red bars in Figure 5a). Furthermore, replacing $\mathcal{L}_\epsilon$ with $\mathcal{L}_{\text{KL}}$ produces noise that appears Gaussian, similar to our method (see ■ in Figure 5a). However, a key difference emerges in t-SNE [23] map in Figure 5b: under $\mathcal{L}_{\text{KL}}$, noises from different input samples exhibit very small variation (see ● in Figure 5b), indicating that the model predicts nearly identical noise regardless of input content. In contrast, our $\mathcal{L}_\epsilon$ encourages input-dependent diversity while preserving Gaussian properties (see ■ in Figure 5a and ● in Figure 5b). For, additional visualizations of the predicted noise using our $\mathcal{L}_\epsilon$, please refer to the supplementary material.

4.4 Effect of Loss Functions

Table 4 shows an analysis of the loss components, $\mathcal{L}_t$ and $\mathcal{L}_\epsilon$. All variants are trained with the baseline losses and $\mathcal{L}_{\text{distill}}$, while selectively enabling or disabling the two additional terms. When $\mathcal{L}_t$ is omitted, the predicted timestep $\hat{t}$ collapses to small values, leading to unstable training; therefore, we instead use a fixed timestep (*i.e.*, 200), and $\mathcal{L}_{\text{distill}}$ is computed accordingly with this fixed timestep.

Table 6: Comparison of IAC-IR with different Gaussianity-objectives.

Measure	PSNR$_\uparrow$	SSIM$_\uparrow$	LIQE$_\uparrow$	MUSIQ$_\uparrow$	CLIPIQA$_\uparrow$	MANIQA$_\uparrow$
KL	21.190	0.652	4.665	**69.896**	0.703	0.573
$\mathcal{G}(\cdot)$ **(Ours)**	**22.509**	**0.672**	**4.668**	69.853	**0.709**	**0.589**

Table 7: User-study comparison of IAC-IR with previous diffusion-based IR methods.

DiffIR	OSEDiff	InvSR	HYPIR	**IAC-IR (Ours)**
6.67	16.67	3.34	20.00	53.34

Compared to the first row, introducing timestep prediction supervised by $\mathcal{L}_t$ (second row) consistently improves performance. Moreover, comparing the first and third rows shows that adding $\mathcal{L}_\epsilon$ also enhances performance, even under a fixed timestep setting. Finally, the last row, which incorporates both $\mathcal{L}_t$ and $\mathcal{L}_\epsilon$ in our IAC framework, achieves the best results, indicating that these loss terms are complementary and synergistic when combined.

Table 5 further analyzes variants of $\mathcal{L}_{\text{distill}}$. While our final model performs score-based distillation using samples corrupted with both the predicted timestep $\hat{t}$ and predicted noise $\hat{\epsilon}$, the variants instead use random timestep or random noise for corruption, following the conventional distillation framework [4,43,45,51,53]. As shown in the table, using the predicted timestep $\hat{t}$ (second row) yields a larger performance gain than using the predicted noise $\hat{\epsilon}$ alone (third row). Importantly, leveraging both predicted factors together (last row) consistently achieves the best performance across all metrics. This validates that integrating predicted corruption into distillation leads to more effective knowledge transfer.

4.5 Effect of $\mathcal{L}_\epsilon$ for the Gaussianity

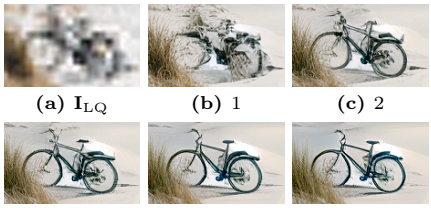
As discussed in Section 3.3, we use $\mathcal{G}(\cdot)$ to encourage the predicted noise $\hat{\epsilon}$ to exhibit Gaussian-like statistics. Although more stringent distribution-matching objectives, such as the KL divergence (KL in Table 6), could also be considered, we find that they consistently lead to inferior reconstruction performance, as shown in Table 6. We attribute this result to the spatially heterogeneous nature of real-world degradations. Specifically, the amount and type of degradation can vary substantially across different image regions; some regions may require strong generative correction, whereas others already preserve sufficient visual information and should remain close to the input. Enforcing the predicted noise to strictly follow a uniform Gaussian distribution across all spatial locations can therefore introduce unnecessary or excessive generation in well-preserved regions, ultimately degrading reconstruction fidelity. In contrast, $\mathcal{G}(\cdot)$ provides a more flexible Gaussianity constraint that encourages appropriate noise behavior without overly constraining the spatially varying restoration process.

4.6 User study

Table 7 presents the results of our user study, conducted to evaluate the perceptual quality of restored images. Here, we recruited 18 participants and presented them with the degraded LQ input together with restored outputs generated by different competing methods. Given the same input image, participants were asked to select the restoration that they considered visually most favorable in

Table 8: Quantitative comparison of IAC-IR with various number of steps. Despite its controllability, single-step version still shows high-quality results.

# of iter.	PSNR↑	SSIM↑	LIQE↑	MUSIQ↑	CLIPIQA↑	MANIQA↑
1	22.509	0.672	4.668	69.853	0.709	0.589
2	21.220	0.643	4.511	70.918	0.718	0.613
3	20.230	0.619	4.560	71.139	0.720	0.626
4	19.445	0.600	4.571	71.111	0.714	0.633
5	18.798	0.585	4.569	70.996	0.706	0.635



(a) $\mathbf{I}_{\text{LQ}}$ **(b)** 1 **(c)** 2 **(d)** 3 **(e)** 4 **(f)** 5

Fig. 5: Visual comparison under different # of iterations.

Table 9: Efficiency comparison with previous diffusion-based IR methods. For the measurement, we restore images sized $\mathbb{R}^{512\times512\times3}$ using a A100 GPU.

Measure	StableSR (50)	DiffBIR (50)	DiffIR	SinSR	OSEDiff	InvSR	HYPIR	**IAC-IR (Ours)**
Latency↓ (s/f)	3460	2720	147	140	177	**117**	220	**110**
GMACs↓	19980	24234	**184**	2649	2265	2123	1784	**1384**

terms of overall image quality, realism, and faithfulness to the input content. As shown in the table, our IAC-IR receives the highest preference rate among all compared methods. This result indicates that the images produced by IAC-IR are more strongly aligned with human visual perception, achieving a favorable balance between perceptual realism and preservation of input-relevant details.

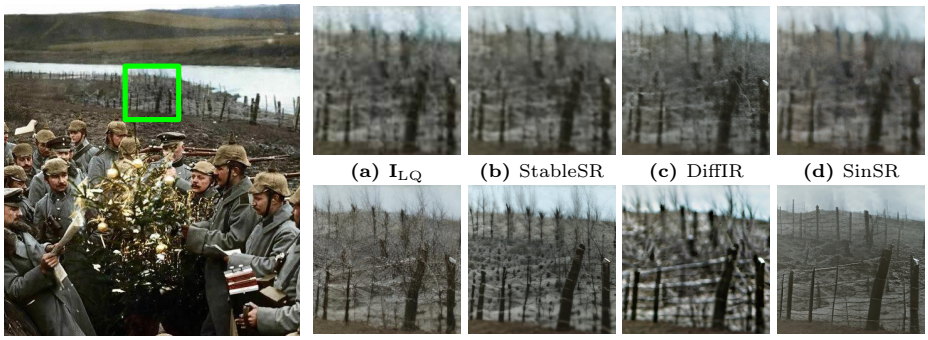
4.7 Discussion on Sampling Steps

Although our IAC-IR operates in a single step, it can be extended to a multi-step setting. Unlike previous approaches [20, 42] that repeatedly apply the denoising network to iteratively refine latent representations, we propose a novel iterative scheme. As our encoder $\mathcal{E}_\theta$ predicts both timestep and noise directly from the input image, we reapply the entire IAC-IR pipeline at each iteration, using the restored image $\mathbf{I}_{\text{IR}}$ from the previous step as the new input. Consequently, instead of heuristically decreasing the timestep according to a predefined scheduler, the network automatically determines the timestep and noise at each iteration based on the updated image.

Table 8 presents quantitative comparisons across different numbers of iterations. As the number of iterations increases, perceptual quality consistently improves. However, because additional iterations progressively modify the original content, distortion-based metrics degrade with more steps. Figure 5 further shows this trade-off: increasing the number of iterations enhances perceptual realism but may introduce greater deviation from the original input. Specifically, although increasing the number of iterations leads to greater distortion, the bicycle appears more perceptually pleasing as the iterations progress.

4.8 Efficiency Comparison

Table 9 compares IAC-IR with prior IR methods, showing that ours achieves the lowest latency. InvSR [55] introduces additional noise prediction, and HYPIR [21]



(a) $\mathbf{I}_{\mathrm{LQ}}$ **(b)** StableSR **(c)** DiffIR **(d)** SinSR

(e) OSEDiff **(f)** InvSR **(g)** HYPIR **(h)** IAC-IR

Fig. 6: Visual comparison on severely degraded historical real-world images.

adopts a larger Transformer backbone, both increasing inference time. Conversely, IAC-IR preserves the SD 2.1 pipeline and adds only a lightweight timestep estimator $\mathcal{T}$ (2% of total parameters), enabling the fastest inference while maintaining superior perceptual quality.

4.9 Limitations

While IAC-IR shows strong perceptual performance, as shown in Table 1 and Figure 2, it does not always achieve the best results in distortion-oriented metrics. This outcome reflects the well-known trade-off between perceptual quality and distortion accuracy, where methods optimized for visually pleasing and realistic outputs may sacrifice pixel-wise fidelity. Understanding the importance of balancing both aspects, as future work, we plan to introduce controllable mechanisms that allow users to adjust the trade-off between perceptual quality and distortion fidelity according to their preferences.

5 Conclusion

In this paper, we propose IAC-IR, a novel diffusion-based image restoration framework that replaces heuristic corruption strategies with input-aware corruption. By predicting an optimal timestep and noise for each degraded input, our method aligns corrupted samples with the pretrained diffusion trajectory, enabling more faithful and stable denoising. We further introduce an input-aware distillation strategy that produces more reliable score supervision compared to conventional random corruption. Extensive experiments show that IAC-IR achieves superior perceptual quality with competitive distortion performance, and offers a promising direction for extending input-aware corruption to other low-level vision tasks such as image editing and image inpainting.

Acknowledgments. This work was supported in part by the IITPgrants [No. RS-2021-II211343, Artificial Intelligence Graduate School Program (Seoul National University), No. RS-2025-02303870, No.2022-0-00156] funded by the Korea government (MSIT) and Mobile eXperience (MX) Business, Samsung Electronics Co., Ltd.

References

1. Cai, J., Zeng, H., Yong, H., Cao, Z., Zhang, L.: Toward real-world single image super-resolution: A new benchmark and a new model. In: ICCV (2019) 2, 8, 10
2. Chung, H., Sim, B., Ye, J.C.: Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction. In: CVPR (2022) 1, 3
3. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: CVPR (2009) 2, 8, 10
4. 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. CVPR (2025) 2, 4, 13
5. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020) 6
6. Du, P., Ning, J., Cui, J., Huang, S., Wang, X., Wang, J.: Geometric structure preserving warp for natural image stitching. In: CVPR (2022) 1
7. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. NeurIPS (2014) 1, 3, 8
8. Hinton, G., Vinyals, O., Dean, J.: Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531 (2015) 2
9. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. NeurIPS (2020) 1, 2, 3
10. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al.: Lora: Low-rank adaptation of large language models. ICLR (2022) 3, 7, 10
11. Kawar, B., Elad, M., Ermon, S., Song, J.: Denoising diffusion restoration models. NeurIPS (2022) 1, 3
12. Kim, J., Lee, J.K., Lee, K.M.: Accurate image super-resolution using very deep convolutional networks. In: CVPR (2016) 1, 3
13. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013) 5
14. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: CVPR (2017) 1, 3
15. Li, Y., Zhang, K., Liang, J., Cao, J., Liu, C., Gong, R., Zhang, Y., Tang, H., Liu, Y., Demandolx, D., et al.: Lsdir: A large scale dataset for image restoration. In: CVPR (2023) 8
16. Liang, J., Zeng, H., Zhang, L.: Details or artifacts: A locally discriminative learning approach to realistic image super-resolution. In: CVPR (2022) 1
17. Liang, J., Zeng, H., Zhang, L.: Efficient and degradation-adaptive network for real-world image super-resolution. In: ECCV (2022) 3
18. Liang, J., Zhang, K., Gu, S., Van Gool, L., Timofte, R.: Flow-based kernel prior with application to blind super-resolution. In: CVPR (2021) 9
19. Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced deep residual networks for single image super-resolution. In: CVPR (2017) 1, 3
20. 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 (2024) 1, 2, 3, 14

21. Lin, X., Yu, F., Hu, J., You, Z., Shi, W., Ren, J.S., Gu, J., Dong, C.: Harnessing diffusion-yielded score priors for image restoration. arXiv preprint arXiv:2507.20590 (2025) 1, 2, 3, 4, 10, 14
22. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017) 10
23. Van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. JMLR (2008) 12
24. Meng, C., Rombach, R., Gao, R., Kingma, D., Ermon, S., Ho, J., Salimans, T.: On distillation of guided diffusion models. In: CVPR (2023) 4
25. Moran, S., Marza, P., McDonagh, S., Parisot, S., Slabaugh, G.: DeepLPF: Deep local parametric filters for image enhancement. In: CVPR (2020) 1
26. Park, J., Lee, K.M.: Bridging the distribution gap to harness pretrained diffusion priors for super-resolution. In: The Fourteenth International Conference on Learning Representations 1
27. Park, J., Son, S., Lee, K.M.: Content-aware local gan for photo-realistic super-resolution. In: ICCV (2023) 1, 3
28. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: CVPR (2022) 1, 2, 3
29. Sauer, A., Lorenz, D., Blattmann, A., Rombach, R.: Adversarial diffusion distillation. arXiv preprint arXiv:2311.17042 (2023) 4
30. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014) 8
31. Song, D.Y., Lee, G., Lee, H., Um, G.M., Cho, D.: Weakly-supervised stitching network for real-world panoramic image generation. In: ECCV (2022) 1
32. Timofte, R., Agustsson, E., Van Gool, L., Yang, M.H., Zhang, L.: Ntire 2017 challenge on single image super-resolution: Methods and results. In: CVPR Workshops (2017) 8
33. Umer, R.M., Micheloni, C.: Deep cyclic generative adversarial residual convolutional networks for real image super-resolution. In: ECCV (2020) 1
34. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. NeurIPS (2017) 3
35. Wang, J., Chan, K.C., Loy, C.C.: Exploring clip for assessing the look and feel of images. In: AAAI (2023) 9
36. Wang, J., Yue, Z., Zhou, S., Chan, K.C., Loy, C.C.: Exploiting diffusion prior for real-world image super-resolution. IJCV (2024) 1, 2, 3
37. Wang, X., Girshick, R., Gupta, A., He, K.: Non-local neural networks. In: CVPR (2018) 3
38. Wang, X., Xie, L., Dong, C., Shan, Y.: Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In: ICCV (2021) 1
39. Wang, X., Yu, K., Dong, C., Loy, C.C.: Recovering realistic texture in image super-resolution by deep spatial feature transform. In: CVPR (2018) 3
40. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Change Loy, C.: Esrgan: Enhanced super-resolution generative adversarial networks. In: ECCV Workshops (2018) 1, 3
41. Wang, Y., Yu, J., Zhang, J.: Zero-shot image restoration using denoising diffusion null-space model. ICLR (2023) 1, 3
42. Wang, Y., Yang, W., Chen, X., Wang, Y., Guo, L., Chau, L.P., Liu, Z., Qiao, Y., Kot, A.C., Wen, B.: Sinsr: diffusion-based image super-resolution in a single step. In: CVPR (2024) 2, 10, 14
43. Wang, Z., Lu, C., Wang, Y., Bao, F., Li, C., Su, H., Zhu, J.: Prolificdreamer: High-fidelity and diverse text-to-3d generation with variational score distillation. NeurIPS (2023) 2, 4, 8, 13

44. Wei, P., Xie, Z., Lu, H., Zhan, Z., Ye, Q., Zuo, W., Lin, L.: Component divide-and-conquer for real-world image super-resolution. In: ECCV (2020) 2, 8, 10
45. Wu, R., Sun, L., Ma, Z., Zhang, L.: One-step effective diffusion network for real-world image super-resolution. NeurIPS (2024) 2, 4, 10, 13
46. Wu, R., Yang, T., Sun, L., Zhang, Z., Li, S., Zhang, L.: Seesr: Towards semantics-aware real-world image super-resolution. In: CVPR (2024) 1, 2, 8, 10
47. Xia, B., Zhang, Y., Wang, S., Wang, Y., Wu, X., Tian, Y., Yang, W., Van Gool, L.: Diffir: Efficient diffusion model for image restoration. In: ICCV (2023) 10
48. Yang, S., Wu, T., Shi, S., Lao, S., Gong, Y., Cao, M., Wang, J., Yang, Y.: Maniqa: Multi-dimension attention network for no-reference image quality assessment. In: CVPR (2022) 9
49. Yang, T., Wu, R., Ren, P., Xie, X., Zhang, L.: Pixel-aware stable diffusion for realistic image super-resolution and personalized stylization. In: ECCV (2024) 2, 3
50. Yi, R., Tian, H., Gu, Z., Lai, Y.K., Rosin, P.L.: Towards artistic image aesthetics assessment: a large-scale dataset and a new method. In: CVPR (2023) 1
51. Yin, T., Gharbi, M., Park, T., Zhang, R., Shechtman, E., Durand, F., Freeman, B.: Improved distribution matching distillation for fast image synthesis. NeurIPS (2024) 4, 8, 13
52. Yin, T., Gharbi, M., Zhang, R., Shechtman, E., Durand, F., Freeman, W.T., Park, T.: One-step diffusion with distribution matching distillation. arXiv preprint arXiv:2311.18828 (2023) 2, 8
53. Yin, T., Gharbi, M., Zhang, R., Shechtman, E., Durand, F., Freeman, W.T., Park, T.: One-step diffusion with distribution matching distillation. In: CVPR (2024) 4, 13
54. Yu, F., Gu, J., Li, Z., Hu, J., Kong, X., Wang, X., He, J., Qiao, Y., Dong, C.: Scaling up to excellence: Practicing model scaling for photo-realistic image restoration in the wild. In: CVPR (2024) 1, 2, 3, 8, 10
55. Yue, Z., Liao, K., Loy, C.C.: Arbitrary-steps image super-resolution via diffusion inversion. CVPR (2025) 2, 4, 7, 10, 11, 14
56. Yue, Z., Wang, J., Loy, C.C.: Resshift: Efficient diffusion model for image super-resolution by residual shifting. NeurIPS (2023) 1, 2, 8, 10
57. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: CVPR. pp. 5728–5739 (2022) 3
58. Zhang, K., Liang, J., Van Gool, L., Timofte, R.: Designing a practical degradation model for deep blind image super-resolution. In: ICCV (2021) 1
59. Zhang, L., Rao, A., Agrawala, M.: Adding conditional control to text-to-image diffusion models. In: ICCV (2023) 3
60. Zhang, W., Zhai, G., Wei, Y., Yang, X., Ma, K.: Blind image quality assessment via vision-language correspondence: A multitask learning perspective. In: CVPR (2023) 9