DS-DEBLUR: DECOUPLED SUPERVISION FOR DIFFUSION-BASED IMAGE DEBLURRING

Anonymous authors
Paper under double-blind review

ABSTRACT

Real-world image deblurring remains challenging due to a fundamental mismatch between blur realism and ground-truth fidelity. Real datasets capture authentically blurred images but contain noisy references, whereas synthetic datasets offer clean supervision but fail to reflect realistic motion patterns. This discrepancy leads to a prior conflict with pretrained diffusion models, whose generative priors are learned from clean images, thereby limiting their effectiveness in real motion blur removal. To address this issue, we propose DS-DeBlur, a diffusion-based image deblurring model built upon a new Real-Synthetic Decoupled Supervision (ReSyDS) paradigm. The paradigm explicitly disentangles the learning of blur realism and supervision fidelity through two complementary paths. The synthetic path leverages clean paired data to unleash diffusion priors and preserve structural fidelity, while the real path directly learns from authentically blurred images to promote perceptual realism and improve generalization without relying on imperfect references. To further support ReSyDS, we introduce an instance-driven realistic blur synthesis method, a physically-grounded approach to generate high-quality and realistic blur–sharp pairs for effective supervision. Extensive experiments demonstrate that the proposed method performs favorably against state-of-the-art deblurring methods.

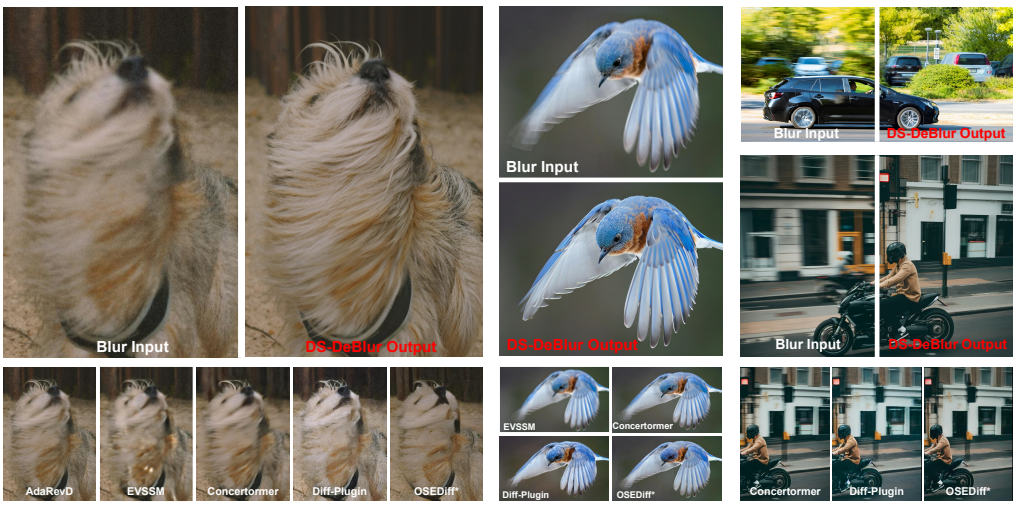

Figure 1: Visual comparison in challenging real-world scenarios: Our DS-DeBlur consistently restores photo-realistic details across diverse blur patterns, establishing a clear superiority over previous approaches. **Zoom in for details.**

1 INTRODUCTION

Image deblurring aims to recover sharp content from spatially-varying motion and defocus blurs arising in dynamic, real-world scenes. Although deep learning-based approaches have achieved

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

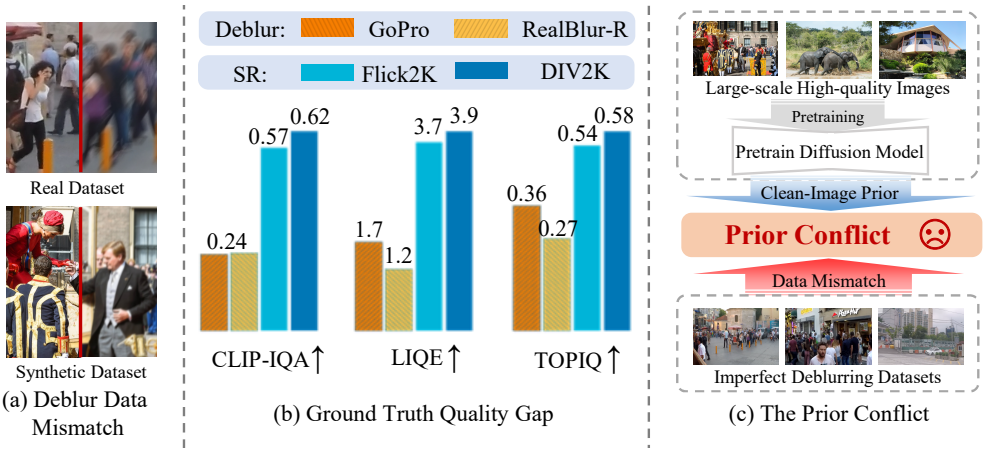

Figure 2: The core motivation for our work: (a) Deblur Data Mismatch: Real datasets offer authentic blur but have relatively low-quality references, while synthetic datasets provide clean references but with unrealistic blur; (b) Ground Truth Quality Gap: This mismatch is empirically quantified by the lower no-reference IQA scores of deblurring GTs compared to super-resolution (SR) GTs; (c) The Prior Conflict: The clean-image priors of pretrained models clash with the imperfect deblurring datasets from (a) and (b), hindering their overall effectiveness.

significant progress Chakrabarti (2016); Tao et al. (2018); Mehri et al. (2021), they often struggle to restore photo-realistic textures, especially in complex scenarios.

Recently, pretrained diffusion models have demonstrated remarkable success in image restoration tasks Yu et al. (2024); Wu et al. (2024c); Lin et al. (2024), benefiting from powerful generative priors Song et al. (2021); Ho et al. (2020); Rombach et al. (2022). However, transferring these priors to deblurring tasks is hindered by a data mismatch between blur realism and ground-truth fidelity. The mismatch is evident in mainstream deblurring datasets, which can be categorized into two types based on the origin of their motion. Real datasets (*e.g.*, GoPro Nah et al. (2017b) and RealBlur Rim et al. (2020)) capture blurs from actual cameras or scene motion, providing realistic spatially-varying degradations. However, their sharp references require short exposures, leading to increased noise and degrading image quality. In contrast, synthetic datasets (*e.g.*, DeblurGAN Kupyn et al. (2018) and GSBlur Lee et al. (2024)) simulate motion on clean images, offering noise-free ground truths, but the synthesized blur is oversimplified and lacks realistic characteristics (Fig. 2a and 2b). Consequently, no single dataset offers both realistic blur formation and high-fidelity ground truth.

The underlying cause of this data mismatch lies in the physics of photometry Zhang et al. (2022); Xiang et al. (2025). In dynamic scenes, capturing a sharp reference requires a short exposure, which drastically reduces photon counts and leads to a low signal-to-noise ratio. Cameras therefore increase the ISO level, amplifying sensor noise and degrading image quality. Conversely, longer exposures suppress noise but inevitably introduce motion blur. This inherent conflict is so fundamental that even state-of-the-art computational photography solutions Lai et al. (2022), such as those in Google Pixel phones, rely on complex fusion methods using long and short exposure images. This data mismatch leads to a prior conflict for pretrained diffusion models Rombach et al. (2022); Labs (2024), which rely on high-quality, clean images for training. When fine-tuned for deblurring, the models are misaligned with imperfect training pairs that contain noisy references or unrealistic blur, failing to fully leverage their generative priors (Fig. 2c). This prior conflict hinders the effectiveness of diffusion models, limiting their ability to restore photo-realistic textures.

To address this challenge, we propose DS-DeBlur, a diffusion-based model for image deblurring that resolves the prior conflict through a novel Real-Synthetic Decoupled Supervision (ReSyDS) paradigm. This paradigm decouples the learning of realism and fidelity into two complementary paths: the synthetic path leverages clean data to unleash diffusion priors and preserves structural fidelity, and the real path uses real blurred images to learn the real-world degradations and improve generalization. To support this paradigm, we introduce two additional components. First, we propose Instance-Driven Realistic Blur Synthesis (InstaBlur), a physically grounded technique for generating high-quality blur-sharp pairs for the synthetic path. Second, the Motion-Strength Modeling

(MSM) module, an adaptive component to handle the complex, spatially-variant blur. In summary, the main contributions are as follows: (i) We propose DS-DeBlur, a diffusion-based model for realistic image deblurring that addresses the prior conflict through the novel Real-Synthetic Decoupled Supervision (ReSyDS) paradigm. (ii) We introduce two additional components: the Instance-Driven Realistic Blur synthesis (InstaBlur) pipeline for generating high-quality synthetic blur-sharp pairs and the Motion-Strength Modeling (MSM) module for adaptive spatially-varying motion removal. (iii) Extensive experiments demonstrate that DS-DeBlur outperforms SOTA methods in terms of generalization and perceptual quality, delivering more visually pleasing results with superior clarity and naturalness.

2 Related Work

Image Deblurring. Early image deblurring methods relied on handcrafted priors, which performed poorly on complex, real-world blur Fergus et al. (2006); Levin et al. (2009); Shan et al. (2008); Michaeli & Irani (2014). The advent of deep learning led to significant progress with CNN and Transformer-based models Abuolaim & Brown (2020); Nimisha et al. (2017) that improved reconstruction via multi-scale Nah et al. (2017a) or attention mechanisms Tsai et al. (2022). However, two persistent challenges remain. First, the reliance on pixel-wise losses often leads to over-smoothed results that lack photo-realistic textures. This motivates the use of generative models, like diffusion models, to restore realistic details. Second, all supervised methods are constrained by the data mismatch in deblurring datasets: no existing paradigm offers both authentic blur and clean ground truths. This limits their generalization to real-world scenarios Xiang et al. (2025). Although unsupervised approaches Nimisha et al. (2018); Lu et al. (2019) exist, they often struggle with training stability. Our work addresses these two challenges directly: we leverage diffusion priors to restore realistic textures and introduce a decoupled supervision paradigm to overcome the data limitation.

Deblurring Datasets. Training data is pivotal for deblurring, yet all existing paradigms reflect the data mismatch between blur realism and ground-truth fidelity. We categorize them into two main types based on the origin of their motion. Real datasets Nah et al. (2017b); Rim et al. (2020; 2022); Zhang et al. (2020); Li et al. (2023); Nah et al. (2019) capture blur from real physical movement. The most common approach, seen in datasets like GoPro Nah et al. (2017b), generates blur by averaging video frames. Although scalable, this method produces blur that lacks realism, and its short-exposure ground truths are inherently noisy. An alternative approach, used by RealBlur Rim et al. (2020), employs specialized hardware like beam-splitters to capture highly authentic motion blur. However, this method is costly, which limits the dataset scale and diversity. Furthermore, it still fails to solve the fidelity problem, providing noisy ground truths compromised by shot noise. Synthetic datasets Kupyn et al. (2018; 2019); Lee et al. (2024) generate blur through algorithmic simulation. Traditional methods Kupyn et al. (2018; 2019) use simple blur kernels, resulting in unrealistic and spatially uniform blur. Modern 3D rendering Lee et al. (2024) can create more diverse patterns, yet it still struggles to replicate authentic motion. Despite providing clean ground truths, these methods ultimately fail to capture realistic blur patterns. In summary, no existing approach simultaneously achieves realism, fidelity, scale, and diversity. This data limitation motivates our work in two ways: developing a new data synthesis pipeline to generate higher-quality data and introducing a novel supervision strategy to more effectively utilize existing imperfect data.

Diffusion Models for Image Restoration. Pretrained diffusion models Ho et al. (2020); Rombach et al. (2022); Labs (2024) have recently advanced image restoration by learning powerful generative priors from large-scale data. Their success has been demonstrated in various tasks, including super-resolution Wu et al. (2024c); Yu et al. (2024) and low-light enhancement Jiang et al. (2024). More recent work has also achieved efficient one-step inference Wang et al. (2024); Zhang et al. (2024); Wu et al. (2024b). However, leveraging these pretrained priors for deblurring remains challenging. Unlike structured degradations like downsampling in super-resolution, deblurring involves complex, spatially-variant artifacts that are more difficult to reverse Xiang et al. (2025), since downsampling merely discards pixels while blurring conflates them. More fundamentally, the task is constrained by the data mismatch caused by the lack of datasets that provide both authentic blur and clean ground truths. This limitation directly leads to a prior conflict when fine-tuning pretrained diffusion models, as the noisy or unrealistic supervision from deblurring datasets is misaligned with their clean-image priors. As a result, the application of these pretrained diffusion models for deblurring remains relatively underexplored Kong et al. (2025b); Liu et al. (2025); Wu et al. (2024a).

3 METHODOLOGY

3.1 ARCHITECTURE OVERVIEW

We introduce **DS-DeBlur**, a one-step diffusion model for realistic image deblurring. As illustrated in Fig. 4, DS-DeBlur is derived by fine-tuning a pretrained diffusion backbone AI (2024). The design is based on the Real-Synthetic Decoupled Supervision (ReSyDS) (Sec. 3.3), which is further supported by two key components: the Instance-Driven Realistic Blur Synthesis (InstaBlur) pipeline (Sec. 3.2) and the Motion-Strength Modeling (MSM) module (Sec. 3.4).

DS-DeBlur is adapted from a standard text-to-image pipeline Rombach et al. (2022) for image deblurring. Instead of starting from pure noise, the process begins by using a VAE encoder $\mathcal{E}$ to encode the blurry image I_{blur} into an initial latent z_{blur}. This latent representation is then refined by the MSM module into z_{refined}, followed by a denoising step to produce the clean latent z_{sharp}. Finally, the VAE decoder $\mathcal{D}$ reconstructs the sharp image $\hat{I}_{\text{sharp}}$.

3.2 INSTANCE-DRIVEN REALISTIC BLUR SYNTHESIS

To address the data limitation of existing datasets Nah et al. (2017b); Rim et al. (2020); Kupyn et al. (2018), we introduce the Instance-Driven Realistic Blur Synthesis (InstaBlur) pipeline. Unlike previous approaches Kupyn et al. (2019); Lee et al. (2024), InstaBlur is a physically-grounded and instance-aware method designed to generate data with high realism, fidelity, scale, and diversity. As illustrated in Fig. 3, the pipeline produces a large-scale dataset of complex, spatially-variant blurs paired with clean ground truths through an efficient multi-stage degradation process. InstaBlur creates realistic, spatially-variant blur by simulating the complex interaction between global camera motion and local object dynamics. To maximize realism, our pipeline applies higher-order degradations to the entire scene, including a global blur to simulate camera shake, a secondary blur for intricate optical effects, and realistic sensor noise (e.g., Gaussian, Poisson). To further enhance data diversity, we randomly select a subset of instances for local blur, resulting in a wide variety of global, local, and mixed motion patterns that closely mimic real-world capture conditions.

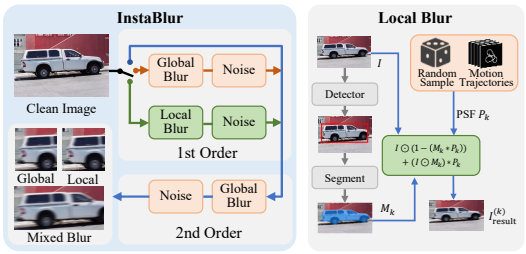

Figure 3: Overview of the InstaBlur pipeline.

Specifically, for local blur, we employ a detection-guided segmentation strategy. Inspired by the observation that real-world motion often arises from dynamic foreground objects Li et al. (2023) (*e.g.*, people or vehicles), this strategy uses a detector and a segment to produce precise pixel-level instance masks $\{M_k\}$ for each image. Second, with these masks providing spatial control, we synthesize physically-grounded motion for each instance. Then, a stochastic trajectory model Kupyn et al. (2018) generates a non-linear motion path, which is then converted into a Point Spread Function (PSF), P_k. To seamlessly blend each blurred instance, we apply the PSF to both the instance region and its mask. This process is applied iteratively for each instance k, with the final composition for a single instance given by:

$$I_{result}^{(k)} = I_{result}^{(k-1)} \odot (1 - (M_k * P_k)) + ((I \odot M_k) * P_k)\,, \tag{1}$$

where I is the clean image, $I_{\text{result}}^{(k-1)}$ is the result from the previous iteration (with $I_{\text{result}}^{(0)} = I$), $*$ is the convolution operation, and $\odot$ represents element-wise multiplication.

By combining these simple yet effective stages, InstaBlur improves upon previous data methods Nah et al. (2017b); Rim et al. (2020); Kupyn et al. (2019); Lee et al. (2024) in four key aspects: *realism*, through physically-grounded motion; *fidelity*, by pairing blurs with clean ground truths; *scale*, via its efficient synthesis process; and *diversity*, through randomized composition. The resulting dataset, with its realistic blur-sharp pairs, provides the synthetic path of our ReSyDS paradigm with the clean supervision needed to resolve the prior conflict and unleash diffusion priors.

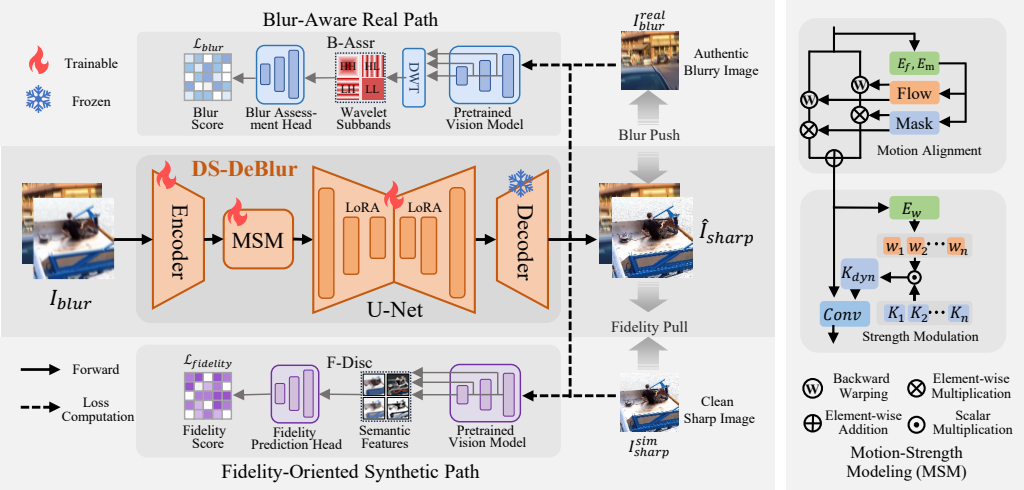

Figure 4: Architecture of DS-DeBlur: A diffusion-based model for realistic image deblurring. DS-DeBlur is trained via Real-Synthetic Decoupled Supervision (ReSyDS) and incorporates the motion-strength modeling module. The ReSyDS paradigm features two complementary paths: The fidelity-oriented synthetic path uses data from our InstaBlur pipeline to unleash diffusion priors and enforce structural fidelity via a Fidelity Discriminator. The blur-aware real path uses authentic blurry images to enhance generalization and perceptual realism via a Wavelet-Based Blur Assessor.

3.3 Real-Synthetic Decoupled Supervision

The core of our training is the Real-Synthetic Decoupled Supervision (ReSyDS) strategy, designed to address the prior conflict in diffusion-based deblurring. As illustrated in Fig. 4, ReSyDS decouples the learning process into two complementary paths: a fidelity-oriented synthetic path and a blur-aware real path. During training, both synthetic ($I^{\text{syn}}_{\text{blur}}$) and real ($I^{\text{real}}_{\text{blur}}$) blurry inputs are processed by the network to produce their respective sharp outputs ($\hat{I}^{\text{syn}}_{\text{sharp}}$, $\hat{I}^{\text{real}}_{\text{sharp}}$). Each path then receives dedicated supervision: the synthetic path leverages clean data pairs to enforce structural fidelity and unleash diffusion priors, while the real path uses authentic blurry images to enhance real-world generalization. Ultimately, ReSyDS leverages precise synthetic guidance while effectively adapting to complex real-world degradation distributions.

Fidelity-Oriented Synthetic Path. This path is designed to pull DS-DeBlur towards clean, high-quality reconstructions. It leverages clean blur-sharp pairs ($I^{\text{syn}}_{\text{blur}}$, $I^{\text{syn}}_{\text{sharp}}$) from our InstaBlur pipeline. The path is supervised by an adversarial objective provided by our Fidelity Discriminator (F-Disc) $\mathcal{D}_f$, which is guided by perceptual features from a pretrained vision model Oquab et al. (2023); Kumari et al. (2022).

During training, the $\mathcal{D}_f$ learns to distinguish DS-DeBlur output $\hat{I}_{\text{sharp}}$ from $I^{\text{syn}}_{\text{sharp}}$, by minimizing $\mathcal{L}_D$. Concurrently, DS-DeBlur is trained with a fidelity loss $\mathcal{L}_{\text{fidelity}}$ to produce outputs that the critic classifies as real:

$$\mathcal{L}_{fidelity} = -\mathbb{E}[\log(\mathcal{D}_f(\hat{I}_{\text{sharp}}))], \tag{2}$$

$$\mathcal{L}_D = -\mathbb{E}[\log(\mathcal{D}_f(I^{\text{syn}}_{\text{sharp}}))] - \mathbb{E}[\log(1 - \mathcal{D}_f(\hat{I}_{\text{sharp}}))]. \tag{3}$$

This adversarial process serves two crucial purposes. First and foremost, it directly resolves the prior conflict discussed in our introduction. The use of a clean ground truth $I^{\text{syn}}_{\text{sharp}}$ from InstaBlur provides a clean supervision signal that unleashes the powerful diffusion priors, thereby forcing DS-DeBlur to learn the properties of natural, artifact-free images. Second, as shown in previous work Zhang et al. (2024); Sauer et al. (2024), this adversarial training also functions as a form of knowledge distillation, enabling efficient one-step inference.

Blur-Aware Real Path. While our InstaBlur pipeline provides rich, diverse, and realistic training data, synthetic data inevitably fails to capture the full complexity of real-world degradations. To complement our synthetic data and enhance real-world generalization, we introduce the real path.

Conversely, this path is designed to push DS-Deblur away from realistic blur artifacts, thereby enhancing its generalization to real-world degradations. It operates on blurry images sourced from real datasets $I_{\text{blur}}^{\text{real}}$, for which no clean ground truths exist.

Consequently, its supervision relies on our Wavelet-Based Blur Assessor (B-Assr) $\mathcal{A}_b$, designed to detect residual blur. To achieve this, $\mathcal{A}_b$ operates in the wavelet domain by applying the Discrete Wavelet Transform (DWT) to features from a pretrained vision model Oquab et al. (2023). This decomposition yields wavelet subbands $\{f_{LL}, f_{LH}, f_{HL}, f_{HH}\}$, enabling a precise analysis of image frequencies. This makes $\mathcal{A}_b$ highly sensitive to the suppression of high-frequency details, a key signature of motion blur Gao et al. (2024), and thus effectively captures directional blur cues.

B-Assr $\mathcal{A}_b$ learns to distinguish the DS-DeBlur restored output $\hat{I}_{\text{sharp}}$ from the authentic blurry input $I_{\text{blur}}^{\text{real}}$, by minimizing $\mathcal{L}_A$. Concurrently, DS-DeBlur is trained with a blur repulsion loss, $\mathcal{L}_{blur}$, to produce outputs that $\mathcal{A}_b$ classifies as sharp:

$$\mathcal{L}_{blur} = -\mathbb{E}[\log(\mathcal{A}_b(\hat{I}_{\text{sharp}}))], \tag{4}$$

$$\mathcal{L}_A = -\mathbb{E}[\log(\mathcal{A}_b(\hat{I}_{\text{sharp}}))] - \mathbb{E}[\log(1 - \mathcal{A}_b(I_{\text{blur}}^{\text{real}}))]. \tag{5}$$

By learning from authentic real-world blur patterns without relying on their compromised ground truths, DS-DeBlur effectively improves its generalization.

Overall Training Objective Our final training objective is a composite loss $\mathcal{L}_{\text{total}}$, that integrates supervisory signals from both the synthetic and real paths of ReSyDS to jointly optimize DS-DeBlur:

$$\begin{aligned}\mathcal{L}_{\text{total}} = \lambda_1 \left\| \hat{I}_{\text{sharp}}^{\text{syn}} - I_{\text{sharp}}^{\text{syn}} \right\|_2^2 + \lambda_2 \,\text{LPIPS}(\hat{I}_{\text{sharp}}, I_{\text{sharp}}) \\ + \lambda_3 \,\mathcal{L}_{fidelity} + \lambda_4 \,\mathcal{L}_{blur}.\end{aligned} \tag{6}$$

The reconstruction losses, which include the L2 norm ($\| \cdot \|_2^2$) and LPIPS Zhang et al. (2018), provide foundational supervision for structural consistency and training stability. Building upon this foundation, the losses from the ReSyDS strategy then drive the model towards higher fidelity and realism.

3.4 Motion-Strength Modeling

To address the complex, spatially-variant nature of real-world blur, the DS-DeBlur model incorporates our Motion-Strength Modeling (MSM) module. Operating in the latent space, MSM decouples the complex deblurring task into two distinct sub-problems: spatial alignment to correct motion distortion, and strength modulation to apply adaptive restoration. This two-stage design enables a more targeted and effective deblurring process, as illustrated in Fig. 4.

Motion Alignment for Spatial Correction. Motion blur arises from directional displacements during exposure, causing spatial distortions. We estimate and correct these using two lightweight predictors: a motion flow estimator $E_f(\cdot) \in \mathbb{R}^{2\times H\times W}$ and a fusion mask estimator $E_m(\cdot) \in \mathbb{R}^{1\times H\times W}$. The aligned latent z'_{blur} is obtained via bidirectional warping Jaderberg et al. (2015):

$$z'_{\text{blur}} = \mathcal{W}(z_{\text{blur}}, o) \odot m + \mathcal{W}(z_{\text{blur}}, -o) \odot (1 - m), \tag{7}$$

where $\mathcal{W}$ is the warping function, $o = E_f(z_{\text{blur}})$ and $m = E_m(z_{\text{blur}})$. This corrects spatial distortions, providing a clean foundation for the subsequent restoration.

Strength Modulation for Adaptive Restoration. Once aligned, the restoration strength is modulated based on the blur severity of the input. To achieve this, we adopt dynamic convolution Chen et al. (2020), where a specialized convolution kernel K_{dyn} is predicted based on the content of the input features. This is accomplished by using a kernel weighting predictor $E_w(\cdot)$, which takes the aligned features z'_{blur} as input and generates softmax-normalized weights $\{w_i\}$ to linearly combine a set of basis kernels:

$$K_{\text{dyn}} = \sum_{i=1}^{N} w_i \cdot K_i, \quad \{w_i\} = \text{Softmax}(E_w(z'_{\text{blur}})). \tag{8}$$

Here, $\{K_i\}$ represents a set of learnable basis kernels. The final refined latent is then produced by convolving the aligned features with the predicted dynamic kernel:

$$z_{\text{refined}} = z'_{\text{blur}} * K_{\text{dyn}}. \tag{9}$$

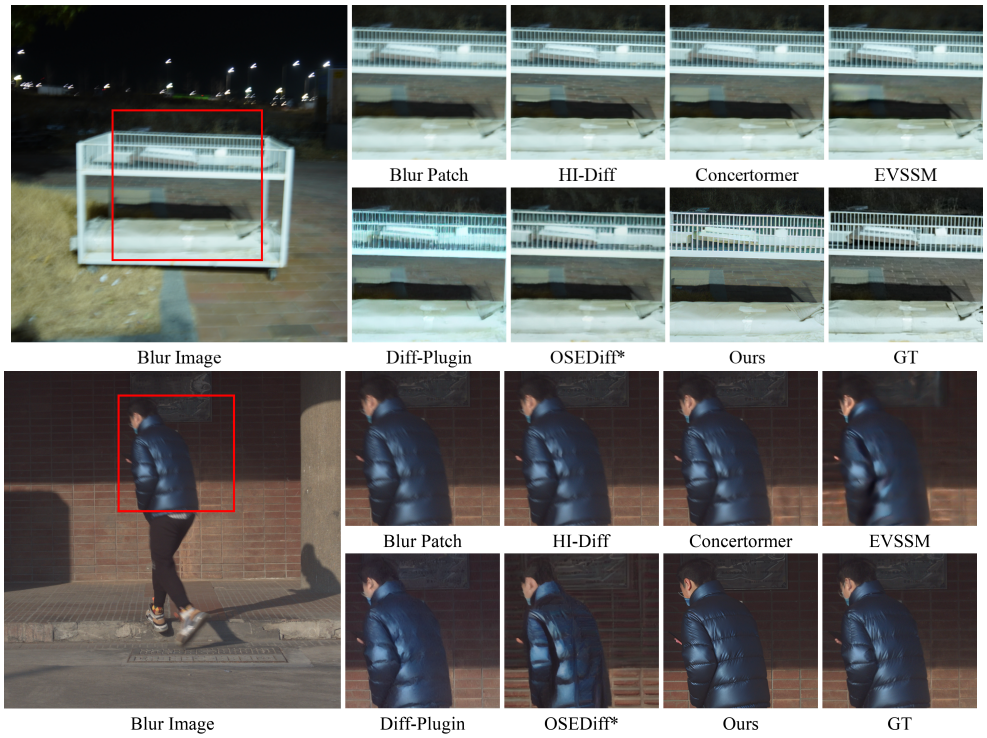

Figure 5: Qualitative comparison of different deblurring methods. **Zoom in for details.**

This enables adaptive restoration, with stronger sharpening for blurrier images and detail preservation for sharper ones.

4 Experiments

4.1 Experimental Settings

Datasets and Metrics. The real path uses GoPro Nah et al. (2017b) training data; the synthetic path uses InstaBlur on clean images. We evaluate on GoPro, RealBlur Rim et al. (2020), REDS Nah et al. (2019), ReLoBlur Li et al. (2023), RWBI Zhang et al. (2020), and FoundIR Li et al. (2025). Paired benchmarks use PSNR, SSIM, and LPIPS Zhang et al. (2018); all benchmarks use CLIP-IQA Wang et al. (2023), NIQE Zhang et al. (2015), MUSIQ Ke et al. (2021), LIQE Zhang et al. (2023), and MANIQA Yang et al. (2022). Data preparation and training details are provided in the supplementary material.

Compared Methods. We compare with DeblurGANv2 Kupyn et al. (2019), FFTFormer Kong et al. (2023), HI-Diff Chen et al. (2023), AdaRevD Xintian Mao & Wang (2024), EVSSM Kong et al. (2025a), Concertormer Kuo et al. (2025), Diff-Plugin Liu et al. (2024), and OSEDiff Wu et al. (2024b). We retrain OSEDiff on GoPro for deblurring (OSEDiff*); all other baselines also use GoPro training data.

4.2 Comparison with State-of-the-Arts

Qualitative Comparison. As illustrated in Fig. 5, we visually compare our proposed approach with several state-of-the-art deblurring methods. When confronted with severe and spatially-varying blurs, existing methods frequently struggle to balance blur removal with detail preservation. As observed in the magnified patches, competing baselines (e.g., HI-Diff Chen et al. (2023), Concertormer Kuo et al. (2025), and EVSSM Kong et al. (2025a)) tend to produce over-smoothed results, failing to recover high-frequency textures. Furthermore, methods like Diff-Plugin Liu et al. (2024) and OSEDiff* Wu et al. (2024b) often introduce noticeable structural distortions or unnatural arti-

Table 1: No-reference quantitative comparison of deblurring methods across six benchmarks. OSEDiff* denotes the retrained OSEDiff on GoPro for deblurring. The best and second-best results of each metric are highlighted in red and blue, respectively.

Datasets	Metrics	DeblurGANv2	FFTformer	HI-Diff	AdaRevD	EVSSM	Concertormer	Diff-Plugin	OSEDiff*	**Ours**
FoundIR	CLIP-IQA↑	0.3336	0.3110	0.3144	0.3080	0.3207	0.3137	0.2630	0.2465	0.5872
	MUSIQ↑	53.374	54.2542	55.614	54.1911	55.3519	55.4099	41.3305	43.5694	70.3436
	NIQE↓	5.3123	5.32	5.4019	5.3702	5.2558	5.3091	4.3725	4.6995	3.4426
	MANIQA↑	0.5501	0.5614	0.5675	0.5658	0.5729	0.5738	0.4326	0.4529	0.6385
	LIQE↑	2.1687	2.1669	2.2463	2.2174	2.2508	2.2427	1.4356	1.3251	4.2585
RealBlur	CLIP-IQA↑	0.2728	0.2149	0.2431	0.2469	0.2221	0.2349	0.2515	0.2276	0.4561
	MUSIQ↑	48.3603	46.5566	50.5487	51.3413	49.4541	49.2257	48.748	42.6545	66.43
	NIQE↓	4.9832	5.2784	5.2351	5.2159	5.1563	5.1185	4.6153	4.7405	3.719
	MANIQA↑	0.5572	0.5397	0.5771	0.5802	0.5535	0.5698	0.5132	0.5103	0.6908
	LIQE↑	1.9087	1.7749	2.0549	2.1739	1.9048	1.9307	1.8969	1.6068	3.6871
ReLoBlur	CLIP-IQA↑	0.3428	0.2436	0.2933	OOM	0.2384	0.2679	0.3374	0.2705	0.6056
	MUSIQ↑	42.0206	31.4301	41.8899	OOM	29.9652	38.3105	43.3617	38.064	64.0227
	NIQE↓	3.9057	3.7683	3.7955	OOM	4.6099	4.1368	3.9702	4.2539	2.9847
	MANIQA↑	0.5316	0.3945	0.5059	OOM	0.3832	0.4772	0.5025	0.4626	0.6085
	LIQE↑	1.5005	1.1419	1.3696	OOM	1.1002	1.2633	1.4568	1.2121	3.2818
GoPro	CLIP-IQA↑	0.2304	0.2631	0.2584	0.2583	0.2633	0.2605	0.2277	0.2123	0.3582
	MUSIQ↑	41.9431	46.0729	45.5391	45.4209	46.3951	45.814	39.0199	42.1592	58.984
	NIQE↓	4.7293	5.0338	5.1862	5.0493	5.0091	5.0329	4.1259	4.2869	3.3715
	MANIQA↑	0.5177	0.5443	0.536	0.5408	0.547	0.5446	0.4706	0.4907	0.6109
	LIQE↑	1.3897	1.5974	1.5473	1.5618	1.615	1.5748	1.2659	1.1797	2.4778
REDS	CLIP-IQA↑	0.3406	0.2425	0.2578	0.2634	0.2523	0.2487	0.3528	0.3323	0.6176
	MUSIQ↑	50.0404	43.0407	47.1816	47.7535	46.982	46.3319	50.5793	48.0371	72.5742
	NIQE↓	4.5637	4.9706	5.3376	5.1185	5.0155	5.228	4.2261	4.1694	3.9608
	MANIQA↑	0.5306	0.5012	0.5081	0.5108	0.4968	0.5035	0.4803	0.4981	0.6614
	LIQE↑	2.1889	1.7358	1.913	1.9945	1.8855	1.8363	2.2785	1.9198	4.5085
RWBI	CLIP-IQA↑	0.2905	0.2641	0.2553	0.2606	0.2775	0.2689	0.2817	0.2493	0.5275
	MUSIQ↑	54.0638	55.7761	54.4542	54.7702	57.7603	55.8511	54.1495	50.0347	70.6149
	NIQE↓	3.8433	3.8621	4.0231	3.9717	3.8012	3.8832	3.2432	3.4446	3.0655
	MANIQA↑	0.6049	0.6184	0.6083	0.6134	0.6317	0.6215	0.5483	0.5625	0.6849
	LIQE↑	2.3485	2.3237	2.2225	2.301	2.5106	2.3236	2.0948	1.8387	4.2221

facts when reconstructing complex geometric patterns, such as the metal grid in the first row. In contrast, our method consistently restores fine details, sharp edges, and authentic textures with high fidelity. Whether reconstructing the rigid, parallel structures of the cart or recovering the intricate folds and lighting variations on the jacket (second row), our approach maintains strict structural consistency and closely approximates the Ground Truth (GT). This demonstrates its strong robustness and superior visual quality against challenging blur degradations. **Additional visual results are provided in the appendix.**

Inference Speed. Table 2 reports the inference speed (sec/image) on an NVIDIA L40 GPU for four methods evaluated on the GoPro Nah et al. (2017b) and RealBlur Rim et al. (2020) datasets. Overall, our method achieves the fastest inference on both benchmarks, taking 0.87 s/image on GoPro and 0.53 s/image on RealBlur. Compared to OSEDiff*, this corresponds to a 14.7% and 26.4% reduction in per-image runtime on GoPro and RealBlur, respectively. Notably, Diff-Plugin, OSEDiff*, and our method are all built upon pre-trained diffusion models; however, our approach is consistently faster, suggesting that the proposed module is lightweight and introduces only minimal computational overhead in practice.

Table 2: Inference speed (seconds per image) on NVIDIA L40.

Dataset	AdaRevD	Diff-Plugin	OSEDiff*	Ours
GoPro	1.96	6.12	1.02	0.87
RealBlur	0.95	2.51	0.72	0.53

Quantitative Evaluation. As summarized in Table 1, our method consistently outperforms competing approaches across No-Reference metrics, demonstrating superior perceptual quality and restoration realism. This advantage is particularly critical for real-world scenarios where ground-truth

images are often unavailable. The full-reference comparison and analysis are provided in the supplementary material.

4.3 Ablation Study

All ablation experiments follow the same training setup as the main experiment and are evaluated on the FoundIR Li et al. (2025).

Table 3: Ablation study of real vs. synthetic supervision.

Variant	Model	Real Data	Synthetic Data	PSNR↑	SSIM↑	LPIPS↓	CLIP-IQA↑	MUSIQ↑	NIQE↓	MANIQA↑	LIQE↑
Real-only	BF-Diff	✓	×	21.70	0.7101	0.2152	0.2823	50.07	4.4378	0.5129	1.7583
Synthetic-only		×	✓	21.19	0.6562	0.2903	0.6489	71.42	3.8679	0.6467	4.5361
ReSyDS		✓	✓	21.51	0.6880	0.2490	0.5872	70.34	3.4426	0.6385	4.2585

Effectiveness of Real vs. Synthetic Supervision. As shown in Table 3, real-only training performs better on full-reference metrics such as PSNR and LPIPS, benefiting from authentic degradations, but noisy GT limits perceptual quality and the use of pretrained diffusion priors. In contrast, synthetic-only training achieves better no-reference metrics such as CLIP-IQA and MANIQA due to cleaner supervision. ReSyDS combines both advantages and achieves the best overall trade-off between structural fidelity and perceptual realism.



Baseline (V1) InstaBlur

Figure 6: Ablation study of InstaBlur.

Table 4: Ablation study of InstaBlur.

Variant	Instance Blur	Noise Simulation	Higher-order Blur	PSNR↑	LPIPS↓	MANIQA↑	LIQE↑
V1	×	×	×	21.14	0.2457	0.6301	4.2245
V2	×	✓	✓	21.48	0.2330	0.6066	3.5557
V3	✓	×	✓	21.31	0.2511	0.6350	4.2185
V4	✓	✓	×	21.43	0.2480	0.6381	4.2515
InstaBlur	✓	✓	✓	21.51	0.2490	0.6385	4.2585

Effectiveness of InstaBlur. Table 4 shows consistent performance drops when removing instance-level blur, noise simulation, or higher-order blur, confirming their complementary contributions. As further shown in Fig. 6, InstaBlur produces better local deblurring than uniform blur, demonstrating the importance of modeling instance-specific motion, complex noise, and diverse degradations for real-world generalization.

Effectiveness of B-Assr and MSM. As shown in Table 5, removing DWT from B-Assr (V1) decreases perceptual metrics such as MANIQA and LIQE, validating the benefit of frequency-domain blur awareness. Removing MSM (V2) degrades PSNR and LPIPS, indicating reduced capability in handling complex motion distortions. The full DS-DeBlur model achieves the highest PSNR, while V2 scores higher on MANIQA and LIQE, reflecting a trade-off among metrics.

Table 5: Ablation of key components.

Variant	B-Assr (w/ DWT)	MSM	PSNR↑	LPIPS↓	MANIQA↑	LIQE↑
V1	×	✓	21.46	0.2319	0.6253	3.8866
V2	✓	×	21.33	0.2545	0.6459	4.3672
DS-DeBlur	✓	✓	21.51	0.249	0.6385	4.2585

5 Conclusion and Limitation

We present DS-DeBlur, a one-step diffusion model for realistic image deblurring. Our ReSyDS paradigm decouples supervision fidelity from blur realism, leveraging clean paired data to preserve structure and pretrained diffusion priors, while using real blurred images to improve perceptual realism and generalization. Together with Instance-Driven Realistic Blur Synthesis, DS-DeBlur achieves state-of-the-art perceptual quality and strong generalization. Its main limitation is the recovery of subtle structures (*e.g.*, small faces and text), which remains a common challenge for pretrained diffusion models Wu et al. (2024b); Yu et al. (2024); Labs (2024); Esser et al. (2024).

AI Use Statement

In preparing this manuscript, generative AI tools were used solely for language editing, grammar proofreading, and LaTeX formatting. All methodologies, theoretical formulations, code implemen-

tations, experimental evaluations, and scientific interpretations were independently conceived, conducted, and verified by the authors.

Ethics Statement

This work adheres to established academic codes of ethics. Our research includes a user study with 20 volunteers; it does not involve demographic profiling or private biometric data collection. All datasets utilized for training and evaluation are publicly accessible and strictly used for academic research under their respective licenses. Care should be taken to prevent the generation of misleading content or malicious misuse of the proposed models.

Reproducibility Statement

All mathematical formulations, algorithmic designs, objective functions, and experimental configurations are comprehensively documented in the main text and supplementary materials. Complete evaluation protocols and metric definitions are provided to ensure maximum transparency. The full source code, trained model checkpoints, and evaluation scripts will be made publicly available upon paper acceptance.

References

Abdullah Abuolaim and Michael S Brown. Defocus deblurring using dual-pixel data. In *ECCV*, 2020.

Stability AI. Sd-turbo. https://huggingface.co/stabilityai/sd-turbo, 2024.

Ayan Chakrabarti. A neural approach to blind motion deblurring. In *ECCV*, 2016.

Yinpeng Chen, Xiyang Dai, Mengchen Liu, Dongdong Chen, Lu Yuan, and Zicheng Liu. Dynamic convolution: Attention over convolution kernels. In *CVPR*, 2020.

Zheng Chen, Yulun Zhang, Ding Liu, Bin Xia, Jinjin Gu, Linghe Kong, and Xin Yuan. Hierarchical integration diffusion model for realistic image deblurring. In *NeurIPS*, 2023.

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, et al. Scaling rectified flow transformers for high-resolution image synthesis. In *ICML*, 2024.

Rob Fergus, Barun Singh, Aaron Hertzmann, Sam T Roweis, and William T Freeman. Removing camera shake from a single photograph. In *Acm Siggraph*, 2006.

Xin Gao, Tianheng Qiu, Xinyu Zhang, Hanlin Bai, Kang Liu, Xuan Huang, Hu Wei, Guoying Zhang, and Huaping Liu. Efficient multi-scale network with learnable discrete wavelet transform for blind motion deblurring. In *CVPR*, 2024.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *NeurIPS*, 2020.

Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. In *NeurIPS*, 2015.

Hai Jiang, Ao Luo, Xiaohong Liu, Songchen Han, and Shuaicheng Liu. Lightendiffusion: Unsupervised low-light image enhancement with latent-retinex diffusion models. In *ECCV*, 2024.

Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. MUSIQ: multi-scale image quality transformer. In *ICCV*, 2021.

Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In *CVPR*, 2023.

Lingshun Kong, Jiangxin Dong, Jinhui Tang, Ming-Hsuan Yang, and Jinshan Pan. Efficient visual state space model for image deblurring. In *CVPR*, 2025a.

Lingshun Kong, Jiawei Zhang, Dongqing Zou, Jimmy Ren, Xiaohe Wu, Jiangxin Dong, and Jinshan Pan. Deblurdiff: Real-world image deblurring with generative diffusion models. In *NeurIPS*, 2025b.

Nupur Kumari, Richard Zhang, Eli Shechtman, and Jun-Yan Zhu. Ensembling off-the-shelf models for gan training. In *CVPR*, 2022.

Pin-Hung Kuo, Jinshan Pan, Shao-Yi Chien, and Ming-Hsuan Yang. Efficient concertormer for image deblurring and beyond. In *ICCV*, 2025.

Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiri Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *CVPR*, 2018.

Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *ICCV*, 2019.

Black Forest Labs. Flux. https://github.com/black-forest-labs/flux, 2024.

Wei-Sheng Lai, YiChang Shih, Lun-Cheng Chu, Xiaotong Wu, Sung-Fang Tsai, Michael Krainin, Deqing Sun, and Chia-Kai Liang. Face deblurring using dual camera fusion on mobile phones. *ACM SIGGRAPH*, 2022.

Dongwoo Lee, Joonkyu Park, and Kyoung Mu Lee. Gs-blur: A 3d scene-based dataset for realistic image deblurring. In *NeurIPS*, 2024.

Anat Levin, Yair Weiss, Fredo Durand, and William T Freeman. Understanding and evaluating blind deconvolution algorithms. In *CVPR*, 2009.

Hao Li, Xiang Chen, Jiangxin Dong, Jinhui Tang, and Jinshan Pan. Foundir: Unleashing million-scale training data to advance foundation models for image restoration. In *ICCV*, 2025.

Haoying Li, Ziran Zhang, Tingting Jiang, Peng Luo, Huajun Feng, and Zhihai Xu. Real-world deep local motion deblurring. In *AAAI*, 2023.

Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Yu Qiao, Wanli Ouyang, and Chao Dong. Diffbir: Toward blind image restoration with generative diffusion prior. In *ECCV*, 2024.

Xiaoyang Liu, Yuquan Wang, Zheng Chen, Jiezhang Cao, He Zhang, Yulun Zhang, and Xiaokang Yang. One-step diffusion model for image motion-deblurring. *arXiv preprint arXiv:2503.06537*, 2025.

Yuhao Liu, Zhanghan Ke, Fang Liu, Nanxuan Zhao, and Rynson W.H. Lau. Diff-plugin: Revitalizing details for diffusion-based low-level tasks. In *CVPR*, 2024.

Boyu Lu, Jun-Cheng Chen, and Rama Chellappa. Unsupervised domain-specific deblurring via disentangled representations. In *CVPR*, 2019.

Armin Mehri, Parichehr B. Ardakani, and Ángel D. Sappa. Mprnet: Multi-path residual network for lightweight image super resolution. In *WACV*, 2021.

Tomer Michaeli and Michal Irani. Blind deblurring using internal patch recurrence. In *ECCV*, 2014.

Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, 2017a.

Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, 2017b.

Seungjun Nah, Sungyong Baik, Seokil Hong, Gyeongsik Moon, Sanghyun Son, Radu Timofte, and Kyoung Mu Lee. Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study. In *CVPR Workshops*, 2019.

Thekke Madam Nimisha, Akash Kumar Singh, and Ambasamudram N Rajagopalan. Blur-invariant deep learning for blind-deblurring. In *ICCV*, 2017.

Thekke Madam Nimisha, Kumar Sunil, and AN Rajagopalan. Unsupervised class-specific deblurring. In *ECCV*, 2018.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.

Jaesung Rim, Haeyun Lee, Jucheol Won, and Sunghyun Cho. Real-world blur dataset for learning and benchmarking deblurring algorithms. In *ECCV*, 2020.

Jaesung Rim, Geonung Kim, Jungeon Kim, Junyong Lee, Seungyong Lee, and Sunghyun Cho. Realistic blur synthesis for learning image deblurring. In *ECCV*, 2022.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, 2022.

Axel Sauer, Dominik Lorenz, Andreas Blattmann, and Robin Rombach. Adversarial diffusion distillation. In *ECCV*, 2024.

Qi Shan, Jiaya Jia, and Aseem Agarwala. High-quality motion deblurring from a single image. *Acm transactions on graphics*, 2008.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *ICLR*, 2021.

Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *CVPR*, 2018.

Fu-Jen Tsai, Yan-Tsung Peng, Yen-Yu Lin, Chung-Chi Tsai, and Chia-Wen Lin. Stripformer: Strip transformer for fast image deblurring. In *ECCV*, 2022.

Jiangang Wang, Qingnan Fan, Qi Zhang, Haigen Liu, Yuhang Yu, Jinwei Chen, and Wenqi Ren. Hero-sr: One-step diffusion for super-resolution with human perception priors. *arXiv preprint arXiv:2412.07152*, 2024.

Jianyi Wang, Kelvin C. K. Chan, and Chen Change Loy. Exploring CLIP for assessing the look and feel of images. In *AAAI*, 2023.

Jia-Hao Wu, Fu-Jen Tsai, Yan-Tsung Peng, Chung-Chi Tsai, Chia-Wen Lin, and Yen-Yu Lin. Id-blau: Image deblurring by implicit diffusion-based reblurring augmentation. In *CVPR*, 2024a.

Rongyuan Wu, Lingchen Sun, Zhiyuan Ma, and Lei Zhang. One-step effective diffusion network for real-world image super-resolution. In *NeurIPS*, 2024b.

Rongyuan Wu, Tao Yang, Lingchen Sun, Zhengqiang Zhang, Shuai Li, and Lei Zhang. Seesr: Towards semantics-aware real-world image super-resolution. In *CVPR*, 2024c.

Yawen Xiang, Heng Zhou, Chengyang Li, Fangwei Sun, Zhongbo Li, and Yongqiang Xie. Deep learning in motion deblurring: current status, benchmarks and future prospects. *The Visual Computer*, 2025.

Qingli Li Xintian Mao and Yan Wang. Adarevd: Adaptive patch exiting reversible decoder pushes the limit of image deblurring. In *CVPR*, 2024.

Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. MANIQA: multi-dimension attention network for no-reference image quality assessment. In *CVPR Workshops*, 2022.

Fanghua Yu, Jinjin Gu, Zheyuan Li, Jinfan Hu, Xiangtao Kong, Xintao Wang, Jingwen He, Yu Qiao, and Chao Dong. Scaling up to excellence: Practicing model scaling for photo-realistic image restoration in the wild. In *CVPR*, 2024.

Aiping Zhang, Zongsheng Yue, Renjing Pei, Wenqi Ren, and Xiaochun Cao. Degradation-guided one-step image super-resolution with diffusion priors. *arXiv preprint arXiv:2409.17058*, 2024.

Kaihao Zhang, Wenhan Luo, Yiran Zhong, Lin Ma, Björn Stenger, Wei Liu, and Hongdong Li. Deblurring by realistic blurring. In *CVPR*, 2020.

Kaihao Zhang, Wenqi Ren, Wenhan Luo, Wei-Sheng Lai, Björn Stenger, Ming-Hsuan Yang, and Hongdong Li. Deep image deblurring: A survey. *IJCV*, 2022.

Lin Zhang, Lei Zhang, and Alan C. Bovik. A feature-enriched completely blind image quality evaluator. *IEEE TIP*, 2015.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018.

Weixia Zhang, Guangtao Zhai, Ying Wei, Xiaokang Yang, and Kede Ma. Blind image quality assessment via vision-language correspondence: A multitask learning perspective. In *CVPR*, 2023.