SP^3: Spherical Priors for Plug-and-Play Restoration

Sean Man
Technion – Israel Institute of Technology
sean.man@campus.technion.ac.il

Ron Raphaeli*
Independent Researcher
ronraphaeli@alumni.technion.ac.il

Matan Kleiner*
Technion – Israel Institute of Technology
matankleiner@campus.technion.ac.il

Or Ronai
Technion – Israel Institute of Technology
or.ronai@campus.technion.ac.il

Abstract

In this paper, we introduce SP^3, a novel Plug-and-Play algorithm that accelerates maximum a posteriori image restoration by replacing denoisers with Spherical Encoders (SE) as generative priors. SP^3 approximates the intractable proximal prior step by utilizing the SE tightly structured latent space as a robust projection onto the natural image manifold. Alternating this projection with a closed-form data-consistency step, via Half-Quadratic Splitting, achieves stable convergence without requiring gradient computation during inference. This unique formulation unlocks "anytime" restoration capabilities, producing sharp, plausible images from the first iteration. Across image restoration tasks, SP^3 often achieves perceptual quality comparable to state-of-the-art zero-shot diffusion and flow methods while being 3-630$\times$ faster. Code is available at the project's webpage.

1 Introduction

Recovering clean images from degraded measurements is a fundamental task in image processing, as photos are often acquired as degraded. A popular technique for solving image restoration is by optimizing the maximum a posteriori (MAP) objective [1]. MAP estimator maximizes the log-likelihood of a data fidelity term and the log-probability of a natural image prior term. After maximizing these two objectives, the resulted image is both natural looking and loyal to the measurements.

Although the data-fidelity term is often straightforward to obtain, specifying an explicit, tractable prior function is challenging. As a result, many approaches have been developed to replace the prior term with easy-to-compute alternatives [1, 2]. One prominent example is the Plug-and-Play (PnP) restoration framework [1], which replaces the explicit prior optimization step within an iterative algorithm with a denoising operator. This framework is particularly attractive because it allows using arbitrary pretrained denoisers. Consequently, better denoisers, such as neural and generative denoisers, can be incorporated with minimal modifications to the underlying reconstruction algorithm [3–6].

With improvements in generative modeling, early methods with a structured latent space, such as generative adversarial network (GAN) [7], variational autoencoders (VAE) [8], and normalizing flows [9], were used as an image prior in restoration tasks [10–16]. These methods often solve MAP optimization problem in the models' latent space, where the prior term is known and easy to sample from, such as a Gaussian. However, they mostly yield poor results [6]. Recent work focuses on utilizing diffusion and flow models [17–23] as generative priors for zero-shot image restoration [24–36]. Yet, their reliance on gradients computed during inference, combined with their iterative nature, significantly increases restoration time.

*denotes equal contribution

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

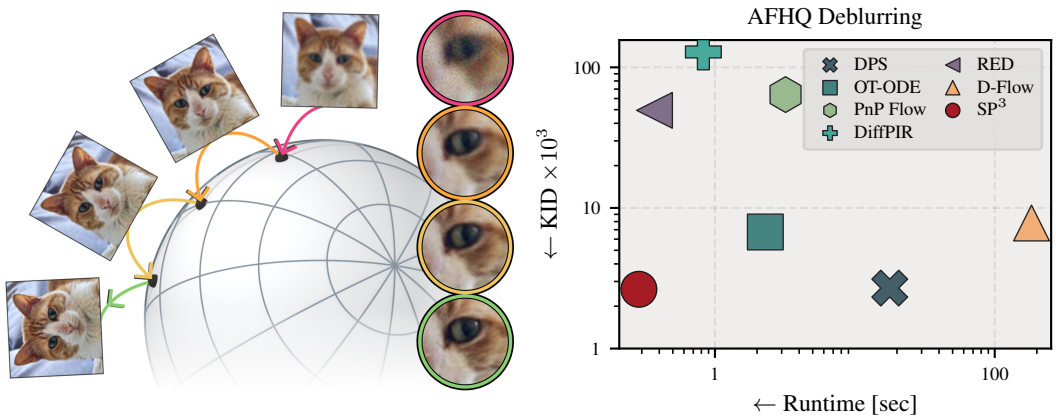


Figure 1: **SP^3**. We introduce a Plug-and-Play image restoration algorithm *(left)* that iteratively projects degraded observations onto a tightly structured latent space using a Spherical Encoder (SE) generative prior. The trajectory along the sphere illustrates our gradient-free, alternating optimization approach, which navigates the image manifold to enforce data consistency. The magnified crops highlight the algorithm's "anytime" generation capability, demonstrating how it stably converges to produce perceptually sharp and plausible details from the very first iteration. *(right)* SP^3 achieves perceptual quality comparable to state-of-the-art zero-shot methods, while being $3 - 630\times$ faster.

In this work, we use the recently proposed Sphere Encoder (SE) [37], a generative model that maps images onto a latent spherical manifold (and vice versa) as a prior for image restoration. Following the original motivation behind Plug-and-Play, enforcing a prior information through a projection operator, we design **Spherical Priors for Plug-and-Play Restoration (SP^3)**. SP^3, schematically illustrated in the left part of Figure 1, alternates, via Half-Quadratic Splitting (HQS), between a prior step and a data fidelity step. The prior step projects the current estimate onto the learned spherical manifold, refines the resulting latent representation, and projects it back to the image manifold, where the data-fidelity step applies a data proximal operator. Importantly, our method stably converges without requiring gradient computation or backpropagation at all.

Unlike diffusion/flow-based methods, SP^3 is extremely fast, achieving $\times 3 - 630$ speed improvement compared to previous work, illustrated in the right part of Figure 1. Moreover, one attractive property of SP^3, emerges from using the SE prior, is anytime restoration – producing a single step reconstruction that could be further refined to improve its perceptual quality (left part of Figure 1). SP^3 achieves these fast and anytime results while producing results on par to different PnP and diffusion/flow based methods, across different noisy restoration problems.

2 Background and related work

2.1 Sphere Encoder

Sphere Encoder (SE) [37] is a recent generative framework that, similar to VAEs [8], relies on an autoencoder [38] to encode images onto a structured latent space and decode latent vectors back to the image space. Unlike VAEs that promote the latent space to follow a Gaussian distribution, SE promotes the latent space to be uniformly distributed on a sphere. By inducing a latent sphere, SE overcomes many of the shortcomings associated with VAEs, especially the posterior hole problem [39–41], which results in unrealistic image samples.

To create the spherical latent space, SE employs a spherifying function (RMS normalization), denoted by $f(\cdot)$, to the output of the encoder E. Namely, a latent vector is given by $v = f(E(x))$. During training, instead of decoding the latent vector v with the decoder D, the SE framework first add noise to the latent vector, followed by spherification, $S(v, \sigma) = f(v + \sigma \cdot e)$, where $e \sim \mathcal{N}(0, I)$ is a white Gaussian noise vector and σ is a scalar that controls the noise magnitude. Adding noise to the latent vectors during training results in a latent sphere that is entirely covered and densely populated.

The training objective of SE compose of three reconstruction loss functions:

$$\mathcal{L}_{\text{SE}} = \mathcal{L}_{\text{L1+perc.}}(D(v_n), x) + \mathcal{L}_{\text{L1+perc.}}(D(v_N), \text{sg}(D(v_n))) + \mathcal{L}_{\text{cos. sim.}}(v, E(D(v_N))), \quad (1)$$

where L1 denotes smoothed L1 loss [42], perc. denotes perceptual loss [43], cos. sim denotes cosine similarity, $\text{sg}(\cdot)$ denotes stop gradient, $v_\epsilon = v + \epsilon$, and $n = \sigma_1 e$, $N = \sigma_2 e$ with $\sigma_2 > \sigma_1$.

In this work, we use the SE framework as a prior for image restoration, and illustrate how its unique features are crucial to the success of SP^3.

2.2 Zero-shot generative priors for image restoration

Image restoration aims to recover a clean natural image, $\hat{x}$, from a degraded measurement y. The degraded measurement is considered to be the result of a degradation operator, A, applied to an unknown clean image x, with the addition of white Gaussian noise $n \sim \mathcal{N}(0, \sigma_n^2 I)$,

$$y = Ax + n. \quad (2)$$

The degradation operator A is linear for many common degradation types, such as inpainting, super-resolution, deblurring, and denoising. Following the notations of Blau & Michaeli [44], the natural image x is a realization of a random variable X with probability density function p_X. The measurement y is a realization of a random variable Y, which relates to X via the conditional probability density function $p_{Y|X}$. In general, an image restoration algorithm is an estimator $\hat{X}$ that generates reconstructions according to $p_{\hat{X}|Y}$, where X and $\hat{X}$ are statistically independent given Y.

Table 1: **SE baseline quantitative comparisons.** Comparisons between SP^3 and baselines that solve MAP optimization problem in SE latent space. Results are given for deblurring on AFHQ-Cat. We highlight the best method in each metric.

	S-GD	S-PGD	SP^3
PSNR [dB] ↑	18.5	23.1	25.7
LPIPS ↓	0.58	0.26	0.12
KID×10^3 ↓	34.3	5.49	2.64
Time [sec]	1.32	1.17	0.28

The success of deep learning in image processing tasks led to a wide adoption of deep networks for image restoration. Early approaches trained regression based networks for image restoration [45–51], while recent approaches trained conditional generative models [52–65]. Despite the impressive results of these methods, they are often constrained by degradation processes seen during training.

An alternative, flexible, approach leverages pretrained models as image priors for zero-shot image restoration. Frameworks such as PnP [1] and RED [2] demonstrate that off-the-shelf deep [3, 4] and generative [5, 66, 6] denoisers can effectively regularize the restoration process. This paradigm has expanded beyond denoisers to incorporate diverse restoration models [67, 68]. Driven by advances in generative modeling, recent frameworks integrates GANs [7], VAEs [8], normalizing flows [9], diffusion [17–20], and flow models [21–23] as powerful image priors.

A common direction to solve a MAP problem in a generative model's latent space [10, 69, 12–14, 16], requires optimizing the latent vector such that applying the model and the degradation operator A, will result in an image similar to the measurement, y. The results of these methods are often poor, as they require propagating gradients through the generative model.

This approach is not easily applied to diffusion/flow models due to their iterative nature and non-trivial latent space. Nevertheless, many methods were developed to apply the image prior learned by diffusion/flow models for zero-shot restoration, often by combining the diffusion model and the acquisition process, via Bayes' rule [24–36]. Their reliance on gradients computed during inference, amalgamated with their iterative nature, results in high-quality restorations at the cost of slow inference.

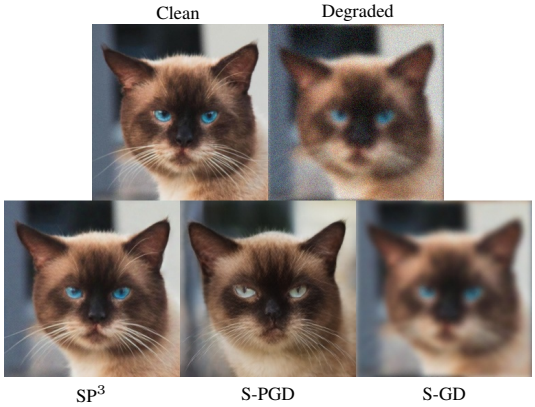


Figure 2: **SE baselines for AFHQ deblurring**. S-GD fails to fully project onto the natural image manifold; S-PGD struggles to remain faithful to the original data. SP^3 balances both constraints, restoring a sharp, high-fidelity image.

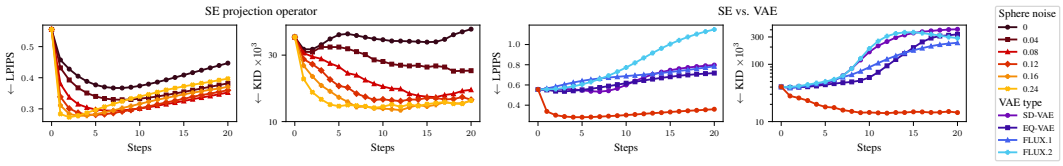


Figure 3: **SE as a projection operator to the clean image manifold.** In both panels, the left plot shows LPIPS, and the right shows KID (on a log scale). Step 0 shows the results of the blurred, noisy input before projections. *(left)* Performance of iteratively applying SE as a projection operator for different values of latent noise. SE projects degraded images to the image manifold, as illustrated by the decrease in the KID score with consecutive steps. *(right)* Different state-of-the-art VAEs are poor projection operators. SE is illustrated with latent noise of 0.12 on the right.

Evaluating image restoration methods requires using perception (whether a reconstructed image looks natural) and distortion (whether a reconstructed image resembles the unknown original image) metrics. It is well established that there is an inherent tradeoff between these metrics [44, 70], therefore, many methods excel at either of these metrics, while sacrificing the other.

In this work, we utilize the recent SE [37] for Plug-and-Play/zero-shot image restoration. In contrast to previous approaches that relied on propagated gradients through the generative model, we use a fast, gradient-free iterative process that takes advantage of SE strong generative prior.

3 Method

To achieve a MAP estimator, we need to obtain a prior over natural images, to guide the restoration process towards a plausible-looking image. Additionally, a measurement-enforcing term is needed to ground the solution in the known, available data.

In this section, we first define the MAP objective within the SE latent space and evaluate two baseline optimization strategies that rely exclusively on the decoder. Next, we analyze the Spherical Encoder's behavior as a projection operator, highlighting its unique stability and manifold-mapping properties in comparison to standard VAEs. Building on these insights, we introduce SP^3: a simple, fast, and effective PnP algorithm. By embedding the SE projection within a Half-Quadratic Splitting (HQS) framework, SP^3 alternates between manifold refinement and proximal data-consistency updates to achieve high-fidelity restoration without gradient backpropagation through the decoder or encoder.

3.1 Naive MAP with a sphere prior

Given a pretrained spherical prior, we can formulate the restoration task as a MAP estimation problem:

$$\arg\max_x \Big\{ p(x|y) \Big\} = \arg\min_x \Big\{ -\log p(y|x) - \log p(x) \Big\}, \tag{3}$$

where the right term results from applying Bayes' theorem and taking the negative logarithm.

Assuming that the decoder D is injective, such that every distinct latent vector v maps to a unique image x, we can reparameterize the problem over the latent space $\mathcal{V}$ rather than the image space $\mathcal{X}$. Furthermore, following the assumption of additive Gaussian noise in the measurements as defined in Equation (2), the objective becomes:

$$\arg\min_v \Big\{ \|AD(v) - y\|_2^2 - \log p(v) \Big\}. \tag{4}$$

This formulation permits us to treat the prior term independently.

We consider two baseline solutions for handling the spherical prior. The first approach strictly enforces the prior by assuming the distribution is uniform over the shell of a sphere $\mathcal{S} = \{v \in \mathcal{V} \mid \|v\|_2^2 = L\}$ and zero elsewhere, $p(v) \propto \delta(\|v\|_2^2 - L)$, allowing us to define $I_{\mathcal{S}}(v)$ as an indicator function that evaluates to 0 if $v \in \mathcal{S}$ and ∞ otherwise. This leads to a constrained optimization problem:

$$\arg\min_v \quad \Big\{ \|AD(v) - y\|_2^2 \Big\} \quad \text{s.t.} \quad I_{\mathcal{S}}(v) = 0, \tag{5}$$

While this is intractable, we approach a local solution using projected gradient descent (S-PGD), where in each iteration the gradient is projected onto the tangent plane to v, and the updated latent vector is retracted back onto the sphere to ensure a valid solution.

An alternative approach uses standard gradient descent (S-GD) with a "soft" enforcement of the prior. We assume a Gaussian distribution with its mean concentrated on the sphere's shell, yielding:

$$\arg\min_{v} \quad \left\{ \|AD(v) - y\|_2^2 + \lambda \left\| \|v\|_2^2 - L \right\|_2^2 \right\}. \tag{6}$$

Notably, in both Equations (5) and (6), the encoder is entirely absent; gradients are propagated only through the decoder. This is an unfortunate omission, as the encoder is an integral part of the prior, having been trained jointly with the decoder to map images into the latent manifold [37]. Our proposed method, SP^3, utilizes both the encoder and the decoder to improve restoration performance, as we discuss in the next subsections.

Shown qualitatively in Figure 2, S-GD fails to recover sharp details, while S-PGD tends to deviate from the measurement constraints. In contrast, SP^3 produces high-fidelity details while remaining faithful to the observed data. Quantitative results in Table 1 further demonstrate that SP^3 outperforms both S-GD and S-PGD across PSNR, LPIPS [71], KID [72], and computational runtime.

3.2 Spherical Encoder as projection operator

As established, formulating the optimization exclusively over the decoder space discards the rich manifold information captured by the encoder. To fully utilize the pretrained spherical prior, we introduce a projection operator that leverages the complete autoencoder architecture.

Denote by $P(\cdot) = D\left(S\left(E\left(\cdot\right), \sigma\right)\right)$ the projection operator onto the image manifold using an SE and latent random noise of strength σ. Such an operator holds several unique properties that make it attractive for image restoration: (1) P projects degraded, out-of-distribution (OOD), images to semantically similar natural images; (2) each application of P produces a probable sample from the image manifold; and (3) successive applications further refine the sample's quality.

We demonstrate those properties on the left side of Figure 3: We take 100 images from the AFHQ-Cat dataset and degrade them using Gaussian blur and additive Gaussian noise. On the degraded inputs, we repeatedly apply P with varying amounts of latent noise. We evaluate the results using LPIPS to measure distortion relative to the clean version of each input and KID to measure perceptual quality relative to the entire dataset.

Empirically, the initial projection validates our first claim: within the first few steps, both LPIPS and KID significantly drop from their starting values across all evaluated noise levels. This indicates a rapid and effective mapping of OOD inputs back toward the natural image distribution.

However, the inclusion of latent noise ($\sigma > 0$) within S is critical to satisfying our second claim. As observed in the KID trajectories, a purely deterministic projection ($\sigma = 0$) slowly diverges, and its perceptual quality remains noticeably worse than configurations with injected latent noise. These higher noise levels effectively stabilize the operator, ensuring it consistently produces more probable, realistic samples from the actual image manifold.

Successive applications of P further validate our third claim. Unconstrained, repeated projections push the sample deeper into the prior manifold, refining global perceptual quality (illustrated by a stable KID score), but inevitably drift from the clean images which created the degraded inputs (illustrated by the increase in LPIPS score).

Crucially, the ability to stably apply P in this manner is not a universal property of autoencoders. On the right side of Figure 3, we compare the SE projection against popular VAEs [73–76]. Instead of refining the sample, successive applications of encoding and decoding using VAEs cause both the LPIPS distortion and the KID perceptual metric to increase dramatically within just a few steps, effectively pushing the sample far from the natural image manifold. This stark contrast highlights the uniqueness of the SE: its tightly structured, bounded latent space, prevents such catastrophic divergence, uniquely enabling the stable, iterative behavior necessary in our case.

This eventual rise in distortion during unconstrained projection motivates our proposed method. We embed P within an alternating PnP framework by interleaving the unconstrained manifold projection

Algorithm 1: SP^3: Spherical Priors for Plug-and-Play Restoration

Input: y; A; E; D; S; prox_λ; K; σ
Output: restored image x_K

$x_0 \leftarrow$ initialization from y;
for $k \leftarrow 1$ **to** K **do**
  $v_k \leftarrow E(x_{k-1})$; `// Encode`
  $v_k \leftarrow S(v_k, \sigma)$; `// Noisy spherify`
  $x_{\text{prior}} \leftarrow D(v_k)$; `// Decode`
  $x_k \leftarrow \text{prox}_\lambda(x_{\text{prior}})$; `// Data proximal`
return x_K;

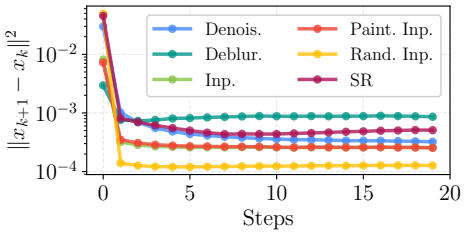


Figure 4: **SP^3 empirical convergence.** MSE of consecutive iterations converges to zero. AFHQ-Cat, y-axis is given in log-scale.

P with a rigorous data-consistency step. This way, we harness P effectiveness without sacrificing data fidelity. Our method, SP^3, effectively balances perceptual realism with adherence to the physical measurement constraints defined in Equation (2).

3.3 SP^3

In previous sections, we showed that the naive baseline solutions, S-PGD and S-GD, yield suboptimal results. Using only the decoder, the methods fail to exploit the spherical prior's learned representations. Disregarding the encoder during the restoration process effectively discards a crucial component.

To address this, we approach the MAP estimation problem from an alternative perspective. Utilizing $I_{\mathcal{S}}(v)$, the strictly constrained optimization problem in Equation (5) can be equivalently rewritten as an unconstrained objective:

$$\arg\min_v \Big\{ \|AD(v) - y\|_2^2 + I_{\mathcal{S}}(v) \Big\}. \tag{7}$$

To decouple the forward measurement model from the non-linear generative prior, we introduce an auxiliary image-space variable x and enforce the constraint $x = D(v)$. Therefore:

$$\arg\min_{v,x} \Big\{ \|Ax - y\|_2^2 + I_{\mathcal{S}}(v) \Big\} \quad \text{s.t.} \quad x = D(v). \tag{8}$$

Following the Half-Quadratic Splitting (HQS) method, we relax this exact constraint by introducing a quadratic penalty term parameterized by $\lambda > 0$. This yields the unconstrained penalty function:

$$\mathcal{L}_\lambda(x, v) = \|Ax - y\|_2^2 + I_{\mathcal{S}}(v) + \lambda \|x - D(v)\|_2^2. \tag{9}$$

We minimize $\mathcal{L}_\lambda(x, v)$ by alternating between updating the latent representation v and the image variable x. Given a fixed intermediate state x_{k-1} from the previous iteration, updating the latent representation requires solving the following subproblem:

$$v_k = \arg\min_v \Big\{ I_{\mathcal{S}}(v) + \lambda \|x_{k-1} - D(v)\|_2^2 \Big\}. \tag{10}$$

Solving Equation (10) exactly is intractable. The objective combines a non-smooth indicator function $I_S(v)$ with a non-linear decoder D, which precludes standard gradient-based optimization. Instead of solving this directly, we approximate the minimizer by accounting for the SE training objectives. During training, the loss in Equation (1) encourages the encoder to behave such that $x \approx P(x)$, while forcing the latent to be on the learned manifold. Therefore, we bypass the optimization in Equation (10) and replace it directly with our projection operator $P(\cdot)$ introduced in Section 3.2.

By applying this operator, the encoder-decoder pair acts as an efficient PnP projector that maps the unregularized state x_{k-1} to a prior-consistent state:

$$x_{\text{prior}} = P(x_{k-1}). \tag{11}$$

Given this projection x_{prior}, the update for x becomes a standard regularized inverse problem:

$$x_k = \arg\min_x \Big\{ \|Ax - y\|_2^2 + \lambda \|x - x_{\text{prior}}\|_2^2 \Big\}. \tag{12}$$

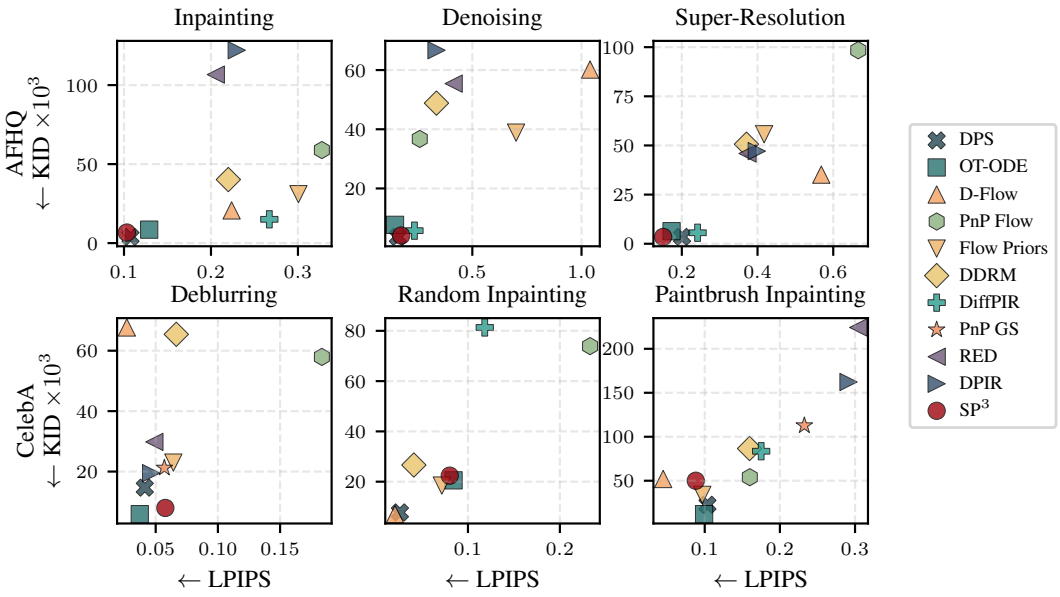


Figure 5: **Perception-distortion evaluation.** Comparison between SP^3 (red sphere) and competing methods on perception (KID) and distortion (LPIPS) metrics on AFHQ (top) and CelebA (bottom). SP^3 often achieves at least comparable perception-distortion results while being significantly faster.

For linear A, the quadratic step in Equation (12) has a closed-form proximal solution:

$$x_k = \text{prox}_\lambda(x_{\text{prior}}) = (A^\top A + \lambda I)^{-1}(A^\top y + \lambda x_{\text{prior}}). \tag{13}$$

Thus, each iteration of our method alternates between ensuring data consistency with the measurements and explicitly projecting the result onto the learned image manifold. This formulation gives rise to our simple and efficient restoration algorithm, summarized in Algorithm 1. Note that during the algorithm, differentiation is not required, thereby significantly reducing computational time.

To conclude this section, we discuss the convergence behavior of SP^3. By construction, our alternating update scheme is an instance of HQS. In classical optimization, if the projection operator $P(x)$ perfectly solves the prior proximal subproblem, HQS is guaranteed to converge to a stationary point of the MAP objective, even in non-convex settings [77, 78].

In the context of PnP methods, theoretically guaranteeing this convergence requires treating the update step as a fixed-point iteration, $x_{k+1} = T(x_k) = \text{prox}_\lambda(P(x_k))$ [79]. Formal convergence to a fixed point $x_\infty \in \text{Fix}(T)$ requires strong assumptions regarding the contractive or averaged nature of the projection operator $P(x)$ (see Appendix D for an extended fixed-point theorem and proofs under these strong assumptions). Because guaranteeing these properties for highly non-linear generative priors is restrictively strong in practice, we validate the convergence of SP^3 empirically. In Figure 4, the empirical sequence generated by our method rapidly stabilizes. Across all evaluated restoration tasks on the AFHQ-Cat dataset, the residual converges to near-zero within the first few iterations, demonstrating that SP^3 robustly reaches a stable fixed point on the image manifold. We further provide additional empirical evidence that, in neighborhoods and directions relevant to the restoration trajectory, the learned SE projector and the complete SP^3 update are compatible with the theoretical assumptions. See Appendix D for further details and discussion.

4 Experiments

We compare SP^3 with zero-shot image restoration methods. The compared methods are divided into three categories: (1) PnP methods utilizing neural denoisers [2, 4, 79]; (2) PnP methods utilizing generative (diffusion/flow based) denoisers [5, 6]; and (3) diffusion/flow based zero-shot methods [24, 25, 31, 33, 32]. Following the recent PnP-Flow [6], we test the different methods on both AFHQ-Cat [80] at 256px resolution and CelebA [81] at 128px resolution. We use the test sets defined by [6]. All the results describe in this section are for 20 SP^3 steps, unless specifically

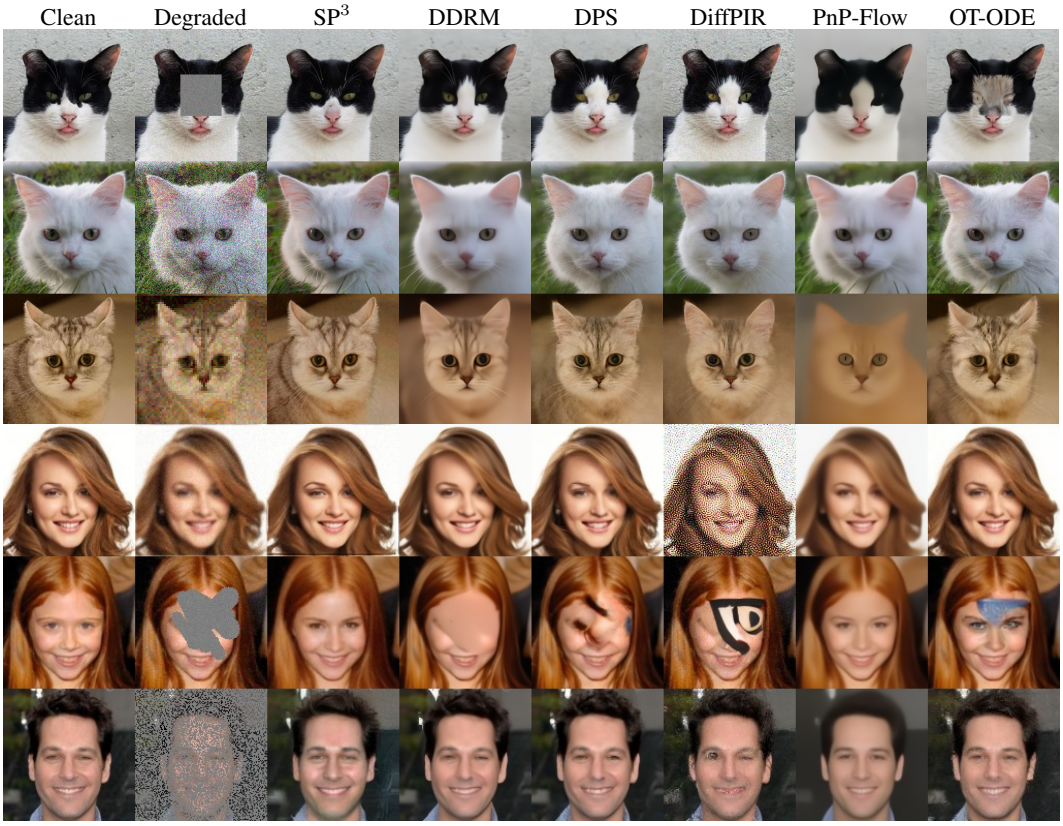


Figure 6: **Qualitative comparison of image restoration.** Results for AFHQ-Cat (top rows) and CelebA (bottom rows). The six rows correspond to the following degradation in descending order: inpainting, denoising, super-resolution, deblurring, paintbrush-inpainting, and random inpainting.

mentioned otherwise. Below we provide details about the experimental setting, where additional implementation details are provided in Appendix C.

Priors. For methods utilizing flow models, we use the priors trained by PnP-Flow [6]. For diffusion-based methods, we use publicly available diffusion models [82, 83]. Classical PnP methods use the DRUnet and GS-DRUnet implemented by the DeepInv library [84]. For SE, we use the official code [85] to train an AFHQ-Cat model using the official hyper-parameters described in [37], and a CelebA model using the same settings for 100 epochs. To ensure fair comparison between SE and diffusion/flow-based methods, we validate that the generative models used as the backbone for all methods mostly achieve similar KID results for generation. See Appendix C for additional details.

Degradations. For each dataset, we evaluate all methods across six noisy image restoration tasks, in the form of Equation (2): (1) Gaussian denoising with $\sigma = 0.4$ for AFHQ-Cat and $\sigma = 0.2$ for CelebA; (2) Deblurring using a 61×61 Gaussian kernel with $\sigma_{\text{blur}} = 3.0$ for AFHQ-Cat and $\sigma_{\text{blur}} = 1.0$ for CelebA; (3) Super-resolution using a $4\times$ bicubic downsampling kernel for AFHQ-Cat and $2\times$ for CelebA; (4) Box-inpainting with a centered square mask of size 80px for AFHQ-Cat and 40px for CelebA; (5) Random pixel inpainting with 70% masked pixels; and (6) paintbrush inpainting. For random inpainting, we add Gaussian noise with $\sigma = 0.02$ on AFHQ-Cat and $\sigma = 0.01$ on CelebA; for all other restoration tasks, we use $\sigma = 0.1$ on AFHQ-Cat and $\sigma = 0.05$ on CelebA.

Metrics. We measure both distortion and perception metrics as they are at odds with each other [44, 70]. We use LPIPS [71] as our main distortion metric, and KID [72] as our perception index. In Appendix A we report the complete set of results in table form, including PSNR as an additional distortion metric that favors smooth reconstructions. Moreover, we report the per-image wall-clock runtime of each method on a single NVIDIA L40S GPU.

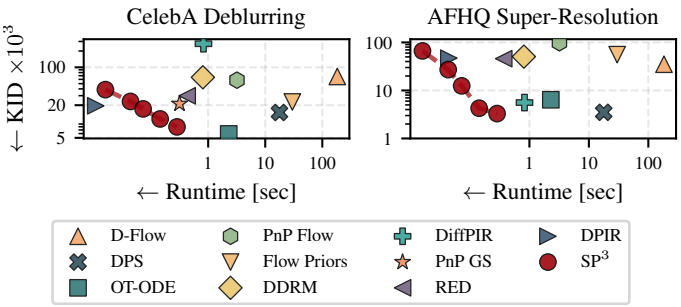


Figure 7: **Perceptual quality vs. running time.** Comparison between SP^3 (red sphere) and competing methods on perception (KID) and running time (sec). SP^3 achieves at least comparable perceptual quality at fraction of the time, achieving 3-630× speedup. SP^3 results are displayed for 1, 3, 5, 10, and 20 steps.

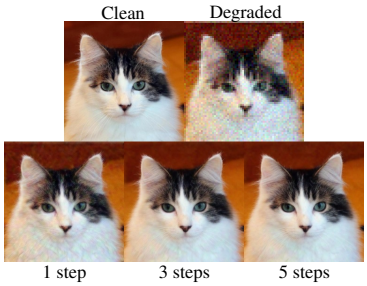


Figure 8: **Anytime results.** Results of anytime restoration for super-resolution on AFHQ-Cat.

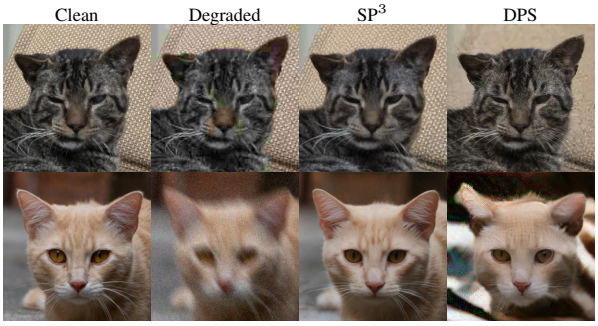


Figure 9: **Nonlinear restoration.** Qualitative comparison between SP^3 and DPS for JPEG artifact removal (top) and nonlinear motion deblurring (bottom) on AFHQ-Cat

Table 2: **Nonlinear degradations.** Quantitative comparison between SP^3 and DPS for JPEG artifact removal and nonlinear motion deblurring on AFHQ-Cat.

	JPEG QF=10		Nonlinear blur	
	SP^3	DPS	SP^3	DPS
PSNR [dB] ↑	27.31	27.22	22.28	20.70
LPIPS ↓	0.097	0.123	0.188	0.254
KID ×10^3 ↓	7.42	3.26	5.15	5.13
Time [sec] ↓	0.7	44.3	8.3	304.9

Comparisons. Figure 5 illustrates quantitative results of SP^3 (red sphere) and the competing methods on a perception-distortion plane [44]. The top part shows results for AFHQ, and the bottom part for CelebA, with each plot showing a different degradation. SP^3 often achieves comparable results to those of the best competing methods, typically DPS [25] and OT-ODE [31]. For a breakdown of SP^3 results compared to competing methods, for both datasets, see Appendix A. Qualitative comparisons illustrated in Figure 6, where SP^3 produces sharp results with fine details that are consistent with the degraded input and loyal to the clean image. See Appendix A for more examples.

Speed & anytime generation. Figures 5 and 6 do not reveal one of the key advantages of SP^3, its speed and anytime capabilities. Because SE projects degraded images onto the clean-image manifold, SP^3 achieves perceptual results comparable to those of competing methods even after a single step. This is illustrated in Figure 7, where we compare the different methods on the perception-runtime plane. Even after a single step (leftmost point), SP^3 is better than several competing methods, and after 20 steps (rightmost point), SP^3 is comparable to the best competing methods while being significantly faster. The speed advantage stems from the small number of steps required, without the need to compute gradients. Concretely, SP^3 using 20 steps is $\times\mathbf{8}$ faster than OT-ODE and $\times\mathbf{60}$ faster than DPS, which are its most direct competitors in terms of visual quality as seen in Figures 5 and 6. Finally, Figure 8 presents an example of the restoration trajectory, demonstrating SP^3 anytime capability. Even after a single step, we obtain a usable, sharp result that can be further refined.

Non-linear degradations. SP^3 is not necessarily restricted to linear degradations, the only requirement is that x_k in Algorithm 1 must satisfy Equation (12). It can accommodate non-linear degradations when an appropriate proximal operator or tractable data-consistency update is available using gradients. To illustrate that, we applied SP^3 to JPEG compression (QF=10) and non-linear blur degradations on the AFHQ-Cat dataset and compared our results with DPS. We initialized the starting state from the degraded image. For JPEG compression, we used a closed-form proximal

operator; for non-linear blur, we used five gradient-descent steps between each prior step. Unlike DPS, the gradient steps are not w.r.t. the prior. Therefore, SP^3 kept its speed advantage, achieving $\times 36$ faster runtime compared to DPS, while being highly competitive across perceptual quality and image fidelity. Qualitative comparisons appear in Figure 9, and quantitative comparisons appear in Table 2. For the implementation details of the non-linear degradations, see Appendix C.

5 Discussion and conclusion

We introduce SP^3, a fast, gradient-free Plug-and-Play algorithm that leverages Spherical Encoders to solve the MAP image restoration problem. By interleaving robust manifold projections with a closed-form data-consistency step, SP^3 achieves anytime generation and perceptual quality which often match state-of-the-art diffusion and flow methods, at a fraction of the computational cost.

We discuss below the role of the SE prior, generalization, robustness, and SP^3's limitations.

Role of the SE prior. Figure 3 illustrates that different from standard autoencoder, SE can be applied stably as projection operator. This unique property makes it a suitable prior for image restoration. To further investigate SE importance within SP^3, we replaced the SE projection with FLUX.1-VAE and EQ-VAE for AFHQ-Cat deblurring. Replacing the SE with either VAE substantially degrades all metrics. These results illustrate that the unique stability and latent-space structure of SE are important for its use as an iterative projection operator. See Appendix B for further details.

Generalization. Given a suitable prior, SP^3 generalizes to additional datasets. We used a SE trained on the entire AFHQ dataset to evaluate SP^3 on the Dog and Wild splits. We compared SP^3 results with DPS, and illustrate that SP^3 generally achieves comparable KID while retaining its benefits of fast, stable, and anytime restoration. We also show that the restoration quality of SP^3 deteriorates when the used prior was trained on limited or out-of-distribution data. This is a similar behavior to other PnP methods that use pretrained generative priors. See Appendix B for details.

Robustness. We show that the stochastic nature of the SP^3 projections does not affect its results, by reporting mean and standard deviation results across ten independent runs with different random seeds for three tasks on each dataset, AFHQ-Cat and CelebA. For all results the standard deviation is at least an order of magnitude smaller than the mean, indicating that SP^3 is robust to the specific random noise realization. Further discussion and details appear in Appendix B.

Limitations. As a MAP estimator and as a direct consequence of the perception-distortion trade-off [44, 70], SP^3 prioritizes perceptual realism over distortion metrics. Although averaging different restoration results over a few seeds can make SP^3 more distortion-oriented (Appendix B), the method still does not provide Minimum Mean Square Error (MMSE) solutions. Additionally, SP^3 is not a posterior sampler like DPS [25]. The current implementation assumes that the degradation operator is known and a tractable data-consistency update is available. Although we demonstrate extensions to selected nonlinear degradations (Section 4), blind restoration with unknown forward operators remains outside the scope of this work. Investigating how the structural latent noise could be actively leveraged to sample from and explore the posterior distribution [86, 87] is a promising direction for future work. Finally, current evaluations are restricted to available pretrained SE priors. Scaling to more diverse and higher-resolution images is an important direction for future work.

Societal impacts. Our method restores clean images from partial measurements, benefiting applications like smartphone photography and noisy imaging. However, it poses dual-use risks, including unauthorized surveillance and deepfake generation.

Acknowledgments and Disclosure of Funding

The authors acknowledge Modal Labs for providing some of the cloud computing platform and GPUs resources ($8\times$ NVIDIA L40S) used for conducting part of the experiments presented in this paper.

References

[1] Singanallur V. Venkatakrishnan, Charles A. Bouman, and Brendt Wohlberg. Plug-and-Play priors for model based reconstruction. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 945–948, December 2013.

[2] Yaniv Romano, Michael Elad, and Peyman Milanfar. The Little Engine That Could: Regularization by Denoising (RED). *SIAM Journal on Imaging Sciences*, 10(4):1804–1844, January 2017.

[3] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep cnn denoiser prior for image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3929–3938, 2017.

[4] Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. Plug-and-Play Image Restoration With Deep Denoiser Prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):6360–6376, October 2022.

[5] Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising Diffusion Models for Plug-and-Play Image Restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1219–1229, 2023.

[6] Ségolène Tiffany Martin, Anne Gagneux, Paul Hagemann, and Gabriele Steidl. Pnp-flow: Plug-and-play image restoration with flow matching. In *The Thirteenth International Conference on Learning Representations*, 2025.

[7] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.

[8] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[9] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real NVP. In *International Conference on Learning Representations*, 2017.

[10] Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G Dimakis. Compressed sensing using generative models. In *International conference on machine learning*, pages 537–546. PMLR, 2017.

[11] Paul Hand, Oscar Leong, and Vlad Voroninski. Phase retrieval under a generative prior. *Advances in Neural Information Processing Systems*, 31, 2018.

[12] Muhammad Asim, Max Daniels, Oscar Leong, Ali Ahmed, and Paul Hand. Invertible generative models for inverse problems: mitigating representation error and dataset bias. In *International conference on machine learning*, pages 399–409. PMLR, 2020.

[13] Sachit Menon, Alexandru Damian, Shijia Hu, Nikhil Ravi, and Cynthia Rudin. Pulse: Self-supervised photo upsampling via latent space exploration of generative models. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*, pages 2437–2445, 2020.

[14] Jay Whang, Qi Lei, and Alex Dimakis. Solving inverse problems with a flow-based noise model. In *International Conference on Machine Learning*, pages 11146–11157. PMLR, 2021.

[15] Xinyi Wei, Hans Van Gorp, Lizeth Gonzalez-Carabarin, Daniel Freedman, Yonina C Eldar, and Ruud JG van Sloun. Deep unfolding with normalizing flow priors for inverse problems. *IEEE Transactions on Signal Processing*, 70:2962–2971, 2022.

[16] Jean Prost, Antoine Houdard, Andrés Almansa, and Nicolas Papadakis. Inverse problem regularization with hierarchical variational autoencoders. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 22894–22905, 2023.

[17] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep Unsupervised Learning using Nonequilibrium Thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning*, pages 2256–2265. PMLR, June 2015.

[18] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models. In *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851. Curran Associates, Inc., 2020.

[19] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising Diffusion Implicit Models. In *International Conference on Learning Representations*, October 2020.

[20] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

[21] Xingchao Liu, Chengyue Gong, and qiang liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*, 2023.

[22] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023.

[23] Michael Albergo, Nicholas M Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *Journal of Machine Learning Research*, 26(209):1–80, 2025.

[24] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising Diffusion Restoration Models. *Advances in Neural Information Processing Systems*, 35:23593–23606, December 2022.

[25] Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion Posterior Sampling for General Noisy Inverse Problems. In *The Eleventh International Conference on Learning Representations*, September 2022.

[26] Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-Shot Image Restoration Using Denoising Diffusion Null-Space Model. (arXiv:2212.00490), December 2022.

[27] Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-Guided Diffusion Models for Inverse Problems. In *International Conference on Learning Representations*, September 2022.

[28] Litu Rout, Negin Raoof, Giannis Daras, Constantine Caramanis, Alexandros G. Dimakis, and Sanjay Shakkottai. Solving Linear Inverse Problems Provably via Posterior Sampling with Latent Diffusion Models, July 2023.

[29] Peiqing Yang, Shangchen Zhou, Qingyi Tao, and Chen Change Loy. PGDiff: Guiding Diffusion Models for Versatile Face Restoration via Partial Guidance. *Advances in Neural Information Processing Systems*, 36:32194–32214, December 2023.

[30] Zongsheng Yue and Chen Change Loy. DifFace: Blind Face Restoration with Diffused Error Contraction. (arXiv:2212.06512), December 2023.

[31] Ashwini Pokle, Matthew J. Muckley, Ricky T. Q. Chen, and Brian Karrer. Training-free linear image inverses via flows. *Transactions on Machine Learning Research*, 2024.

[32] Heli Ben-Hamu, Omri Puny, Itai Gat, Brian Karrer, Uriel Singer, and Yaron Lipman. D-flow: Differentiating through flows for controlled generation. In *International Conference on Machine Learning*, pages 3462–3483. PMLR, 2024.

[33] Yasi Zhang, Peiyu Yu, Yaxuan Zhu, Yingshan Chang, Feng Gao, Ying N Wu, and Oscar Leong. Flow priors for linear inverse problems via iterative corrupted trajectory matching. *Advances in Neural Information Processing Systems*, 37:57389–57417, 2024.

[34] Ron Raphaeli, Sean Man, and Michael Elad. Silo: Solving inverse problems with latent operators. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10570–10580, October 2025.

[35] Sean Man, Guy Ohayon, Ron Raphaeli, Matan Kleiner, and Michael Elad. Elad: Blind face restoration using expectation-based likelihood approximation and diffusion prior. In *Proceedings of the SIGGRAPH Asia 2025 Conference Papers*, pages 1–12, 2025.

[36] Tomer Garber and Tom Tirer. Zero-shot image restoration using few-step guidance of consistency models (and beyond). In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 2398–2407, 2025.

[37] Kaiyu Yue, Menglin Jia, Ji Hou, and Tom Goldstein. Image generation with a sphere encoder. *arXiv preprint arXiv:2602.15030*, 2026.

[38] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507, 2006.

[39] Bin Dai and David Wipf. Diagnosing and enhancing VAE models. In *International Conference on Learning Representations*, 2019.

[40] Danilo Jimenez Rezende and Fabio Viola. Taming vaes. *arXiv preprint arXiv:1810.00597*, 2018.

[41] Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, and Ian Goodfellow. Adversarial autoencoders. In *International Conference on Learning Representations*, 2016.

[42] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.

[43] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2414–2423, 2016.

[44] Yochai Blau and Tomer Michaeli. The Perception-Distortion Tradeoff. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6228–6237, 2018.

[45] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015.

[46] Jian Sun, Wenfei Cao, Zongben Xu, and Jean Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 769–777, 2015.

[47] Richard Zhang, Phillip Isola, and Alexei A Efros. Colorful image colorization. In *European conference on computer vision*, pages 649–666. Springer, 2016.

[48] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, July 2017.

[49] Idan Kligvasser, Tamar Rott Shaham, and Tomer Michaeli. xunit: Learning a spatial activation function for efficient image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2433–2442, 2018.

[50] Javier Gurrola-Ramos, Oscar Dalmau, and Teresa E. Alarcón. A Residual Dense U-Net Neural Network for Image Denoising. *IEEE Access*, 9:31742–31754, 2021.

[51] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. SwinIR: Image Restoration Using Swin Transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1833–1844, 2021.

[52] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.

[53] Andreas Lugmayr, Martin Danelljan, Luc Van Gool, and Radu Timofte. SRFlow: Learning the Super-Resolution Space with Normalizing Flow. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*, pages 715–732, Cham, 2020. Springer International Publishing.

[54] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-ESRGAN: Training Real-World Blind Super-Resolution With Pure Synthetic Data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1905–1914, 2021.

[55] Guy Ohayon, Theo Adrai, Gregory Vaksman, Michael Elad, and Peyman Milanfar. High Perceptual Quality Image Denoising With a Posterior Sampling CGAN. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1805–1813, 2021.

[56] Jay Whang, Mauricio Delbracio, Hossein Talebi, Chitwan Saharia, Alexandros G. Dimakis, and Peyman Milanfar. Deblurring via Stochastic Refinement. pages 16293–16303, 2022.

[57] Sean Man, Guy Ohayon, Theo Adrai, and Michael Elad. High-Perceptual Quality JPEG Decoding via Posterior Sampling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1272–1282, 2023.

[58] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-Image Diffusion Models. In *ACM SIGGRAPH 2022 Conference Proceedings*, SIGGRAPH '22, pages 1–10, New York, NY, USA, July 2022. Association for Computing Machinery.

[59] Yuchao Gu, Xintao Wang, Liangbin Xie, Chao Dong, Gen Li, Ying Shan, and Ming-Ming Cheng. VQFR: Blind Face Restoration with Vector-Quantized Dictionary and Parallel Decoder. In Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner, editors, *Computer Vision – ECCV 2022*, pages 126–143, Cham, 2022. Springer Nature Switzerland.

[60] Shangchen Zhou, Kelvin Chan, Chongyi Li, and Chen Change Loy. Towards Robust Blind Face Restoration with Codebook Lookup Transformer. *Advances in Neural Information Processing Systems*, 35:30599–30611, December 2022.

[61] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11461–11471, 2022.

[62] Mauricio Delbracio and Peyman Milanfar. Inversion by direct iteration: An alternative to denoising diffusion for image restoration. *Transactions on Machine Learning Research*, 2023. Featured Certification, Outstanding Certification.

[63] Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Wanli Ouyang, Yu Qiao, and Chao Dong. DiffBIR: Towards Blind Image Restoration with Generative Diffusion Prior. (arXiv:2308.15070), April 2024.

[64] Guy Ohayon, Tomer Michaeli, and Michael Elad. Posterior-Mean Rectified Flow: Towards Minimum MSE Photo-Realistic Image Restoration. (arXiv:2410.00418), October 2024.

[65] Noam Elata, Hyungjin Chung, Jong Chul Ye, Tomer Michaeli, and Miki Elad. Invfusion: Bridging supervised and zero-shot diffusion for inverse problems. In *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pages 35823–35850. Curran Associates, Inc., 2025.

[66] Morteza Mardani, Jiaming Song, Jan Kautz, and Arash Vahdat. A Variational Perspective on Solving Inverse Problems with Diffusion Models. In *The Twelfth International Conference on Learning Representations*, October 2023.

[67] Yuyang Hu, Mauricio Delbracio, Peyman Milanfar, and Ulugbek Kamilov. A restoration network as an implicit prior. In *The Twelfth International Conference on Learning Representations*, 2024.

[68] Matthieu Terris, Ulugbek S Kamilov, and Thomas Moreau. Fire: Fixed-points of restoration priors for solving inverse problems. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 23185–23194, 2025.

[69] Xingang Pan, Xiaohang Zhan, Bo Dai, Dahua Lin, Chen Change Loy, and Ping Luo. Exploiting deep generative prior for versatile image restoration and manipulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):7474–7489, 2021.

[70] Dror Freirich, Tomer Michaeli, and Ron Meir. A Theory of the Distortion-Perception Tradeoff in Wasserstein Space. In *Advances in Neural Information Processing Systems*, volume 34, pages 25661–25672. Curran Associates, Inc., 2021.

[71] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–595, 2018.

[72] Mikołaj Bińkowski, Danica J. Sutherland, Michael Arbel, and Arthur Gretton. Demystifying MMD GANs. In *International Conference on Learning Representations*, February 2018.

[73] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-Resolution Image Synthesis With Latent Diffusion Models. pages 10684–10695, 2022.

[74] Theodoros Kouzelis, Ioannis Kakogeorgiou, Spyros Gidaris, and Nikos Komodakis. Eq-vae: Equivariance regularized latent space for improved generative image modeling. *arXiv preprint arXiv:2502.09509*, 2025.

[75] Black Forest Labs. Flux. 1 kontext: Flow matching for in-context image generation and editing in latent space. *arXiv preprint arXiv:2506.15742*, 2025.

[76] Black Forest Labs. FLUX.2: Analyzing and enhancing the latent space of FLUX – representation comparison, 2025.

[77] Jérôme Bolte, Shoham Sabach, and Marc Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Math. Program.*, 146(1–2):459–494, August 2014.

[78] Mila Nikolova and Michael K. Ng. Analysis of half-quadratic minimization methods for signal and image recovery. *SIAM Journal on Scientific Computing*, 27(3):937–966, 2005.

[79] Samuel Hurault, Arthur Leclaire, and Nicolas Papadakis. Gradient step denoiser for convergent plug-and-play. In *International Conference on Learning Representations*, 2022.

[80] Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. Stargan v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8188–8197, 2020.

[81] Shuo Yang, Ping Luo, Chen-Change Loy, and Xiaoou Tang. From facial parts responses to face detection: A deep learning approach. In *Proceedings of the IEEE international conference on computer vision*, pages 3676–3684, 2015.

[82] Chen Henry Wu and Fernando De la Torre. A latent space of stochastic diffusion models for zero-shot image editing and guidance. In *ICCV*, 2023.

[83] rdruce. ddpm-celeb-128. https://huggingface.co/rdruce/ddpm-celeb-128. Accessed: 2026-05-07.

[84] Julián Tachella, Matthieu Terris, Samuel Hurault, Andrew Wang, Leo Davy, Jérémy Scanvic, Victor Sechaud, Romain Vo, Thomas Moreau, Thomas Davies, Dongdong Chen, Nils Laurent, Brayan Monroy, Jonathan Dong, Zhiyuan Hu, Minh-Hai Nguyen, Florian Sarron, Pierre Weiss, Paul Escande, Mathurin Massias, Thibaut Modrzyk, Brett Levac, Tobías I. Liaudat, Maxime Song, Johannes Hertrich, Sebastian Neumayer, and Georg Schramm. Deepinverse: A python package for solving imaging inverse problems with deep learning. *Journal of Open Source Software*, 10(115):8923, 2025.

[85] Meta. Sphere encoder in pytorch. https://github.com/facebookresearch/sphere-encoder.

[86] Noa Cohen, Hila Manor, Yuval Bahat, and Tomer Michaeli. From Posterior Sampling to Meaningful Diversity in Image Restoration. October 2023.

[87] Elias Nehme and Tomer Michaeli. Generative AI for Solving Inverse Problems in Computational Imaging. *XRDS*, 31(2):32–37, January 2025.

[88] Omkar M. Parkhi, Andrea Vedaldi, Andrew Zisserman, and C. V. Jawahar. Cats and dogs. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2012.

[89] R. Tyrrell Rockafellar and Roger J-B Wets. *Variational Analysis*, volume 317. Springer Science & Business Media, 2009.

[90] Heinz H Bauschke and Patrick L Combettes. Correction to: convex analysis and monotone operator theory in hilbert spaces. In *Convex analysis and monotone operator theory in Hilbert spaces*, pages C1–C4. Springer, 2020.

A Additional Comparisons

Figures S1-S6 and Figures S7-S12 illustrate qualitative comparisons of all methods on AFHQ-Cat and CelebA, respectively.

Tables S1 and S2, and Tables S3 and S4 provide PSNR and LPIPS for distortion metrics, and KID for perception metric, for all methods AFHQ-Cat and CelebA, respectively.

Table S5 provides running time (in seconds) of all methods for a single image.

A.1 Qualitative comparisons



Figure S1: Qualitative comparison of image restoration results on AFHQ-Cat (deblurring).

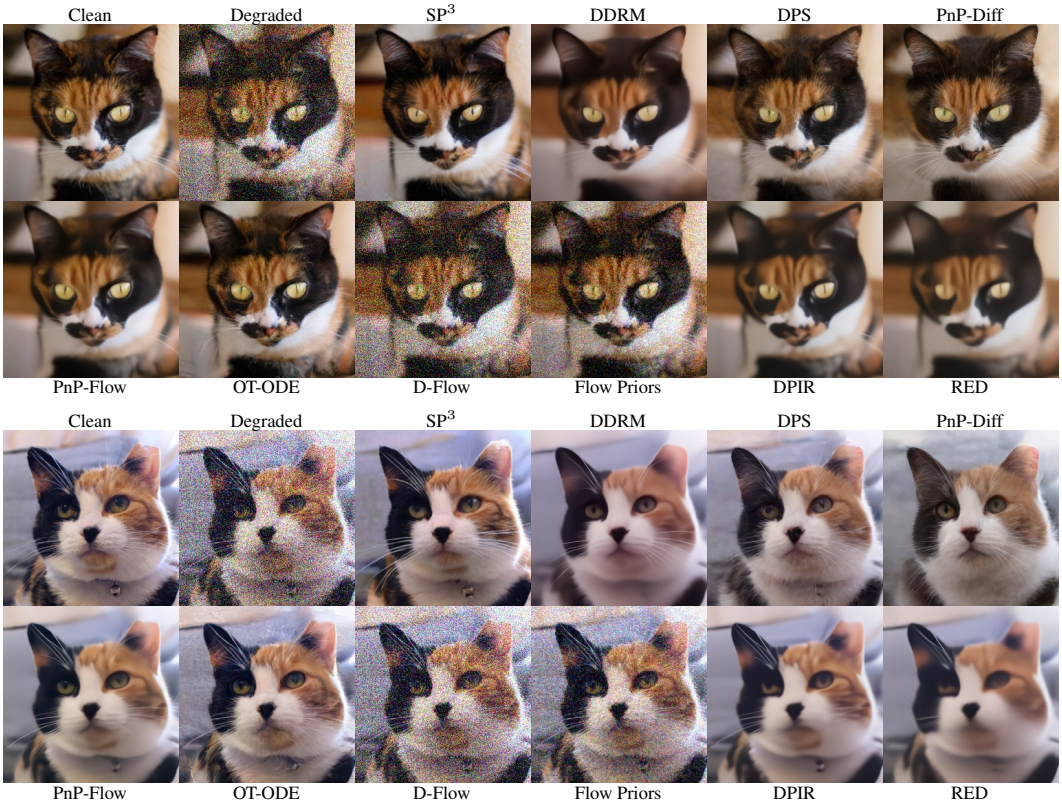


Figure S2: Qualitative comparison of image restoration results on AFHQ-Cat (denoising).

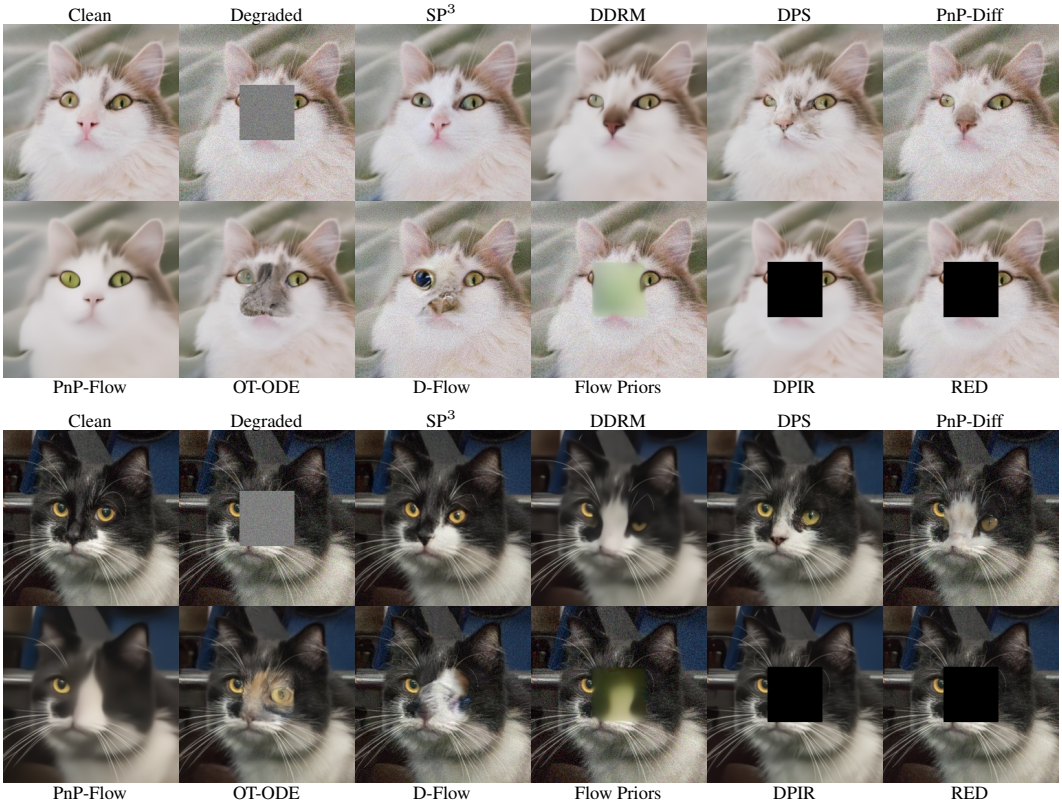


Figure S3: Qualitative comparison of image restoration results on AFHQ-Cat (inpainting).



Figure S4: Qualitative comparison of image restoration results on AFHQ-Cat (paintbrush-inpainting).

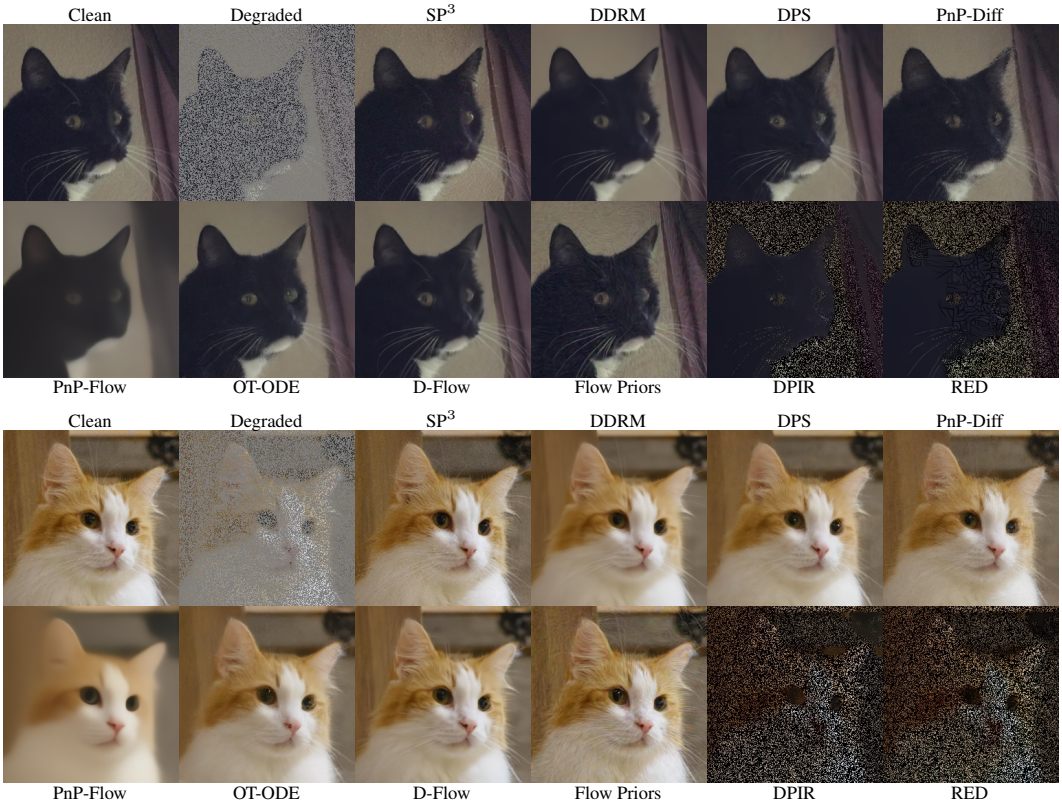


Figure S5: Qualitative comparison of image restoration results on AFHQ-Cat (random inpainting).

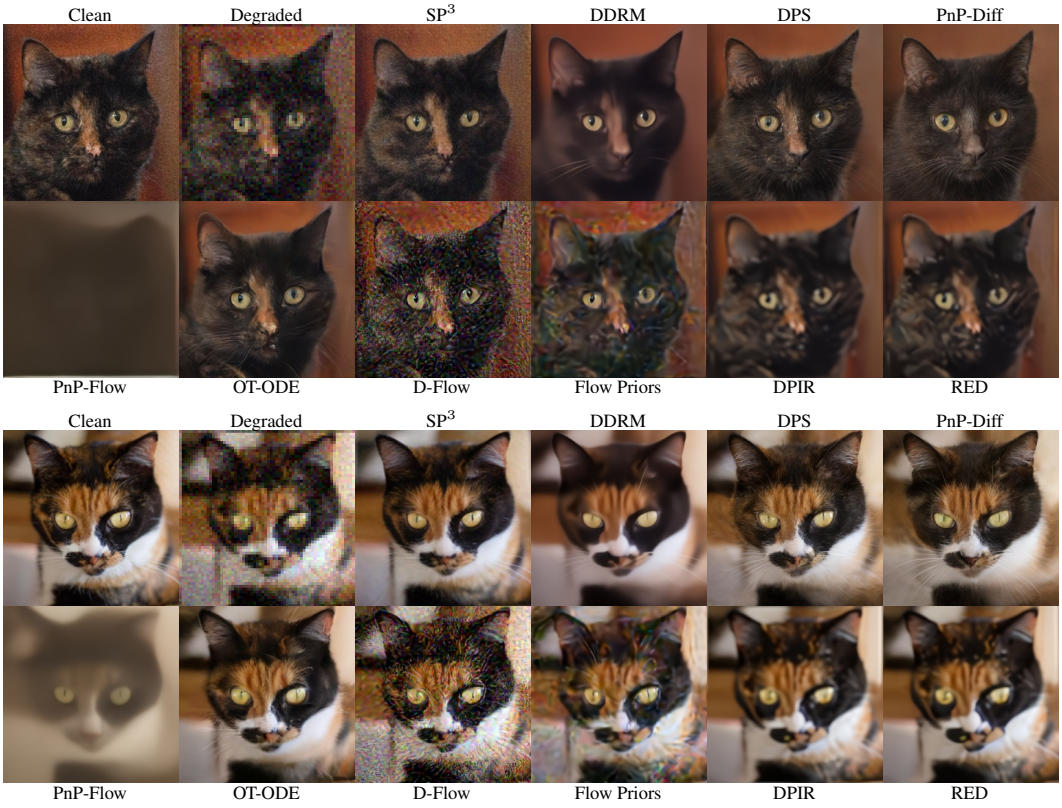


Figure S6: Qualitative comparison of image restoration results on AFHQ-Cat (super-resolution).



Figure S7: Qualitative comparison of image restoration results on CelebA (deblurring).

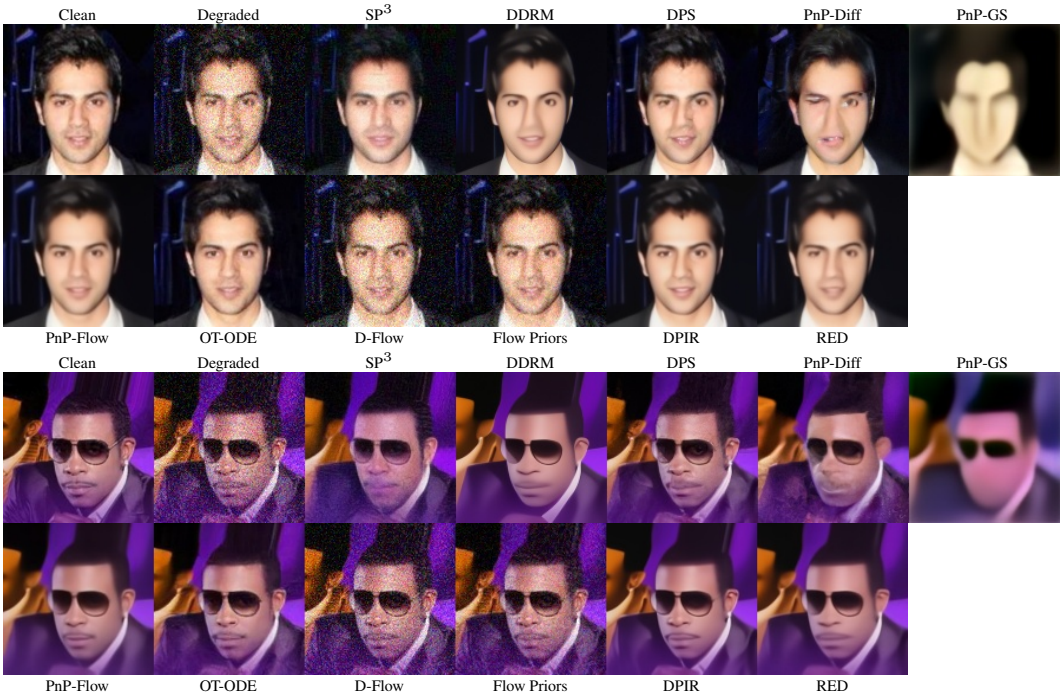


Figure S8: Qualitative comparison of image restoration results on CelebA (denoising).

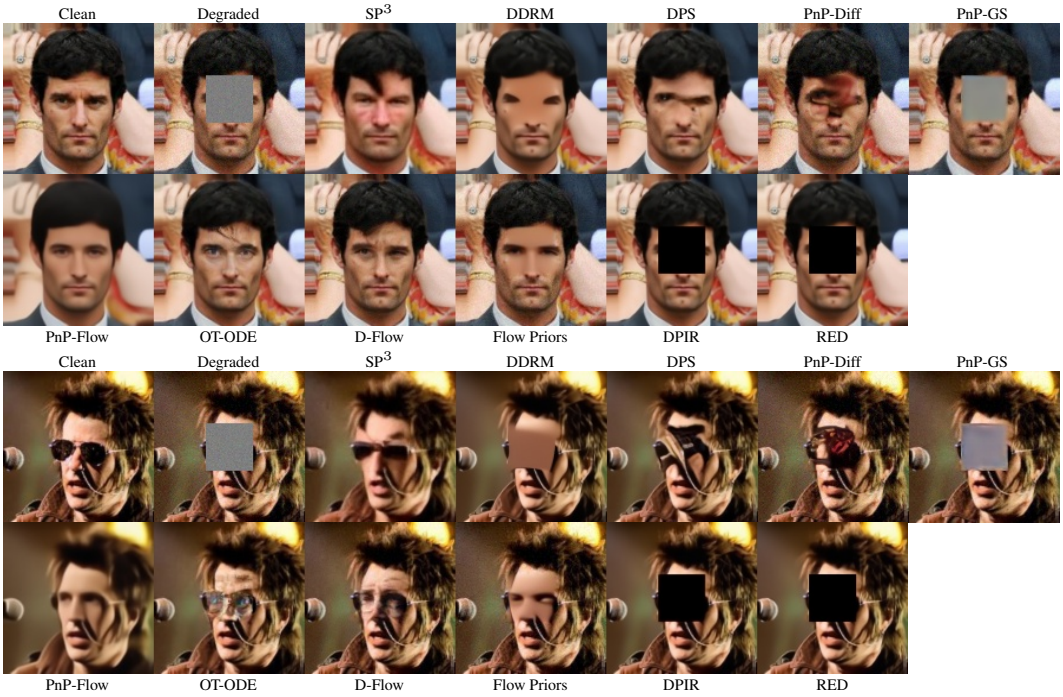


Figure S9: Qualitative comparison of image restoration results on CelebA (inpainting).



Figure S10: Qualitative comparison of image restoration results on CelebA (paintbrush-inpainting).

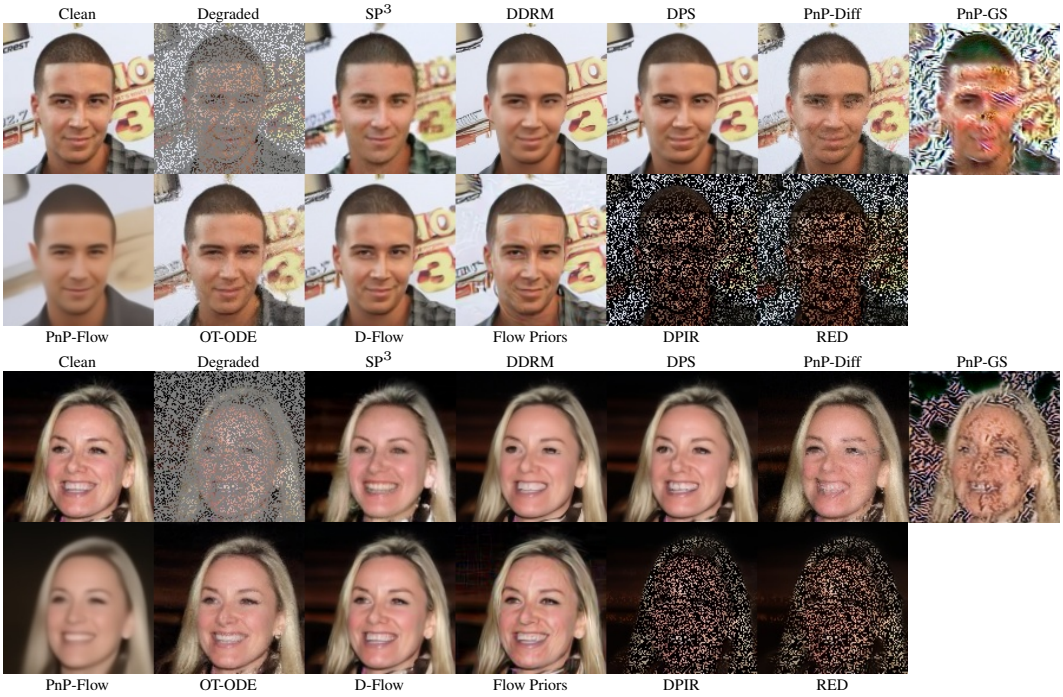


Figure S11: Qualitative comparison of image restoration results on CelebA (random-inpainting).



Figure S12: Qualitative comparison of image restoration results on CelebA (super-resolution).

A.2 Quantitative results

On AFHQ-Cat (Table S1 and Table S2), SP^3 achieves the best KID for deblurring and super-resolution and the best LPIPS on four of the six tasks. SP^3 achieves lower PSNR, reflecting its stronger emphasis on perceptual realism rather than image fidelity.

On CelebA (Table S3 and Table S4), SP^3 offers a favorable quality-efficiency trade-off, but not a general quality advantage over diffusion-based methods.

Method	**Denoising**			**Deblurring**			**Inpainting**		
	KID ↓	**LPIPS** ↓	**PSNR** ↑	**KID** ↓	**LPIPS** ↓	**PSNR** ↑	**KID** ↓	**LPIPS** ↓	**PSNR** ↑
DPS	3.63	0.159	26.38	2.69	0.178	24.72	3.91	0.107	25.85
OT-ODE	7.71	0.146	27.32	6.74	0.155	26.79	8.53	0.129	22.95
DiffPIR	5.74	0.235	23.05	128.68	0.584	16.71	15.06	0.267	23.72
PnP-Flow	36.74	0.259	29.49	64.01	0.446	25.91	58.89	0.328	25.57
DDRM	48.83	0.336	26.74	41.93	0.325	27.34	40.21	0.220	26.54
RED	55.42	0.414	25.99	49.44	0.398	28.25	106.65	0.206	17.20
DPIR	66.70	0.336	27.97	40.84	0.401	28.08	122.08	0.230	17.27
Flow Priors	38.86	0.700	19.33	21.18	0.297	26.46	30.95	0.301	24.34
D-Flow	60.17	1.039	14.97	7.81	0.170	27.20	20.70	0.224	23.71
SP^3	4.09	0.174	22.84	2.56	0.118	25.75	6.68	0.103	25.12

Table S1: Quantitative evaluation on AFHQ-Cat. KID is scaled by $\times 10^3$.

Method	**Paintbrush Inp.**			**Random Inp.**			**Superresolution**		
	KID ↓	**LPIPS** ↓	**PSNR** ↑	**KID** ↓	**LPIPS** ↓	**PSNR** ↑	**KID** ↓	**LPIPS** ↓	**PSNR** ↑
DPS	4.53	0.103	26.92	3.15	0.100	29.26	3.50	0.201	24.32
OT-ODE	9.10	0.120	24.14	4.34	0.118	28.80	6.40	0.173	26.34
DiffPIR	16.69	0.263	24.26	9.55	0.159	27.26	5.59	0.242	22.60
PnP-Flow	57.43	0.321	26.34	66.44	0.438	25.16	98.44	0.666	18.09
DDRM	40.20	0.213	27.55	20.93	0.114	32.32	50.64	0.371	25.42
RED	59.57	0.159	18.19	360.24	1.171	8.31	45.91	0.373	27.95
DPIR	96.43	0.198	18.29	333.30	1.139	7.97	47.11	0.398	27.42
Flow Priors	23.60	0.278	25.23	16.51	0.145	29.77	55.70	0.417	24.41
D-Flow	19.14	0.217	24.40	6.64	0.061	32.24	35.15	0.568	21.28
SP^3	6.91	0.099	25.73	3.71	0.085	27.62	3.47	0.151	24.24

Table S2: Quantitative evaluation on AFHQ-Cat. KID is scaled by $\times 10^3$.

Method	Denoising			Deblurring			Inpainting		
	KID ↓	LPIPS ↓	PSNR ↑	KID ↓	LPIPS ↓	PSNR ↑	KID ↓	LPIPS ↓	PSNR ↑
OT-ODE	7.70	0.044	30.47	5.90	0.037	32.80	8.29	0.050	27.57
DPS	17.52	0.041	31.30	14.67	0.041	32.09	16.16	0.055	25.14
DPIR	44.45	0.071	32.02	19.48	0.046	35.42	121.67	0.250	14.88
PnP-Flow	45.58	0.114	29.95	57.95	0.184	27.90	52.23	0.136	27.61
DiffPIR	64.02	0.109	26.47	272.39	0.951	10.73	74.04	0.105	22.61
RED	62.99	0.126	30.02	29.82	0.049	34.81	136.61	0.261	14.85
Flow Priors	86.68	0.234	24.73	22.96	0.064	31.52	12.27	0.058	29.65
DDRM	103.42	0.113	29.89	65.43	0.067	32.95	68.42	0.089	26.80
D-Flow	95.72	0.390	20.72	67.70	0.027	30.60	48.90	0.025	30.74
PnP-GS	187.39	0.418	21.46	21.22	0.057	33.41	128.87	0.155	22.97
SP^3	57.10	0.147	25.45	7.95	0.057	31.13	34.38	0.052	27.23

Table S3: Quantitative evaluation on CelebA. KID is scaled by $\times 10^3$.

Method	Paintbrush Inp.			Random Inp.			Superresolution		
	KID ↓	LPIPS ↓	PSNR ↑	KID ↓	LPIPS ↓	PSNR ↑	KID ↓	LPIPS ↓	PSNR ↑
OT-ODE	12.01	0.099	22.93	20.53	0.084	27.61	6.28	0.047	31.30
DPS	22.72	0.104	21.49	7.79	0.026	31.91	15.86	0.045	31.63
DPIR	162.28	0.291	12.54	425.53	1.018	8.45	25.84	0.067	32.75
PnP-Flow	54.00	0.160	25.06	73.93	0.233	23.32	80.71	0.272	22.14
DiffPIR	83.56	0.175	19.07	81.34	0.118	26.70	49.54	0.067	28.76
RED	224.26	0.304	12.47	434.20	1.015	8.87	29.05	0.062	33.65
Flow Priors	33.91	0.097	24.97	18.45	0.072	29.05	33.75	0.119	28.53
DDRM	86.61	0.160	23.21	26.63	0.041	31.04	87.84	0.091	30.51
D-Flow	51.96	0.045	27.43	7.08	0.021	33.28	71.54	0.044	28.62
PnP-GS	112.76	0.233	19.45	343.59	0.638	12.46	24.44	0.072	31.33
SP^3	50.00	0.088	23.48	22.39	0.080	26.06	13.80	0.067	29.46

Table S4: Quantitative evaluation on CelebA. KID is scaled by $\times 10^3$.

Method	Runtime [sec]
OT-ODE	2.29
DPS	17.77
DPIR	0.044
PnP-Flow	3.20
DiffPIR	0.83
RED	0.37
Flow Priors	29.88
DDRM	0.81
D-Flow	183.18
PnP-GS	0.31
SP^3 (1 step)	0.016
SP^3 (3 step)	0.043
SP^3 (5 step)	0.073
SP^3 (10 step)	0.14
SP^3 (20 step)	0.28

Table S5: **Average inference time per image.** Reported per-image wall-clock runtimes were measured using a single NVIDIA L40S GPU for all methods.

B Additional results

Section B.1 illustrates the importance of SE in the SP^3 framework.

Section B.2 provides the results of SP^3 on additional datasets – AFHQ-Dog and AFHQ-Wild splits.

Section B.3 provides results of SP^3 using SE trained on limited or out-of-distribution data, showing its robustness to such cases.

Section B.4 provides the mean and std results of SP^3 over ten independent runs.

Finally, Section B.5 illustrates that by averaging over multiple different runs, SP^3 can achieve better distortion performance, on the expanse of perceptual realism.

B.1 Importance of SE in the SP^3 framework

To fairly compare between FLUX.1-VAE, EQ-VAE and SE in the SP^3 framework, we kept initialization, number of iterations, and data-consistency formulation fixed. We tuned the data-fidelity regularization parameters for FLUX.1-VAE and EQ-VAE to achieve best performance. Table S6 illustrates the results, given for debulrring on AFHQ-Cat.

Prior	**KID** $\downarrow$	**LPIPS** $\downarrow$	**PSNR** $\uparrow$
FLUX.1-VAE	43.13	0.522	25.05
EQ-VAE	246.11	0.516	21.13
SE	2.56	0.118	25.75

Table S6: Comparison of SP^3 with different autoencoders for deblurring on AFHQ-Cat. KID is scaled by $\times 10^3$; PSNR is reported in dB.

B.2 Results on additional datasets

Dataset	Task	Method	KID ↓	LPIPS ↓	PSNR [dB] ↑
Dog	Denoising	SP^3	2.66	0.170	23.16
		DPS	3.33	0.139	27.73
	Deblurring	SP^3	2.52	0.102	26.93
		DPS	2.79	0.151	26.31
	Inpainting	SP^3	3.09	0.090	25.00
		DPS	2.98	0.099	26.91
	Paintbrush Inp.	SP^3	3.15	0.084	26.38
		DPS	3.28	0.096	28.19
	Random Inp.	SP^3	2.29	0.073	28.83
		DPS	3.06	0.085	31.64
	Super-resolution	SP^3	2.51	0.128	25.06
		DPS	3.45	0.175	25.78
Wild	Denoising	SP^3	16.77	0.194	21.78
		DPS	18.02	0.170	25.19
	Deblurring	SP^3	18.02	0.137	24.08
		DPS	17.96	0.195	23.04
	Inpainting	SP^3	16.93	0.105	24.92
		DPS	17.21	0.103	26.07
	Paintbrush Inp.	SP^3	17.28	0.103	25.34
		DPS	17.54	0.100	26.79
	Random Inp.	SP^3	17.44	0.087	26.51
		DPS	17.99	0.115	27.61
	Super-resolution	SP^3	17.59	0.162	23.15
		DPS	18.13	0.220	22.85

Table S7: Quantitative comparison of SP^3 and DPS on AFHQ-Dog and AFHQ-Wild. For DPS, we use a separate diffusion prior trained on each class, whereas SP^3 uses a single unified SE prior. KID is scaled by $\times 10^3$.

B.3 Robustness to SE trained on limited or out-of-distribution data

SP^3 is a PnP method that uses a generative prior to project images onto a learned manifold. If these projections fail, for instance since the SE was trained on limited or out-of-domain data, the restoration quality of SP^3 will naturally deteriorate. Importantly, this phenomenon is not unique to SP^3, but applies more broadly to PnP methods that use pretrained generative priors, such as DPS.

To evaluate the behavior of SP^3 in such cases, we used a SE trained on AFHQ-Cat to restore cat images from the Oxford-IIIT Pet dataset [88]. The images were cropped around the cats' faces and degraded with blur. Although these images belong to a similar semantic domain as AFHQ-Cat, they exhibit a distribution shift. We also compare against DPS using a diffusion model trained on AFHQ-Cat. As shown in Table S8, SP^3 achieves better perceptual quality (KID) and better image fidelity (PSNR) than DPS, while DPS achieves slightly better LPIPS on these images. The results of DPS and SP^3 on AFHQ are given below for reference.

Method	**KID** $\downarrow$	**LPIPS** $\downarrow$	**PSNR** $\uparrow$
DPS Oxford	32.0	0.24	24.5
SP^3 Oxford	25.3	0.27	25
DPS AFHQ	2.69	0.178	24.72
SP^3 AFHQ	2.56	0.118	25.75

Table S8: Quantitative results for SP^3 where the SE prior was trained on AFHQ-Cat and evaluated on Oxford-IIIT Pet degraded with blur. KID is scaled by $\times 10^3$.

B.4 SP^3 robustness to noise realization

Experiment	PSNR ↑	LPIPS ↓	KID ↓
AFHQ Denoising	22.91 ± 0.03	0.174 ± 0.0004	3.84 ± 0.26
AFHQ Inpainting	25.15 ± 0.03	0.103 ± 0.0004	7.06 ± 0.19
AFHQ Super-resolution	24.32 ± 0.03	0.151 ± 0.0007	3.58 ± 0.28
CelebA Deblurring	31.08 ± 0.03	0.057 ± 0.0005	8.55 ± 0.30
CelebA Paintbrush Inp.	23.37 ± 0.17	0.091 ± 0.0025	49.88 ± 1.02
CelebA Random Inp.	26.08 ± 0.03	0.081 ± 0.0015	21.30 ± 0.74

Table S9: Mean results, given with standard deviation across ten independent runs, for the detailed degradations. KID is scaled by 10^3.

B.5 Perceptual realism over distortion metrics

SP^3 acts as a MAP estimator, it balances the image prior against fidelity to the observed measurements. Consequently, in the current configuration, SP^3 prioritizes perceptual realism over distortion metrics, which is a direct consequence of the perception-distortion tradeoff [44, 70]. While it might limit the use of SP^3 in fidelity-critical domains, it makes SP^3 well suited for restoring natural images, where perceptual realism is often more important to the end user.

Importantly, SP^3 can be adapted to be more distortion-oriented. By averaging multiple independently generated restorations, we can trade perceptual quality for improved distortion metrics. In Table S10, we report results obtained by averaging 10 independently restored images. Compared with the single-restoration results reported in the manuscript, PSNR improves across all tasks.

Experiment	**Method**	**PSNR** $\uparrow$	**LPIPS** $\downarrow$	**KID** $\downarrow$
AFHQ Inpainting	SP^3 Averaged	25.90	0.112	10.32
	SP^3	25.12	0.103	6.68
AFHQ Denoising	SP^3 Averaged	24.14	0.180	7.48
	SP^3	22.84	0.174	4.09
AFHQ Super-resolution	SP^3 Averaged	25.45	0.209	7.53
	SP^3	24.24	0.151	3.47
CelebA Deblurring	SP^3 Averaged	31.15	0.082	12.74
	SP^3	31.13	0.057	7.95
CelebA Paintbrush Inp.	SP^3 Averaged	24.79	0.092	56.90
	SP^3	23.48	0.088	50.00
CelebA Random Inp.	SP^3 Averaged	26.23	0.101	29.56
	SP^3	26.06	0.080	22.39

Table S10: Perceptual realism over distortion metrics. Averaging results were obtained using 10 independently restored images with different seeds. KID is scaled by $\times 10^3$.

C Implementation details

SP^3 (Algorithm 1) is primarily governed by the data-fidelity update strength and the injected latent noise scale, σ. For problems using a proximal data-fidelity update, its strength is controlled by the regularization coefficient λ. Here, σ denotes the relative noise strength with respect to the maximum noise level used to train the prior (equivalent to r in [37]). In addition to these parameters, the number of forward steps can be adjusted; however, this serves as a flexible parameter rather than a strict constraint due to the anytime restoration capabilities of SP^3.

We tailor the initialization x_0 to the degradation. For denoising and Gaussian deblurring, we apply the adjoint operator to the observation. For inpainting, we use a GPU-friendly, convolution-based approximation to median filtering. For super-resolution, we initialize using bicubic upsampling. For JPEG artifact removal and nonlinear motion deblurring, we initialize directly from the degraded image, $x_0 = y$.

The hyper-parameters for the evaluated inverse problems are listed in Tables S11 and S12. These configurations were selected empirically by grid search on a held-out validation set.

Problem	Steps	Init Mode	λ	η	σ
Denoising	20	$A^\top$	0.04	–	0.08
Gaussian Deblurring	20	$A^\top$	1.50	–	0.32
Inpainting	20	Conv	0.30	–	0.10
Paintbrush Inpainting	20	Conv	0.30	–	0.10
Random Inpainting	20	Conv	1.60	–	0.05
Super-resolution	20	Bicubic	2.20	–	0.15
JPEG Artifact Removal	20	y	4.00	–	0.15
Nonlinear Motion Blur	20	y	–	0.02	0.10

Table S11: Hyper-parameters for AFHQ-Cat. λ controls the proximal data-fidelity update; η is the learning rate for nonlinear-blur data-fidelity gradient steps.

Problem	Steps	Init Mode	λ	σ
Denoising	20	$A^\top$	0.10	0.05
Gaussian Deblurring	20	$A^\top$	0.80	0.05
Inpainting	20	Conv	0.45	0.05
Paintbrush Inpainting	20	Conv	0.80	0.15
Random Inpainting	20	Conv	0.45	0.05
Super-resolution	20	Bicubic	1.20	0.05

Table S12: Hyper-parameters for CelebA.

Each experiment used a single NVIDIA L40S GPU or NVIDIA A40 GPU, each with 49GB of memory.

C.1 Comparisons between different priors

To ensure fair comparison between Spherical Encoder and diffusion/flow-based methods, we validate that the generative models used as the backbone for all methods mostly achieve similar KID results for generation, as illustrated in Table S13.

C.2 Licences

AFHQ [80] is released under CC BY-NC 4.0; CelebA [81] is released under a custom non-commercial research license.

Prior	AFHQ-Cat	CelebA
Diffusion	14.1	104.4*
Flow	10.78	11.4
Sphere Encoder	9.5	10.4

Table S13: KID $\times 10^3$ of different priors on AFHQ-Cat and CelebA. *Despite the high KID number achieved by the diffusion model prior [83], diffusion based methods (*e.g.*, DPS) achieve state-of-the-art results.

D Theoretical Analysis and Convergence

While classical Half-Quadratic Splitting guarantees convergence for specific regularizers [78], extending these guarantees to highly non-linear deep generative priors requires treating the alternating updates as a non-convex optimization problem. Under the assumption that the objective satisfies the Kurdyka-Łojasiewicz property, alternating minimization schemes are guaranteed to converge to a critical point [77]. More recently, similar theoretical extensions have been specifically proven for HQS-based Plug-and-Play architectures [79].

To formally analyze the convergence behavior of SP^3 as a fixed-point iteration, we follow standard fixed-point and convex optimization theory [89, 90]. While the assumptions required for formal fixed-point convergence are restrictive and rarely strictly satisfied by deep generative models, they provide the mathematical framework that motivates our empirical analysis in Section 3.

D.1 Preliminaries and Notations

Let $\mathcal{X} = \mathbb{R}^n$ be a finite-dimensional image space, $\mathcal{V} = \mathbb{R}^m$ be a finite-dimensional latent space, and $\mathcal{Y} = \mathbb{R}^k$ be a finite-dimensional measurement space, with a linear forward operator $A : \mathcal{X} \to \mathcal{Y}$ containing a non-trivial null-space.

Our algorithm, SP^3, updates the image state via the operator T:

$$T(x) = \text{prox}_{\eta\ell}(P(x)), \tag{S1}$$

where $P(x) = D(S(E(x)))$ is the generative prior mapping defined in the main text. The data-fidelity proximal operator is defined for a convex loss $\ell(x) = \|Ax - y\|^2$ as:

$$\text{prox}_{\eta\ell}(z) = \arg\min_{x \in \mathcal{X}} \Big\{ \ell(x) + \frac{1}{2\eta}\|x - z\|^2 \Big\}. \tag{S2}$$

Note that setting $\eta = \frac{1}{2\lambda}$ exactly recovers the update step defined in Equation (12).

Definition 1 (Nonexpansive and Averaged Operators). An operator $\mathcal{T} : \mathcal{X} \to \mathcal{X}$ is nonexpansive if it is 1-Lipschitz continuous ($\|\mathcal{T}(x) - \mathcal{T}(y)\| \leq \|x - y\|$). It is α-averaged if there exists a nonexpansive operator R and a constant $\alpha \in (0, 1)$ such that $\mathcal{T} = (1 - \alpha)I + \alpha R$. An operator is firmly nonexpansive if it is exactly $1/2$-averaged.

Lemma 1 (Composition). *The composition of an α_1-averaged operator and an α_2-averaged operator is an α_3-averaged operator, where $\alpha_3 \in (0, 1)$ [90].*

D.2 Proof of Convergence

Because $\ell(x)$ is closed, proper, and convex, its proximal operator $\text{prox}_{\eta\ell}$ is firmly nonexpansive ($1/2$-averaged). If we assume the deep prior P acts as an averaged operator, we can guarantee convergence to a fixed point.

Theorem 1 (Convergence to a fixed point). *Assume that the prior mapping P is α-averaged on $\mathcal{X}$ and the fixed-point set $\text{Fix}(T) = \{x : T(x) = x\}$ is nonempty. Then, the sequence generated by $x_{k+1} = T(x_k)$ converges to a fixed point $x_\infty \in \text{Fix}(T)$.*

Proof. By Lemma 1, the composition of the $1/2$-averaged $\text{prox}_{\eta\ell}$ and the α-averaged P means the update operator T is β-averaged for some $\beta \in (0, 1)$. A fundamental property of a β-averaged operator is that for any $x, y \in \mathcal{X}$:

$$\|T(x) - T(y)\|^2 \leq \|x - y\|^2 - \frac{1 - \beta}{\beta}\|(I - T)(x) - (I - T)(y)\|^2. \tag{S3}$$

Let $x^* \in \text{Fix}(T)$. Substituting $y = x^*$ (noting $T(x^*) = x^*$) and our iteration $x_{k+1} = T(x_k)$, we obtain:

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \frac{1 - \beta}{\beta}\|x_k - x_{k+1}\|^2. \tag{S4}$$

Because the subtracted term is non-negative, $\|x_{k+1} - x^*\| \leq \|x_k - x^*\|$. Therefore, the sequence $\{x_k\}$ is Fejér monotone, bounded, and non-increasing in distance to any fixed point.

Summing Equation (S4) from $k = 0$ to $K - 1$ yields a telescoping sum bounding the residuals:

$$\frac{1-\beta}{\beta} \sum_{k=0}^{K-1} \|x_k - x_{k+1}\|^2 \leq \|x_0 - x^*\|^2. \tag{S5}$$

Because the infinite sum is bounded by a finite constant, the individual terms must approach zero ($\|x_k - x_{k+1}\| \to 0$), demonstrating the sequence is asymptotically regular. In a finite-dimensional space, the Krasnosel'skii-Mann theorem dictates that a sequence which is both bounded and asymptotically regular under an averaged operator strictly converges to a point in the fixed-point set. Thus, $x_k \to x_\infty \in \mathrm{Fix}(T)$. ■

Remark 1 (On Linear Convergence). If the data fidelity term $\ell(x)$ was strongly convex, its proximal operator would be a strict contraction, yielding a linear convergence rate via the Banach Fixed Point Theorem. However, in typical image restoration tasks (*e.g.*, deblurring, inpainting), the forward operator A has a non-trivial null-space, meaning $\ell(x)$ is not strongly convex. Consequently, theoretical bounds on the convergence rate are difficult to establish, which motivates the empirical convergence validation provided in Figure 4.

D.3 Additional empirical evidence of SP^3 averagedness behavior

Theorem 1 is conditioned on the prior operator (P) being globally averaged, which is rarely the case. Thus, the theorem only provides intuition for the convergence behavior we observe empirically in Figure 4.

Although proving that P is an averaged operator is beyond the scope of our work, we performed a new empirical analysis on its local behavior along restoration trajectories. At iterations $k \in \{0, 1, 2, 3, 5, 10, 20\}$, we formed symmetric perturbation pairs:

$$x_k^- = x_k - \delta/2, \qquad x_k^+ = x_k + \delta/2, \tag{S6}$$

where $\delta = \varepsilon \frac{u}{\mathrm{RMS}(u)}$, ε is the magnitude in range 0.5/255,1/255,2/255, and u is a sampled perturbation. We used both isotropic and high-frequency perturbations, where the data-proximal step provides less stabilization. We separately evaluated the SE prior operator and the complete SP^3 update ($T = \mathrm{prox}_\lambda(P)$). We tested both the deployed stochastic SE and the deterministic variant ($\sigma = 0$). For the stochastic SE, each perturbation pair used the same latent-noise realization. For each pair, we first measured the local sensitivity,

$$g_F = \frac{|F(x_k^+) - F(x_k^-)|_2}{|x_k^+ - x_k^-|_2}, \qquad F \in P, T. \tag{S7}$$

Across 16 images, two perturbation types, three perturbation radii, seven iterations, two noise samples, and one deterministic setting, each measured for both P and T (32,256 measurements in total), none of the tested pairs was expansive (*i.e.*, $g_F < 1$). The largest observed 95th-percentile g_F was below $\sim$0.4 for both P and T. Thus, it illustrates that stability is not merely introduced by the data-proximal step. Moreover, the stochastic SE was generally more contractive than the deterministic variant (0.1 vs. 0.2 for P, and 0.08 vs. 0.17 for T, averaged across iterations).

We then directly evaluated the averagedness inequality:

$$|F(x_k^-) - F(x_k^+)|_2^2 \leq |x_k^- - x_k^+|_2^2 - \frac{1-\alpha}{\alpha} |(I - F)(x_k^-) - (I - F)(x_k^+)|_2^2. \tag{S8}$$

All sampled pairs satisfied this inequality for $\alpha = 0.9$ and $\alpha = 0.99$, for both P and T, with both stochastic and deterministic SE variants, across all tested iterations and perturbation scales. At the final iteration, the smallest compatible α for the sampled pairs was tightly concentrated at $\sim$0.5, across the three perturbation radii.

These trajectory-local measurements do not establish global averagedness. Nevertheless, these experiments provide evidence that, in neighborhoods and directions relevant to the restoration trajectory, the learned SE projector and the complete SP^3 update behave as locally contractive operators compatible with the theorem's averagedness assumption. Figures S13 and S14 show the SE prior's gain and local averagedness.

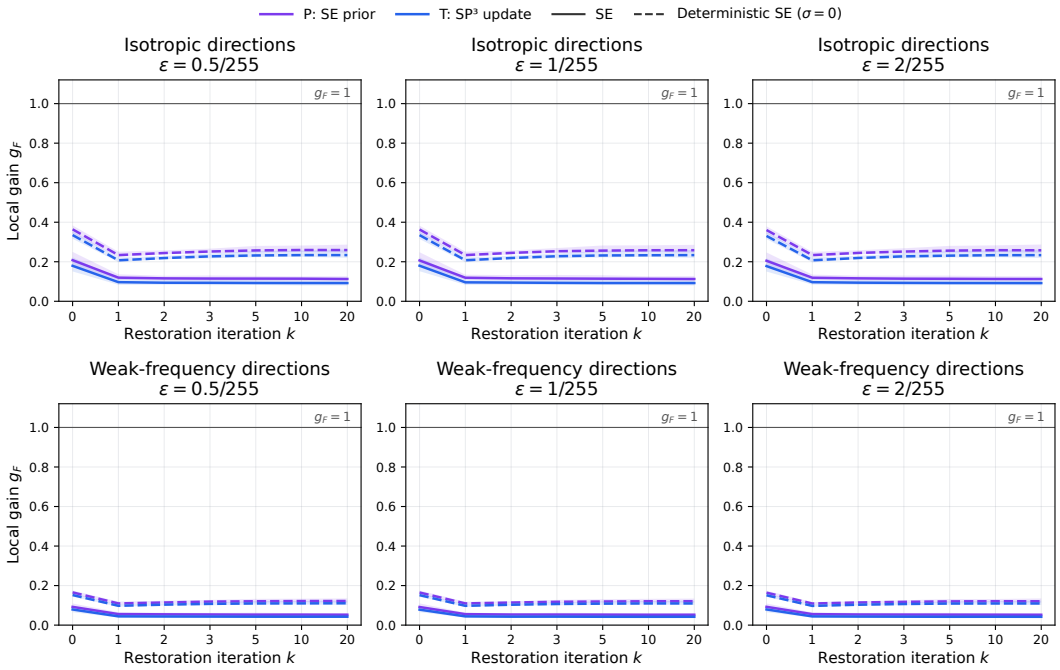


Figure S13: **Local contraction across perturbation scales**. Local secant gains of the SE prior P (purple) and the complete SP^3 update T (blue) along restoration trajectories. Rows show isotropic and weak-frequency perturbations; columns show perturbation radii $\epsilon \in \{0.5, 1, 2\}/255$. Curves indicate median gains and shaded bands span the 5th–95th percentiles across 16 images, sampled directions, and latent-noise draws. Solid and dashed curves denote SE and deterministic SE ($\sigma = 0$), respectively. Gains remain well below unity across all tested scales, with SE generally more contractive than its deterministic variant, supporting local stability of the prior itself as well as the complete update.

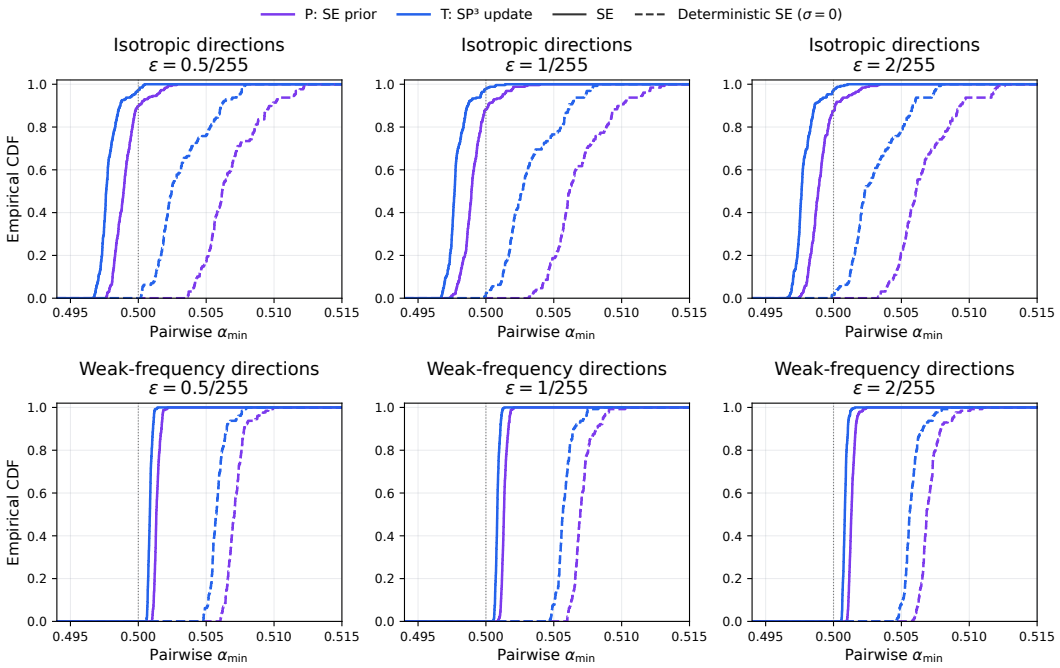


Figure S14: **Local compatibility with the averagedness inequality**. Empirical cumulative distributions of the smallest compatible averagedness parameter $\alpha_{\min}$ for each sampled perturbation pair at restoration iteration $k = 20$. Rows show isotropic and weak-frequency perturbations; columns show perturbation radii $\epsilon \in \{0.5, 1, 2\}/255$. Purple and blue curves represent the SE prior P and the complete SP^3 update T; solid and dashed curves denote SE and deterministic SE ($\sigma = 0$), respectively. The dotted vertical line marks $\alpha = 0.5$. Pairwise thresholds concentrate near 0.5 across all tested scales, supporting trajectory-local compatibility with the averagedness assumption without establishing global averagedness.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Section 3 presents the main algorithm, including the SE prior analysis and convergence analysis (expanded in Appendix D); Section 4 presents qualitative and quantitative experiments, including runtime and anytime restoration demonstrations.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Section 5 presents the method's limitations, and the theoretical analysis limitations are elaborated in Section 3 and Appendix D.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: Appendix D includes a complete set of notations and assumptions as well as a complete proof for the fixed-point convergence theorem. Section 3 includes the setup and an informal version of the results, while clearly referring the reader to the relevant supplemental material.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The full algorithm is described in Algorithm 1 with hyper-parameters per dataset and task in Appendix C. The full experimental setup is described in Section 4. The priors trained for the paper are based on publicly available data and configurations as described in Section 4. All competing methods are based on publicly available implementations as described in Section 4. We plan to publish the code and checkpoints.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The datasets used in the paper are publicly available; The code will be released upon acceptance; The paper includes all needed information to reproduce the results.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: All relevant experimental settings are presented in Section 4 and Appendix C.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: We evaluate our method using standard metrics on fixed benchmark datasets. In accordance with established precedent in the image restoration literature, we report the exact point estimates over the full test sets to ensure direct and fair comparability with baseline methods. We additionally provide the mean results with standard deviation for key experiments in the Supplementary Materials.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The compute resources used and needed are described in Appendix C

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We've read the NeurIPS code of ethics and ensured the writing process adhered to its rules.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: In Section 5 we included a societal impact paragraph, describing possible positive and negative impact of our work.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The paper introduces a new restoration framework, not a new model or data.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: The paper cites all relevant papers, code packages, and datasets used for the experimental section of the paper. Explicit licenses for datasets are provided in Appendix C.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.

- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: the paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: The paper does not relate to LLMs in any shape or form.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.