Generative Manifold Distillation

Aligning Restoration Trajectories with Natural Image Prior

Yuyang Hu[1,2⋆], Mojtaba Sahraee-Ardakan[1], Arpit Bansal[1], Kangfu Mei[1], Christian Qi[1], Peyman Milanfar[1], and Mauricio Delbracio[1]

[1] Google
[2] Washington University in St. Louis
{yuyanghu, mojtabaa, arpitbansal, kangfumei, christianqi, milanfar, mdelbra}@google.com, h.yuyang@wustl.edu

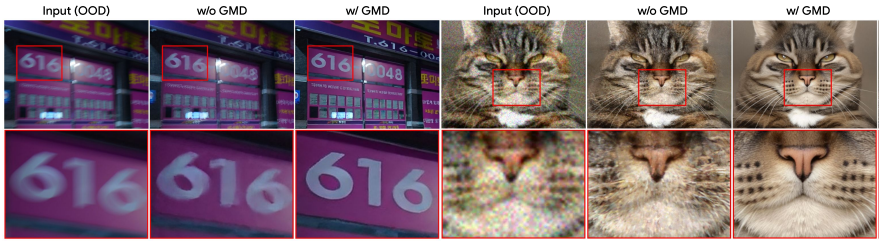


Fig. 1: Bridging the Domain Gap with GMD. (Top) Real-world deblurring and (Bottom) 4x super-resolution examples. While the baseline (LDM) fails on out-of-distribution degradations, GMD recovers high-fidelity details without target-domain clean image or architectural modifications.

Abstract. Pre-trained image restoration models often fail on out-of-distribution (OOD) real-world degradations. Adapting to these domains is challenging as real-world data lacks paired ground truth, and unsupervised methods often require unstable architectural changes. We propose Generative Manifold Distillation (GMD), which reframes domain adaptation as geometric manifold alignment. GMD operates in a strictly unpaired setting, requiring *only low-quality (LQ) target observations*. By leveraging the flow-matching dynamics of a frozen text-to-image foundation model, GMD projects off-manifold restorations onto the natural image manifold to generate high-quality pseudo-targets. To ensure stability, a quality-gated manifold filter rejects off-manifold samples, while source-anchored trajectory regularization prevents error accumulation. Ultimately, GMD distills a powerful generative prior into an efficient restoration network. Experiments demonstrate that GMD seamlessly adapts to new distributions using only LQ inputs, drastically improving perceptual quality with zero architectural modifications or added inference latency.

Keywords: Image Restoration · Unsupervised Domain Adaptation · Generative Model · Manifold Alignment

⋆ This work was done during an internship at Google.

1 Introduction

Image restoration leveraging diffusion models has recently achieved impressive results across tasks like super-resolution [19, 46, 51], deblurring [3, 40, 56], and inpainting [5, 25]. These models benefit from powerful learned generative priors, demonstrating high fidelity under in-distribution settings. However, their performance often collapses when applied to real-world images with complex, unknown out-of-distribution (OOD) degradations [40, 53]. Because acquiring paired ground-truth for these OOD samples is practically impossible, researchers face a fundamental challenge: *how can we adapt an efficient restoration model, pre-trained on synthetic data, to a new unlabeled real-world domain?*

Previous approaches to this problem fall into two distinct paradigms: (1) Traditional Unpaired Domain Adaptation (UDA): Methods relying on CycleGAN-like frameworks [2, 24, 52] require intrusive architectural modifications, such as adding domain discriminators or auxiliary feature extractors. These changes are complex to tune and break the seamless deployment of standard pre-trained models. (2) Zero-Shot Foundation Models: Large text-to-image models (e.g., the 12B parameter FLUX [18]) can act as strong external priors via zero-shot projection techniques (like SDEdit [32] or RF-Solver [49]). While perceptually impressive, applying a massive foundation model at test time introduces additional inference latency. Furthermore, without task-specific finetuning to control them, these strong priors frequently dominate the output, leading to semantic hallucinations and content drift in zero-shot results.

These limitations highlight a clear gap: we need the powerful external generative prior of a large-scale foundation model to overcome severe OOD degradations, but we simultaneously require the strict structural fidelity and blazing-fast inference speed of a standard, unmodified restoration network.

To achieve this, we propose **Generative Manifold Distillation (GMD)**, a framework that reframes domain adaptation as an elegant *offline distillation* task. GMD completely decouples the heavy generative prior from online inference. Instead of running a foundation model at test time, we treat it as an offline T2I generative prior. In **Stage 1**, we perform a deterministic, single-pass automated projection to map unlabeled OOD inputs onto the natural image manifold, creating a dataset of high-quality pseudo-targets. Crucially, a strictly quality-gated manifold filter rejects off-manifold samples. In **Stage 2**, we perform **source-anchored unsupervised distillation**, fine-tuning the efficient baseline model on these filtered targets while rigidly anchoring it to the original synthetic source data to prevent error accumulation.

Contributions.

(1) Zero-Latency Prior Distillation: GMD absorbs the massive external prior of a foundation model entirely offline. This translates the prohibitive computational cost of generative priors into a one-time training setup, yielding significant performance improvement with *zero added inference latency.*

(2) Mitigating Generative Hallucinations: By framing adaptation as a dual-task optimization, the fast student model interpolates the target's natural texture manifold while relying on the source data for strict fidelity. This effectively

avoids the teacher model's hallucinations.
(3) Architecture-Agnostic Unsupervised Adaptation: GMD adapts pre-trained models strictly within the Unpaired Domain Adaptation paradigm. It requires absolutely zero manually labeled target data and zero modifications to the underlying inference architecture.

Table 1: Adaptation Paradigm Comparison. GMD bridges the gap between traditional UDA and zero-shot generative priors, enabling high-quality OOD restoration without architectural changes or high inference latency.

Method	Modifies Base Architecture	Found. Model Prior	Test-Time Found. Model	Inference Latency	Hallucination Risk
Unpaired UDA	✗ (Yes)	✗ (No)	✓ (No)	✓ (Fast)	✗ (High)
Zero-Shot	✓ (No)	✓ (Yes)	✗ (Yes)	✗ (Slow)	✗ (High)
GMD (Ours)	✓ **(No)**	✓ **(Yes)**	✓ **(No)**	✓ **(Fast)**	✓ **(Low)**

2 Background

Image restoration, which aims to recover a high-quality image from its degraded observation, has recently been improved by advances in generative modeling, particularly diffusion-based approaches.

2.1 Image Restoration with Diffusion Models

Diffusion models have become powerful tools for conditional image generation and restoration. Conditional variants [5, 9, 11, 20, 22, 25, 27–30, 40, 45–47, 51, 56, 60] directly learn to sample from the posterior distribution conditioned on the low-quality input. This formulation achieves state-of-the-art performance across diverse restoration tasks, including super-resolution [11, 20, 22, 46, 47, 60], deblurring [40, 56]. However, these models remain sensitive to domain shifts. Trained primarily on synthetic degradations, they often fail to generalize to complex real-world degradations, leading to substantial performance degradation [40].

2.2 Unsupervised Domain Adaptation for Image Restoration

A key challenge in real-world image restoration is lacking of large-scale, paired datasets. Unsupervised domain adaptation aims to bridge this gap by adapting a model trained on a labeled source domain to an unlabeled target domain.

One prominent line of work learns restoration mappings from unpaired data, often via CycleGAN-like adversarial frameworks [2, 7, 14, 15, 39, 63, 67, 68]. Crucially, while these approaches eliminate the need for strictly paired images, they still fundamentally require a curated set of high-quality (HQ) clean images from the target domain to establish the desired output distribution. Furthermore,

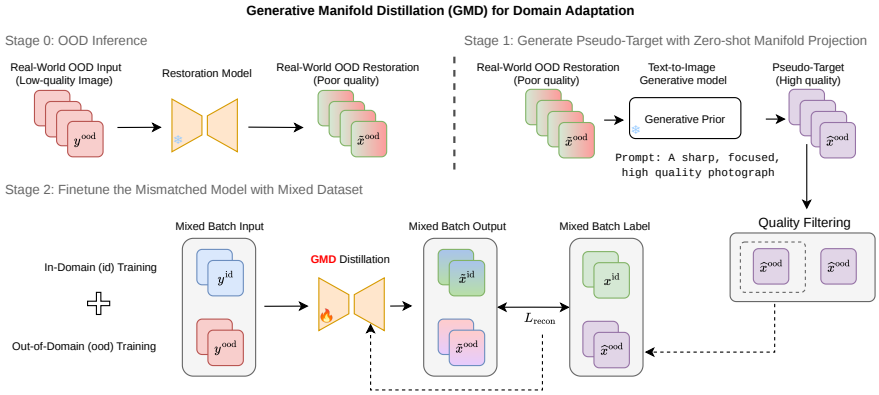


Fig. 2: Overview of the GMD Framework. (a) The Challenge: An ID model fails on out-of-distribution $\mathcal{D}_{\text{ood}}$ data. Zero-shot foundation models can fix this, but suffer from crippling test-time latency and hallucinations. (b) The GMD Solution: A post-training adaptation process. Stage 1 constructs high-quality pseudo-targets *offline* via zero-shot manifold projection. Stage 2 distills the fast student model on a mix of in-distribution anchors and filtered targets, achieving the generative prior with zero inference latency.

establishing these bidirectional mappings suffers from training instability and relies on complex, task-specific architectures [2, 4, 23].

Another strategy explicitly learns a realistic degradation model [57] to synthesize training data, or converts out-of-distribution degradations into in-distribution ones prior to restoration [37, 58]. Alternatively, post-processing methods like DeSRA [61] detect and delete generated artifacts after restoration. However, modeling complex degradations remains challenging, and region-based artifact deletion does not actively adapt the model to the target domain. In contrast, GMD performs a global, offline manifold projection to adapt the model to new OOD domains without test-time post-processing or extra inference latency.

Despite progress, existing UDA methods remain highly task-specific and data-restrictive [2,4,23]. There is a growing need for a general, model-agnostic approach that can adapt pre-trained restoration models using *strictly low-quality (LQ) target observations*, without requiring any target-domain HQ data or architectural modifications.

2.3 Generative Models as Image Priors

Large-scale generative models, such as text-to-image diffusion or rectified flow models [10, 12, 18, 42], offer powerful priors over natural images. Zero-shot restoration methods leverage these priors by mapping degraded inputs to an intermediate noisy state, followed by guided reconstruction. This is typically achieved either via stochastic perturbation (e.g., SDEdit [32]) or deterministic forward ODE inversion [34, 43, 49].

However, deploying these zero-shot strategies is often impractical for high-quality real-world image restoration. They suffer from a severe fidelity-realism trade-off—where the strong, unconstrained prior frequently causes semantic hallucinations and detail loss [11, 38]—and their massive parameter counts introduce prohibitive inference costs that make them computationally slow and expensive [31].

Instead of deploying a massive foundation model online, GMD fundamentally repurposes it as an offline teacher. By treating the foundation model as an offline manifold projection operator, we distill its massive prior into a fast, dedicated restoration network using only unannotated target data. This entirely bypasses the test-time computational bottleneck and actively regularizes against structural hallucinations.

3 Method: Generative Manifold Distillation (GMD)

The fundamental objective of GMD is to adapt an efficient restoration model $f_{\boldsymbol{\theta}}$, pre-trained on a labeled in-distribution (ID) source domain, to a distinct out-of-distribution (OOD) target domain. Crucially, we assume the strictest unsupervised setting for the target domain: we have access *only* to unannotated, low-quality (LQ) degraded observations. Formally, our data regimes are:

- **Source Domain (Paired):** $\mathcal{D}_{\text{id}} = \{(\mathbf{y}_i^{\text{id}}, \mathbf{x}_i^{\text{id}})\}_{i=1}^{N_{\text{id}}}$
- **Target Domain (LQ-Only):** $\mathcal{D}_{\text{ood}} = \{\mathbf{y}_j^{\text{ood}}\}_{j=1}^{N_{\text{ood}}}$

From a representation learning perspective, let $\mathcal{M} \subset \mathbb{R}^d$ represent the low-dimensional manifold of natural, high-fidelity images. A robustly trained ID restorer accurately approximates the projection onto this manifold, mapping $\mathbb{E}_{\mathbb{P}_S}[\mathbf{x}|\mathbf{y}] \in \mathcal{M}$. However, when subjected to real-world OOD data, the model suffers from **Manifold Drift**. Because the unmodeled OOD degradation shifts the input distribution, the restorer's mapping is pushed into undefined regions of the latent space. Consequently, its outputs exhibit high divergence from the target manifold, characterized by residual artifacts and unnatural textures: $D_{KL}(f_{\boldsymbol{\theta}_{\text{id}}}(\mathbb{P}_T(\mathbf{y}))\|\mathcal{M}) > \epsilon$.

To systematically correct this drift relying *only on the LQ inputs* $\mathbf{y}^{ood}$, our framework formulates domain adaptation as a manifold alignment process comprising three phases: Initial Off-Manifold Inference, Generative Prior-Guided Orthogonal Projection, and Trajectory-Regularized Distillation.

3.1 Stage 0: Initial Off-Manifold Inference

For each unlabeled degraded image $\mathbf{y}^{\text{ood}} \in \mathcal{D}_{\text{ood}}$, we first obtain an initial prediction:

$$\tilde{\mathbf{x}}^{\text{ood}} = f_{\boldsymbol{\theta}_{\text{id}}}(\mathbf{y}^{\text{ood}}). \tag{1}$$

Empirically, $\tilde{\mathbf{x}}^{\text{ood}}$ successfully recovers the global structural topology of the scene, but it has been displaced off $\mathcal{M}$. Because the base model lacks the representational capacity for OOD noise profiles, it deposits the image in a low-density "artifact

space" just outside the boundaries of the natural image manifold. This off-manifold estimate serves as the structural anchor for our subsequent projection.

3.2 Stage 1: Generative Prior-Guided Orthogonal Projection

To pull the off-manifold estimate $\tilde{\mathbf{x}}^{\text{ood}}$ back onto $\mathcal{M}$ without altering its semantic identity, we utilize a frozen large-scale text-to-image generative model as a **Prior Projection Operator** $\Pi_{\mathcal{G}} : \mathbb{R}^d \to \mathcal{M}$. Let $p_{\text{data}}(\mathbf{x})$ denote the true distribution of natural images, whose high-density support defines the manifold $\mathcal{M}$. The off-manifold input can be modeled as $\tilde{\mathbf{x}}^{\text{ood}} = \mathbf{x}_{\text{clean}} + \boldsymbol{\epsilon}_{\text{art}}$, where $\boldsymbol{\epsilon}_{\text{art}}$ represents unmodeled OOD artifacts pushing the image into a low-probability region ($p_{\text{data}}(\tilde{\mathbf{x}}^{\text{ood}}) \approx 0$).

(1) Manifold Lifting (Forward ODE): We first lift $\tilde{\mathbf{x}}^{\text{ood}}$ into the generative prior's latent distribution via forward-time Ordinary Differential Equation (ODE) dynamics under null conditioning $\mathbf{c}_\emptyset$:

$$\frac{d\mathbf{x}}{d\tau} = \mathbf{v}(\mathbf{x}(\tau), \tau, \mathbf{c}_\emptyset), \quad \mathbf{x}(0) = \tilde{\mathbf{x}}^{\text{ood}}, \quad \tau \in [0, 1]. \tag{2}$$

In the latent space, this forward mapping acts as an artifact-marginalization step. By integrating the deterministic flow to the high-noise state $\mathbf{z} := \mathbf{x}(1)$, the unstructured artifact penalty $\boldsymbol{\epsilon}_{\text{art}}$ is asymptotically absorbed into the isotropic noise prior. Simultaneously, the low-frequency spatial coordinates of the scene are encoded into the trajectory's self-attention maps.

(2) Orthogonal Projection (Reverse ODE): We then reconstruct a refined, manifold-aligned image $\widehat{\mathbf{x}}^{\text{ood}}$ by solving the guided reverse ODE:

$$\widehat{\mathbf{x}}^{\text{ood}} = \Pi_{\mathcal{G}}(\tilde{\mathbf{x}}^{\text{ood}}) = \int_1^0 \mathbf{v}_{\text{cfg}}(\mathbf{x}(\tau), \tau, \mathbf{c}_{\text{prompt}}) d\tau \tag{3}$$

where $\mathbf{v}_{\text{cfg}}$ represents the classifier-free guided velocity field.

Mathematical Interpretation of the Projection: In generative models, the reverse velocity field $\mathbf{v}_{\text{cfg}}$ inherently points toward high-density regions of the data distribution. Because the degraded input $\tilde{\mathbf{x}}^{\text{ood}}$ contains unmodeled OOD artifacts, it initially lies in a low-density region off the natural image manifold $\mathcal{M}$. Integrating the reverse ODE continuously pulls the trajectory toward the high-density regions, mathematically acting as a projection operator that guarantees the output lands on $\mathcal{M}$.

Preventing Identity Shift via Attention Injection: While the reverse ODE guarantees $\widehat{\mathbf{x}}^{\text{ood}} \in \mathcal{M}$, an unconstrained trajectory can easily project to an arbitrary high-density point, resulting in severe *identity shift* (e.g., hallucinating a completely different valid image on the manifold). To ensure this projection is *orthogonal*—finding the point on $\mathcal{M}$ structurally closest to the input—we employ attention injection [49]. By replacing the self-attention keys and values in the reverse trajectory with those cached from the forward pass, we rigidly lock the spatial layout. This forces the generative flow to shed the off-manifold artifacts while strictly preserving the original semantic identity.

3.3 Stage 1b: Quality-Gated Manifold Filtering

The generative prior is not perfect; severe OOD degradations can cause the projection to miss the target manifold entirely. To ensure the distillation process is driven by high-confidence signals, we admit a pseudo-pair $(\mathbf{y}_i^{\text{ood}}, \widehat{\mathbf{x}}_i^{\text{ood}})$ into the filtered training set $\mathcal{D}_{\text{ood}}^{\text{sel}}$ only if the projection successfully reaches a high-density region of $\mathcal{M}$, measured by a reliability threshold α:

$$\mathcal{D}_{\text{ood}}^{\text{sel}} = \left\{ (\mathbf{y}_i^{\text{ood}}, \widehat{\mathbf{x}}_i^{\text{ood}}) \mid s(\widehat{\mathbf{x}}_i^{\text{ood}}) > \alpha \right\} \tag{4}$$

where an objective quality metric $s(\cdot)$ serves as a proxy for manifold proximity. By rejecting off-manifold samples that fall below threshold α, this quality-gated manifold filter prevents the student restorer from internalizing unnatural hallucinations and erroneous trajectories.

3.4 Trajectory-Regularized Manifold Distillation

In the final phase, we solve for the updated student parameters $\boldsymbol{\theta}$ to internalize the generative prior's orthogonal projection mapping. If we trained $f_{\boldsymbol{\theta}}$ exclusively on the target domain pseudo-pairs $\mathcal{D}_{\text{ood}}^{\text{sel}}$, the model would suffer from manifold collapse—learning to output visually pleasing textures while forgetting the strict deterministic physical mappings required for accurate restoration.

Analysis of the Dual-Objective: We formulate this fine-tuning as a constrained manifold alignment problem:

$$\mathcal{L} = \underbrace{\frac{1}{B_{\text{id}}} \sum_{i=1}^{B_{\text{id}}} \left\| f_{\boldsymbol{\theta}}(\mathbf{y}_i^{\text{id}}) - \mathbf{x}_i^{\text{id}} \right\|^2}_{\substack{\text{Trajectory Regularization} \\ (\mathcal{L}_{\text{id}})}} + \lambda \underbrace{\frac{1}{B_{\text{ood}}} \sum_{j=1}^{B_{\text{ood}}} \left\| f_{\boldsymbol{\theta}}(\mathbf{y}_j^{\text{ood}}) - \widehat{\mathbf{x}}_j^{\text{ood}} \right\|^2}_{\substack{\text{Manifold Alignment} \\ (\mathcal{L}_{\text{ood}})}} \tag{5}$$

The parameter λ balances perceptual alignment with structural fidelity. Geometrically, $\mathcal{L}_{\text{ood}}$ pulls the student's OOD predictions toward the natural image manifold $\mathcal{M}$, while $\mathcal{L}_{\text{id}}$ acts as a structural anchor. By maintaining a high mixing ratio of ID data, we restrict the model to valid physical restorations, preventing error accumulation. This dual objective forces the student to seamlessly interpolate between its physically accurate source knowledge and the target's natural textures, effectively distilling the generative prior into an efficient, single-pass restorer.

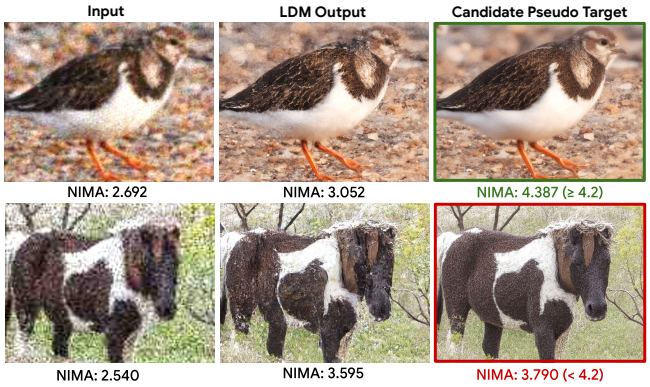


Fig. 3: Selection of projected targets based on image quality assessment. Top row: A successful selection where the generated target scored above the NIMA threshold (≥ 4.0). Bottom row: A rejection case where the projection quality is insufficient.

4 Experiments

In this section, we evaluate GMD on synthetic and real-world deblurring and super-resolution. We test its ability to perform domain adaptation using only unlabeled out-of-distribution data, demonstrating that it adapts a pre-trained model to a new domain with zero test-time overhead.

4.1 Experimental Setup

Unpaired Adaptation Data Requirements. We emphasize that across all experiments, adaptation to the target domain is strictly unsupervised. The models are provided **only with unlabeled, low-quality (LQ) degraded images** from the target domain's training split. At no point does the model have access to target-domain high-quality (HQ) counterparts, paired data.

We evaluate GMD on the adaptation tasks summarized in Table 2. Our base restoration model, $f_{\boldsymbol{\theta}_{id}}$, is a 1.3B parameter Latent Diffusion Model (LDM) based on MMDiT backbone. We test the following adaptation scenarios:

- **Deblurring:** Adapting a model trained on GoPro [36] to REDS [35] and RealBlur-J [41] datasets.
- **Synthetic SR:** Adapting a SR model trained on a *weak* degradation domain (w/ small blur, low noise) to a *strong* domain (larger blur, heavier noise).
- **Real-World SR:** Adapting a SR model trained on synthetic data to the real-world DPED-iPhone dataset [13].

See the supplement for detailed dataset configurations.

Our base restoration model ($f_{\boldsymbol{\theta}_{id}}$) is a 1.3B parameter Latent Diffusion Model (LDM). While 1.3B is large, it remains highly efficient relative to the massive 12B foundation model. It is first pre-trained on its in-distribution dataset for 500K

Table 2: Summary of unsupervised domain adaptation tasks. We evaluate GMD across four scenarios to test generalization against various distribution shifts. Crucially, the target domain utilizes **strictly unpaired, low-quality (LQ) images** during adaptation.

Task	**Source Domain** ($\mathcal{D}_{id}$) *(Paired LQ-HQ Data)*	**Target Domain** ($\mathcal{D}_{ood}$) *(Unpaired LQ Only)*	**Domain Shift**
Deblurring	GoPro [36]	REDS [35]	Video-based motion blur.
Deblurring	GoPro [36]	RealBlur-J [41]	Real-world low-light & blur.
SR	Synthetic-SR$^{(weak)}$ [53]	Synthetic-SR$^{(strong)}$ [53]	Degradation Intensity.
SR	Synthetic-SR [53]	DPED-iPhone [13]	Sensor noise & compression.

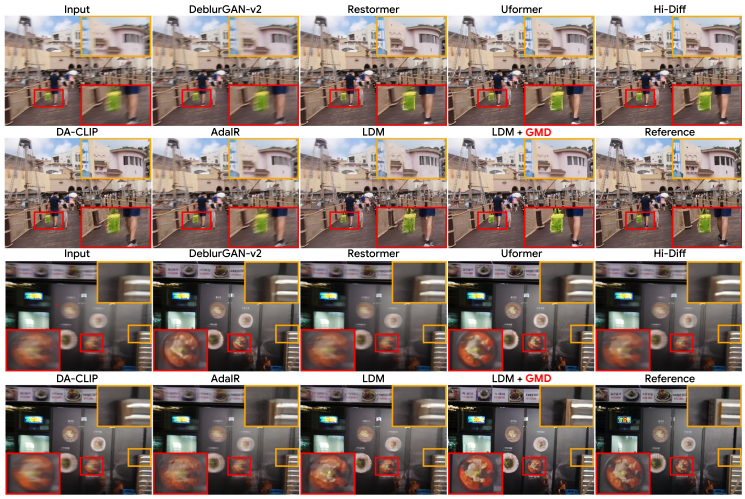


Fig. 4: Qualitative comparison on the REDS (top) and RealBlur-J (bottom) deblurring dataset. The GoPro-trained baseline leaves residual blur. GMD successfully adapts to the out-of-distribution domains, producing perceptually superior restorations.

iterations. The GMD adaptation process then operates in two offline phases. Phase 1 (unsupervised pseudo-target generation) takes approximately ∼3 hours on 4 TPUv5p chips. Phase 2 (knowledge distillation / fine-tuning) takes approximately ∼4 hours on 32 TPUv5p chips. This tractable, one-time offline cost completely eliminates any test-time computational overhead, allowing the baseline model to adapt to new domains with zero added inference latency.

Teacher Model Details: The pseudo-target generation (Sec. 3.2) uses a 12B pre-trained FLUX.1.dev [18] Rectified Flow model as the generative prior. We solve the forward (inversion) and reverse (generation) ODEs with an Euler solver with $N = 50$ steps, classifier-free guidance ($w = 3.5$), and attention injection [49] to improve content preservation.

Baselines: We compare GMD against several models: *(1) SOTA open sourced Methods* (for deblurring: DeblurGAN-v2 [17], Restormer [64], Uformer [54], MPR-

Table 3: Quantitative comparison for real-world deblurring adaptation (GoPro → REDS and RealBlur-J). GMD achieves state-of-the-art performance across all perceptual quality metrics. By vertically stacking the domains, we cleanly present the distortion-perception balance.

Method	Perceptual Quality					Distortion	
	LPIPS↓	NIMA↑	MUSIQ↑	FID↓	CLIPIQA↑	PSNR↑	SSIM↑
Evaluation on REDS Dataset [35]							
DeblurGANv2 [17]	0.190	4.350	53.17	35.77	0.313	27.07	0.805
Restormer [64]	0.220	4.401	53.07	36.42	0.270	26.58	0.801
Uformer [54]	0.197	4.315	54.24	35.11	0.283	27.20	0.825
MPRNet [65]	0.203	4.322	53.64	36.24	0.271	26.87	0.811
Hi-Diff [3]	0.205	4.305	54.04	39.26	0.277	**27.21**	**0.831**
DA-CLIP [26]	0.223	4.294	48.43	42.90	0.258	25.82	0.764
AdaIR [6]	0.285	4.137	40.97	49.67	0.227	25.77	0.774
LDM-Deblur (Baseline)	0.183	4.325	57.60	37.67	0.306	24.08	0.678
LDM-Deblur w/ GMD (Ours)	**0.179**	**4.460**	**63.67**	**31.64**	**0.404**	24.35	0.682
*Fully Supervised**	0.169	4.578	65.04	32.49	0.462	24.51	0.686
Evaluation on RealBlur-J Dataset [41]							
DeblurGANv2 [17]	0.147	4.093	47.82	23.76	0.291	**27.20**	**0.839**
Restormer [64]	0.150	4.060	47.71	23.97	0.245	27.07	0.824
Uformer [54]	0.149	4.148	49.56	23.31	0.256	27.14	0.837
MPRNet [65]	0.149	4.088	47.77	25.05	0.239	26.99	0.833
Hi-Diff [3]	0.145	4.142	50.07	22.28	0.254	27.12	0.831
DA-CLIP [26]	0.239	3.859	38.96	42.26	0.236	20.53	0.680
AdaIR [6]	0.233	3.861	39.44	51.45	0.230	25.92	0.781
LDM-Deblur (Baseline)	0.145	4.081	50.71	25.74	0.214	26.74	0.780
LDM-Deblur w/ GMD (Ours)	**0.132**	**4.480**	**61.33**	**20.33**	**0.354**	26.88	0.796
*Fully Supervised**	0.110	4.487	52.57	17.63	0.293	27.96	0.829

Net [65], Hi-Diff [3], DA-CLIP [26], and AdaIR [6]); *(2) in-distribution baseline*, the base LDM $f_{\boldsymbol{\theta}_{id}}$ trained *only* on in-distribution data, representing the performance *lower bound* without domain adaptation; and *(3) Fully supervised*, the same base LDM fully-supervised trained on the *target* dataset with groundtruth labels, serving as a practical performance upper bound without domain gap.

Evaluation metrics: Performance is evaluated using distortion metrics (PSNR, SSIM [55]) and a suite of perceptual metrics (LPIPS [66], FID [8], and non-reference metrics (NIMA [48], MUSIQ [16], NIQE [44], BRISQUE [33], CLIP-IQA [50], MANIQA [62]).

Human evaluation. Besides the image quality evaluation metrics, we also conducted user studies to validate our method against leading baselines, as shown in Figure 7. To ensure statistical reliability, we used 50 raters for the deblur study and 60 for the super-resolution study. We report win rates with 95% confidence intervals. Further human evaluation details are in the supplement.

4.2 Main Results

Deblurring adaptation: GoPro → REDS / RealBlur-J. Table 3 shows that the baseline model trained on GoPro dataset degrades significantly on REDS and RealBlur-J datasets due to domain mismatch. GMD consistently improves the base model across all perceptual quality metrics and achieves the SOTA performance. While distortion-oriented methods like Hi-Diff [3] and DeblurGAN-v2 [17] offer strong PSNR/SSIM, GMD delivers superior perceptual realism. As shown in Figures 4, GMD outputs are consistently sharper and visually closer to the reference than both the unadapted baseline and other SOTA methods. These visual improvements are further validated by a human preference study (Figure 7), where GMD is overwhelmingly favored over all the leading baselines.
Synthetic SR adaptation: Synthetic-SR$^{(weak)}$ → Synthetic-SR$^{(strong)}$. To evaluate generalization across degradation intensities, we pre-train a 4× SR model on *weak* degradations (RealESRGAN-style [53] on DIV2K [1] with smaller blur kernels and lower noise levels). We then adapt this model to a *strong* degradation domain (more severe blur and noise) using unlabeled, heavily-degraded images from the Flickr2K [21] dataset. The adapted model is then evaluated on the strongly-degraded DIV2K test set, which exhibits a clear domain gap from the pre-trained distribution. As shown in Table 4, the adapted model significantly improves performance, increasing MUSIQ by 3.88, PSNR by 0.29dB, and reducing FID by 1.81. This demonstrates GMD's capacity to generalize across domains without target-domain ground truth.

Table 4: Quantitative Results for SR Adaptation. GMD effectively adapts models to out-of-distribution domains, consistently improving perceptual quality across synthetic (DIV2K) and real-world (DPED) degradations.

4× SR: Weak → Strong (DIV2K)					
Method	**LPIPS↓**	**MUSIQ↑**	**FID↓**	**CLIPIQA↑**	**PSNR/SSIM↑**
LDM-SR w/o GMD	0.386	54.41	30.11	0.442	22.03 / .564
LDM-SR w/ GMD (Ours)	**0.368**	**58.29**	**28.30**	**0.488**	**22.32** / **.578**
*Fully Supervised**	0.233	66.73	15.06	0.602	22.80 / .590
2× SR: DIV2K → DPED-iPhone					
Method	**NIQE↓**	**BRISQUE↓**	**MANIQA↑**	**MUSIQ↑**	**NIMA↑**
DA-CLIP [26]	7.682	25.79	0.403	37.42	3.897
StableSR [51]	4.475	19.77	0.607	58.82	4.426
SeeSR [59]	5.110	19.61	0.619	**60.19**	4.430
LDM-SR (Baseline)	5.467	19.02	0.543	49.30	4.158
LDM-SR w/ GMD (Ours)	**4.300**	**16.35**	**0.627**	59.45	**4.452**

Real-world SR adaptation: Synthetic-SR → DPED-iPhone. Table 4 highlights adaptation from synthetic SR (RealESRGAN-style [53] degradation on DIV2K [1]) to the real-world DPED-iPhone dataset. GMD improves perceptual metrics significantly—e.g., NIQE drops from 5.47 to 4.30, and MANIQA

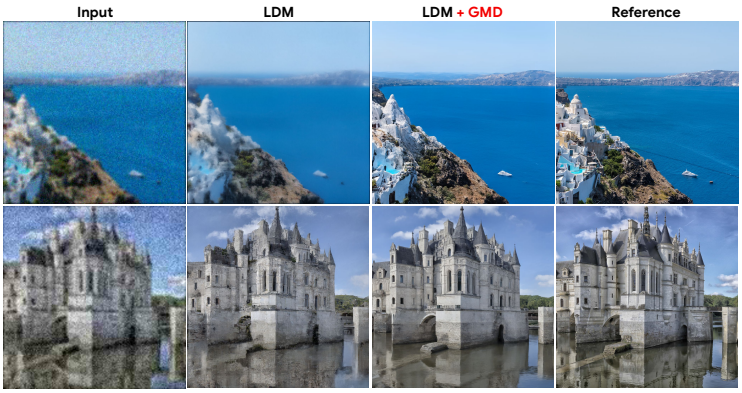


Fig. 5: Visual comparison on Synthetic Super-Resolution (Weak → Strong). The baseline model, pre-trained only on weak degradations, fails to generalize to the heavy noise and blur in the target domain. GMD successfully adapts to the stronger degradation setting, producing clean and sharp restorations.

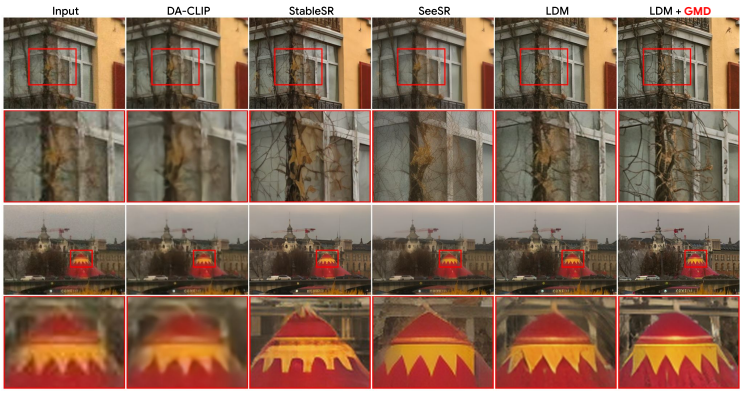
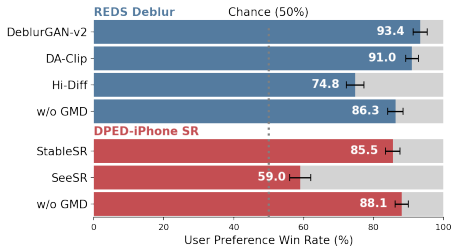


Fig. 6: Qualitative comparison of real-world SR on the DPED-iPhone dataset. Compared to state-of-the-art baselines, GMD produces sharper details and more natural textures under real-world degradations without introducing artifacts.

rises from 0.543 to 0.627—surpassing strong baselines such as DA-CLIP [26], StableSR [51] and SeeSR [59]. These gains confirm that GMD effectively adapts models pretrained on synthetic degradations to real-world low-resolution data, without paired labels. Visual examples in Figure 6 further demonstrate GMD's advantage: compared to prior methods, our adapted outputs exhibit more natural textures, reduced artifacts and has fewer hallucinations while preserving semantic content. Our human preference study (Figure 7) confirms these visual gains, with GMD overwhelmingly outperforming all leading baselines.

Due to the page limit, additional visual results across all datasets are available in the supplement.



"Click on the image that you think is of highest quality (fewer defects, distortions, artifacts, excessive blur, etc.). If both have the same quality, choose the one that appears more natural or pleasing to your eye overall. If both images are equally appealing, click on 'Equally Good/Bad'."

Fig. 7: Human preference study. (Left) Visual comparisons for deblurring on REDS dataset and super resolution on DPED-iPhone dataset. (Right) The evaluation prompt. GMD consistently outperforms baselines in visual quality. Error bars indicate 95% CI.

Table 5: Test-time FLUX projection vs. offline distillation (GMD).

Method	Quality						Compute	
	LPIPS↓	NIMA↑	MUSIQ↑	FID↓	CLIPIQA↑	PSNR↑ / SSIM↑	**Params**	**Latency**
Baseline (GoPro)	0.183	4.325	57.60	37.67	0.306	24.08 / 0.678	1.3B	1.17s
+ SDEdit (online)	0.201	4.445	63.50	40.09	0.404	20.18 / 0.619	1.3B+12B	1.17s+2.40s
+ FLUX.1 Kontext (online)	0.180	4.448	63.60	32.10	0.400	21.50 / 0.630	1.3B+12B	1.17s+6.50s
+ RF-Solver (online)	0.186	4.439	63.47	33.57	0.404	22.15 / 0.643	1.3B+12B	1.17s+4.79s
w/ GMD (offline)	**0.179**	**4.460**	**63.67**	**31.64**	**0.404**	**24.35 / 0.682**	**1.3B**	**1.17s**

4.3 Ablation Studies

We analyze: (1) the importance of quality-gated filtering; (2) the impact of the source-anchoring ratio during fine-tuning; and (3) the role of attention injection for zero-shot manifold alignment.

Table 6: Importance of quality-gated filtering. Filtering targets (Stage 1) is crucial to prevent the student from overfitting to generative artifacts and semantic drift.

Variant	PSNR / SSIM↑	LPIPS↓	MUSIQ↑	FID↓	CLIPIQA↑
w/o filter	23.51 / .652	0.293	50.17	41.26	0.272
w/ filter (Ours)	**24.35 / .682**	**0.179**	**63.67**	**31.64**	**0.404**

Effect of target filtering and Content Drift Mitigation. High-quality pseudo-supervision is essential for successful adaptation. Table 6 examines the impact of the NIMA-based quality gate used in Stage 1. Disabling this mechanism causes a dramatic degradation in performance: LPIPS worsens from 0.179 to 0.293, and FID increases from 31.64 to 41.26. Notably, the model trained without filtering performs substantially worse than even the unadapted baseline (Baseline FID: 37.67). This highlights the necessity of quality gating to prevent the model from overfitting to artifacts in the generated data, ensuring that the adaptation process is driven by reliable supervision.

Effect of data mixing ratio. We study the impact of the mixed-supervision strategy in Table 7. The 1.0 ratio (in-distribution-only) is the baseline. Training

Table 7: Effect of source-anchoring ratio. Ratio of source ID data to OOD target data during fine-tuning. A high ratio (0.9) provides necessary structural regularization.

Ratio	1.0	0.95	0.9 (Ours)	0.6	0.3	0.0
LPIPS↓	0.183	0.188	**0.179**	0.187	0.198	0.217
MUSIQ↑	57.60	58.10	**63.67**	62.15	62.03	52.61
FID↓	37.67	36.75	**31.64**	39.09	39.73	53.56
CLIPIQA↑	0.306	0.320	**0.404**	0.379	0.375	0.287

solely on pseudo-targets (0.0 ratio) performs poorly (FID 53.56), as the model overfits to pseudo-label artifacts and suffers from catastrophic forgetting. The best performance is achieved with a 0.9 ratio (90% in-distribution data). This confirms the necessity of our mixed-supervision strategy: the in-distribution data provides strong regularization, while the small portion of pseudo-targets guides adaptation.

Zero-shot Manifold Alignment. Finally, we assess the visual fidelity of the pseudo-targets generated in Stage 1 (Zero-shot Manifold Alignment). Qualitative comparisons in Figure 8 demonstrate that methods lacking Attention Injection—specifically SDEdit and the "No Attention" pipeline—struggle to maintain structural consistency, often hallucinating new geometries or altering semantic identities. In contrast, RF-Solver leverages attention keys and values from the forward process to guide generation, enforcing strict structural fidelity. The pronounced structural failures of the "No Attention" variant render it unsuitable for supervising a restoration model, as it would inevitably bias the network toward learning geometric distortions.

Generality across teacher and student architectures. We verify GMD's framework generality by evaluating different teacher and student capacities. Specifically, we ablate student model flexibility by adapting a lightweight, non-diffusion CNN model (**SwinIR**) under our framework, and ablate teacher capacity by adapting the model using a 9.2× smaller Stable Diffusion (**SD**) teacher. Please refer to Section 4 in the Supplement for detailed quantitative results, which confirm that GMD generalizes effectively across different restoration backbones and generative prior sources.

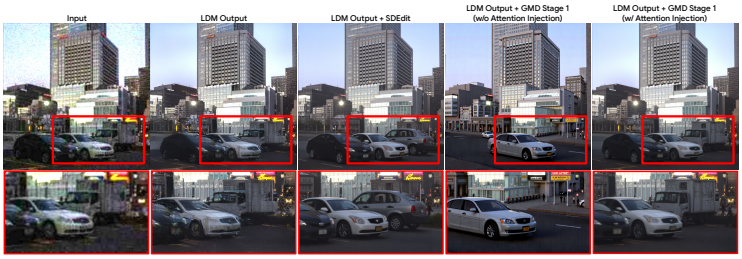


Fig. 8: Visual comparison of pseudo-target generation strategies (Stage 1). We evaluate the intrinsic quality of different generation methods. Standard baselines like SDEdit [32] or the flow matching inversion without Attention Injection help remove artifacts but suffer from severe content drift and structural distortion (e.g., altering vehicle geometry). In contrast, our adopted method with Attention Injection effectively restores details while strictly preserving the original structural layout.

5 Conclusion

We introduced GMD, an intuitive framework that overcomes the limitations of prior paradigms by transforming unsupervised domain adaptation into an offline generative distillation task. Instead of modifying model architectures, relying on unstable adversarial training, or incurring the prohibitive inference latency of zero-shot foundation models, we leverage a frozen T2I generative model offline to create high-quality pseudo-targets. A source-anchored distillation process ensures the student model learns real-world textures while actively smoothing over the teacher's semantic hallucinations. Extensive experiments demonstrate that GMD effectively bridges the domain gap, achieving state-of-the-art perceptual quality and surpassing the generative fidelity ceiling, all while maintaining zero inference-time overhead.

Acknowledgements

The authors would like to thank our colleagues Keren Ye, Viraj Shah and Ashwini Pokle for helpful discussions.

References

1. Agustsson, E., Timofte, R.: Ntire 2017 challenge on single image super-resolution: Dataset and study. In: CVPRW. pp. 1122–1131. IEEE Computer Society (2017)
2. Chen, L., Tian, X., Xiong, S., Lei, Y., Ren, C.: Unsupervised blind image deblurring based on self-enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 25691–25700 (2024)
3. Chen, Z., Zhang, Y., Liu, D., Gu, J., Kong, L., Yuan, X., et al.: Hierarchical integration diffusion model for realistic image deblurring. NIPS **36** (2024)
4. Cheng, J., Chen, W.T., Lu, X., Yang, M.H.: Unpaired deblurring via decoupled diffusion model. arXiv preprint arXiv:2502.01522 (2025)

5. Corneanu, C., Gadde, R., Martinez, A.M.: Latentpaint: Image inpainting in latent space with diffusion models. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 4334–4343 (2024)
6. Cui, Y., Zamir, S.W., Khan, S., Knoll, A., Shah, M., Khan, F.S.: Adair: Adaptive all-in-one image restoration via frequency mining and modulation. In: The Thirteenth International Conference on Learning Representations (2025)
7. Dong, J., Roth, S., Schiele, B.: Learning spatially-variant map models for non-blind image deblurring. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4886–4895 (2021)
8. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. NIPS **30** (2017)
9. Hsiao, C.W., Liu, Y.L., Yang, C.K., Kuo, S.P., Jou, K., Chen, C.P.: Ref-ldm: A latent diffusion model for reference-based face image restoration. NIPS **37**, 74840–74867 (2024)
10. Hu, Y., Delbracio, M., Milanfar, P., Kamilov, U.: A restoration network as an implicit prior. In: International Conference on Learning Representations. vol. 2024, pp. 52376–52400 (2024)
11. Hu, Y., Mei, K., Sahraee-Ardakan, M., Kamilov, U.S., Milanfar, P., Delbracio, M.: Kernel density steering: Inference-time scaling via mode seeking for image restoration. In: The Thirty-ninth Annual Conference on Neural Information Processing Systems (2025)
12. Hu, Y., Peng, A., Gan, W., Milanfar, P., Delbracio, M., Kamilov, U.S.: Stochastic deep restoration priors for imaging inverse problems. Proceedings of Machine Learning Research **267**, 24621–24652 (2025)
13. Ignatov, A., Kobyshev, N., Timofte, R., Vanhoey, K., Van Gool, L.: Dslr-quality photos on mobile devices with deep convolutional networks. In: Proceedings of the IEEE international conference on computer vision. pp. 3277–3285 (2017)
14. Jiang, R., Han, Y.: Uncertainty-aware variate decomposition for self-supervised blind image deblurring. In: Proceedings of the 31st ACM International Conference on Multimedia. pp. 252–260 (2023)
15. Jiangxin, D., ROTH, S., SCHIELE, B.: Learning spatially-variant map models for non-blind image deblurring. In: CVF Conference on Computer Vision and Pattern Recognition, Nashville, USA. pp. 4884–4893 (2021)
16. Ke, J., Wang, Q., Wang, Y., Milanfar, P., Yang, F.: Musiq: Multi-scale image quality transformer. In: ICCV. pp. 5148–5157 (2021)
17. Kupyn, O., Martyniuk, T., Wu, J., Wang, Z.: Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In: ICCV (Oct 2019)
18. Labs, B.F.: Flux. https://github.com/black-forest-labs/flux (2024)
19. Li, H., Yang, Y., Chang, M., Chen, S., Feng, H., Xu, Z., Li, Q., Chen, Y.: Srdiff: Single image super-resolution with diffusion probabilistic models. Neurocomputing **479**, 47–59 (2022)
20. Li, X., Ren, Y., Jin, X., Lan, C., Wang, X., Zeng, W., Wang, X., Chen, Z.: Diffusion models for image restoration and enhancement–a comprehensive survey. ARXIV (2023)
21. Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution. In: CVPRW (2017)
22. Lin, X., He, J., Chen, Z., Lyu, Z., Dai, B., Yu, F., Qiao, Y., Ouyang, W., Dong, C.: Diffbir: Toward blind image restoration with generative diffusion prior. In: ECCV. pp. 430–448. Springer (2024)

23. Liu, C., Qi, L., Pan, J., Qian, X., Yang, M.H.: Learning deblurring texture prior from unpaired data with diffusion model. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 14195–14204 (2025)
24. Lu, B., Chen, J.C., Chellappa, R.: Unsupervised domain-specific deblurring via disentangled representations. In: CVPR. pp. 10225–10234 (2019)
25. Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., Van Gool, L.: Repaint: Inpainting using denoising diffusion probabilistic models. In: CVPR. pp. 11461–11471 (2022)
26. Luo, Z., Gustafsson, F.K., Zhao, Z., Sjölund, J., Schön, T.B.: Controlling vision-language models for multi-task image restoration. In: ICLR (2024), https://openreview.net/forum?id=t3vnnLeajU
27. Mei, K., Delbracio, M., Talebi, H., Tu, Z., Patel, V.M., Milanfar, P.: Codi: Conditional diffusion distillation for higher-fidelity and faster image generation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9048–9058 (2024)
28. Mei, K., Figueroa, L., Lin, Z., Ding, Z., Cohen, S., Patel, V.M.: Latent feature-guided diffusion models for shadow removal. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 4313–4322 (2024)
29. Mei, K., Nair, N.G., Patel, V.M.: Bi-noising diffusion: Towards conditional diffusion models with generative restoration priors. arXiv preprint arXiv:2212.07352 (2022)
30. Mei, K., Talebi, H., Ardakani, M., Patel, V.M., Milanfar, P., Delbracio, M.: The power of context: How multimodality improves image super-resolution. In: CVPR (2025)
31. Mei, K., Tu, Z., Delbracio, M., Talebi, H., Patel, V.M., Milanfar, P.: Bigger is not always better: Scaling properties of latent diffusion models. TMLR (2024)
32. Meng, C., He, Y., Song, Y., Song, J., Wu, J., Zhu, J.Y., Ermon, S.: Sdedit: Guided image synthesis and editing with stochastic differential equations. In: ICLR (2022)
33. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE Transactions on image processing **21**(12), 4695–4708 (2012)
34. Mokady, R., Hertz, A., Aberman, K., Pritch, Y., Cohen-Or, D.: Null-text inversion for editing real images using guided diffusion models. In: CVPR. pp. 6038–6047 (2023)
35. Nah, S., Baik, S., Hong, S., Moon, G., Son, S., Timofte, R., Lee, K.M.: Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study. In: CVPRW (June 2019)
36. Nah, S., Kim, T.H., Lee, K.M.: Deep multi-scale convolutional neural network for dynamic scene deblurring. In: CVPR (July 2017)
37. Pham, B.D., Tran, P., Tran, A., Pham, C., Nguyen, R., Hoai, M.: Blur2blur: Blur conversion for unsupervised image deblurring on unknown domains. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2804–2813 (2024)
38. Qi, C., Tu, Z., Ye, K., Delbracio, M., Milanfar, P., Chen, Q., Talebi, H.: Spire: Semantic prompt-driven image restoration. In: European Conference on Computer Vision. pp. 446–464. Springer (2024)
39. Ren, D., Zhang, K., Wang, Q., Hu, Q., Zuo, W.: Neural blind deconvolution using deep priors. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3341–3350 (2020)
40. Ren, M., Delbracio, M., Talebi, H., Gerig, G., Milanfar, P.: Multiscale structure guided diffusion for image deblurring. In: ICCV. pp. 10721–10733 (2023)

41. Rim, J., Lee, H., Won, J., Cho, S.: Real-world blur dataset for learning and benchmarking deblurring algorithms. In: ECCV. pp. 184–201. Springer (2020)
42. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: CVPR. pp. 10684–10695 (2022)
43. Rout, L., Chen, Y., Ruiz, N., Caramanis, C., Shakkottai, S., Chu, W.S.: Semantic image inversion and editing using rectified stochastic differential equations. In: ICLR (2025), https://openreview.net/forum?id=Hu0FSOSEyS
44. Saad, M.A., Bovik, A.C.: Blind quality assessment of videos using a model of natural scene statistics and motion coherency. In: 2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR). pp. 332–336. IEEE (2012)
45. Saharia, C., Chan, W., Chang, H., Lee, C., Ho, J., Salimans, T., Fleet, D., Norouzi, M.: Palette: Image-to-image diffusion models. In: ACM SIGGRAPH 2022 Conference Proceedings. pp. 1–10 (2022)
46. Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D.J., Norouzi, M.: Image super-resolution via iterative refinement. PAMI **45**(4), 4713–4726 (2022)
47. Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D.J., Norouzi, M.: Image super-resolution via iterative refinement. PAMI **45**(4), 4713–4726 (2022)
48. Talebi, H., Milanfar, P.: Nima: Neural image assessment. TIP **27**(8), 3998–4011 (2018)
49. Wang, J., Pu, J., Qi, Z., Guo, J., Ma, Y., Huang, N., Chen, Y., Li, X., Shan, Y.: Taming rectified flow for inversion and editing. In: ICML (2025), https://openreview.net/forum?id=uDreZphNky
50. Wang, J., Chan, K.C., Loy, C.C.: Exploring clip for assessing the look and feel of images. In: AAAI. pp. 2555–2563 (2023)
51. Wang, J., Yue, Z., Zhou, S., Chan, K.C., Loy, C.C.: Exploiting diffusion prior for real-world image super-resolution. IJCV **132**(12), 5929–5949 (2024)
52. Wang, W., Zhang, H., Yuan, Z., Wang, C.: Unsupervised real-world super-resolution: A domain adaptation perspective. In: ICCV. pp. 4298–4307 (2021). https://doi.org/10.1109/ICCV48922.2021.00428
53. Wang, X., Xie, L., Dong, C., Shan, Y.: Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In: ICCV. pp. 1905–1914 (2021)
54. Wang, Z., Cun, X., Bao, J., Zhou, W., Liu, J., Li, H.: Uformer: A general u-shaped transformer for image restoration. In: CVPR. pp. 17683–17693 (June 2022)
55. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. TIP **13**(4), 600–612 (2004)
56. Whang, J., Delbracio, M., Talebi, H., Saharia, C., Dimakis, A.G., Milanfar, P.: Deblurring via stochastic refinement. In: CVPR. pp. 16293–16303 (2022)
57. Wolf, V., Lugmayr, A., Danelljan, M., Van Gool, L., Timofte, R.: Deflow: Learning complex image degradations from unpaired data with conditional flows. In: CVPR. pp. 94–103 (2021)
58. Wu, J.H., Tsai, F.J., Peng, Y.T., Tsai, C.C., Lin, C.W., Lin, Y.Y.: Id-blau: Image deblurring by implicit diffusion-based reblurring augmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 25847–25856 (2024)
59. Wu, R., Yang, T., Sun, L., Zhang, Z., Li, S., Zhang, L.: Seesr: Towards semantics-aware real-world image super-resolution. In: CVPR. pp. 25456–25467 (2024)
60. 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. pp. 13095–13105 (2023)

61. Xie, L., Wang, X., Chen, X., Li, G., Shan, Y., Zhou, J., Dong, C.: Desra: detect and delete the artifacts of gan-based real-world super-resolution models. In: Proceedings of the 40th International Conference on Machine Learning. pp. 38204–38226 (2023)
62. 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. pp. 1191–1200 (2022)
63. Yi, Z., Zhang, H., Tan, P., Gong, M.: Dualgan: Unsupervised dual learning for image-to-image translation. In: Proceedings of the IEEE international conference on computer vision. pp. 2849–2857 (2017)
64. Zamir, S.W., Arora, A., Khan, S., Hayat, Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: CVPR (2022)
65. Zamir, S.W., Arora, A., Khan, S., Hayat, Khan, F.S., Yang, M.H., Shao, L.: Multi-stage progressive image restoration. In: CVPR (2021)
66. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: CVPR. pp. 586–595 (2018)
67. Zhang, Y., Wang, C., Tao, D.: Neural maximum a posteriori estimation on unpaired data for motion deblurring. IEEE Transactions on Pattern Analysis and Machine Intelligence (2023)
68. Zhao, S., Zhang, Z., Hong, R., Xu, M., Yang, Y., Wang, M.: Fcl-gan: A lightweight and real-time baseline for unsupervised blind image deblurring. In: Proceedings of the 30th ACM International Conference on Multimedia. pp. 6220–6229 (2022)