TRUSTWORTHY IMAGE SUPER-RESOLUTION VIA GENERATIVE PSEUDOINVERSE

Andreas Floros[⋆], **Seyed-Mohsen Moosavi-Dezfooli**[‡] **& Pier Luigi Dragotti**[⋆]
[⋆]Imperial College London, [‡]Apple
{andreas.floros18, pld}@imperial.ac.uk, smoosavi@apple.com

ABSTRACT

We consider the problem of trustworthy image restoration, taking the form of a constrained optimization over the prior density. To this end, we develop generative models for the task of image super-resolution that respect the degradation process and that can be made asymptotically consistent with the low-resolution measurements, outperforming existing methods by a large margin in that respect.

1 INTRODUCTION

Image Super-Resolution (SR) defines ill-posed inverse problems where the task is to recover a high-resolution image, $\mathbf{x}$, from Low-Resolution (LR) measurements, $\mathbf{y}$, which are obtained from a degradation process described by some function $\mathcal{D} : \mathcal{X} \to \mathcal{Y}$. Generative technologies are becoming increasingly more common for solving such restoration tasks. In this context, a natural concern that arises is regarding the trustworthiness of these systems. We ask the following question:

How can we maximize the quality of restored images and simultaneously minimize hallucinations?

We take a step toward answering the above, first formalizing the notion of consistency with respect to the measurements, which is a necessary condition for trustworthiness, and then providing an algorithm which is asymptotically consistent. Specifically, we develop normalizing flows that allow exploration of the data manifold while ensuring consistency. We then employ latent diffusion models for this purpose, to conditionally sample the missing details and recover high quality images that are simultaneously faithful to the originals. Chiefly, our approach differs from existing flow-based (Ardizzone et al., 2019; Liang et al., 2021) and diffusion-based (Saharia et al., 2022; Gao et al., 2023; Moser et al., 2024) SR methods in two key ways:

- We decouple the SR task, initially modeling the physics of the degradation function and then focusing on optimizing for restoration performance as a separate matter. This construction enables tighter control on consistency errors and, simultaneously, it allows our generative systems to solely focus on synthesizing the missing information.
- We recognize the strengths and weaknesses of normalizing flows and diffusion models, combining them in a way such that they complement each other and the restoration task is carried out in a compact and disentangled latent space. In doing so, we pair the interpretability of flows with the generative capacity of diffusion to tackle SR more effectively.

2 NORMALIZING FLOWS

Flow-based models are bijections with easy to compute or approximate Jacobian determinants. Their construction allows for density estimation and variational inference tasks via the change of variables formula. Of note is the affine coupling architecture (Dinh et al., 2017), which first partitions the data space, $\mathcal{X}$, into two parts, $\mathcal{U}$ and $\mathcal{V}$, applying element-wise and affine invertible transformations in an alternating fashion. For arbitrary $\boldsymbol{\phi}, \boldsymbol{\psi} : \mathcal{V} \to \mathcal{U}$ and $\boldsymbol{\eta}, \boldsymbol{\rho} : \mathcal{U} \to \mathcal{V}$, a single layer takes the form:

$$\mathbf{u} \leftrightarrow [\exp \circ \boldsymbol{\psi}](\mathbf{v})\mathbf{u} + \boldsymbol{\phi}(\mathbf{v}), \tag{1a}$$

$$\mathbf{v} \leftrightarrow [\exp \circ \boldsymbol{\rho}](\mathbf{u})\mathbf{v} + \boldsymbol{\eta}(\mathbf{u}). \tag{1b}$$

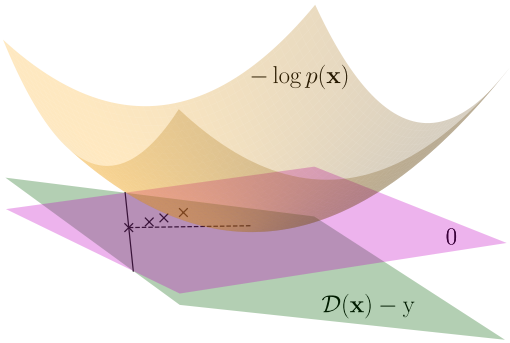
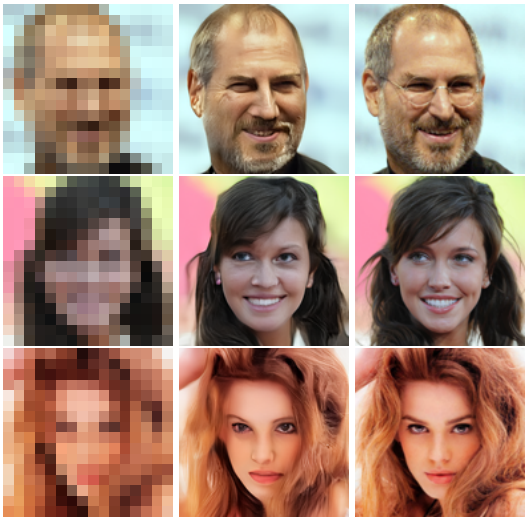


Figure 1: Linear SR problem in $\mathbb{R}^2$. We have $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $p(\mathrm{y}|\mathbf{x}) = \delta_{\mathcal{D}\mathbf{x}-\mathrm{y}}$. The intersection of the planes $\mathrm{z} = \mathcal{D}(\mathbf{x}) - \mathrm{y}$ and $\mathrm{z} = 0$ defines the feasibility set. The marked points outside of this set correspond to solutions that optimize the prior at the cost of consistency. We deem these to not be trustworthy as they violate the constraints of the problem.



(a) LR (b) SR (c) Original

Figure 2: Results of our $16 \times 16 \rightarrow 128 \times 128$ SR.

3 Diffusion Models

Diffusion is a process by which a Gaussian distribution is gradually morphed into the desired data distribution p. A popular formulation is given by Denoising Diffusion Probabilistic Models (DDPMs) (Ho et al., 2020), which reverse a stochastic process described by the following law:

$$p_{t|t-1}(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t \mathbf{I}), \tag{2}$$

where $t \in [1, T]$ with $\mathbf{x}_0 \sim p$ and β_t is the noise schedule. Reversing such a process is possible via a Gaussian denoiser that is optimal in the Mean Squared Error (MSE) sense (Anderson, 1982; Efron, 2011). In practice, one trains a single noise predicting neural network, $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\cdot, t)$, over all time-steps t. Notably, the structure of Equation 2 yields an efficient method for sampling $\mathbf{x}_t|\mathbf{x}_0$, avoiding multi-step simulations. Letting $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$, $\alpha_t = 1 - \beta_t$, this allows training with a weighted Evidence Lower Bound (ELBO) (Ho et al., 2020):

$$\min_{\boldsymbol{\theta}} \mathbb{E}_{t,\mathbf{x}_0,\boldsymbol{\epsilon}} \lambda_t \|\boldsymbol{\epsilon}_{\boldsymbol{\theta}} \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t) - \boldsymbol{\epsilon}\|_2^2. \tag{3}$$

4 Proposed Method

Let δ denote the Dirac delta and let $\mathcal{D} : \mathcal{X} \rightarrow \mathcal{Y}$ be the degradation associated with our SR problem. We are interested in sampling from the posterior, $p(\cdot|\mathbf{y})$, given evidence, $p(\mathbf{y}|\mathbf{x}) = \delta_{\mathcal{D}(\mathbf{x})-\mathbf{y}}$, and a finite data budget for approximating the prior density $p(\mathbf{x})$. In practice, black-box learning systems risk violation of the evidence constraints. Contrary to such approaches, we instead endeavor to design SR systems in a physics-aware manner, such that they respect consistency, defined as follows.

Definition 1 (Consistency)**.** *For $\mathbf{y} = \mathcal{D}(\mathbf{x})$, we say that $\mathbf{x}'$ is consistent or feasible when $\mathcal{D}(\mathbf{x}') = \mathbf{y}$.*

The solvers we are looking for naturally take the form of generalized inverses, extending the notion from the linear case to potentially non-linear operators. Thus, we will parameterize our neural networks as such. We give our definition of generalized inverses and a motivating example below.

Definition 2 (Generalized Inverse)**.** *A (stochastic) generalized reflexive inverse for $\mathcal{D} : \mathcal{X} \rightarrow \mathcal{Y}$ is any operator, $\mathcal{D}^\dagger : \mathcal{Y} \rightarrow \mathcal{X}$, satisfying $\mathcal{D} \circ \mathcal{D}^\dagger \circ \mathcal{D} \stackrel{d}{=} \mathcal{D}$ and $\mathcal{D}^\dagger \circ \mathcal{D} \circ \mathcal{D}^\dagger \stackrel{d}{=} \mathcal{D}^\dagger$ (reflexive).*

Remark 1. *Consider linear SR problems under a Gaussian prior, detailed in Figure 1. The maximum a posteriori, consistent solution is obtained by application of Moore-Penrose pseudoinverse.*

(a) LR (b) SR3 (c) IDM (d) DiWa (e) Ours (f) Original

Figure 3: $8\times$ SR. Based on the restored eyes, nose and lips, our method surpasses the competitors.

Algorithm 1 Degradation flow training

Input: Data, $p(\mathbf{x})$, degradation function $\mathcal{D}$
Parameters: Initial flow, $\mathbf{f}_{\boldsymbol{\xi}}$, $\sigma > 0$
Output: Learned $\boldsymbol{\xi}$

repeat
  $\mathbf{x} \sim p$
  $[\mathbf{y}, \mathbf{z}] = \mathbf{f}_{\boldsymbol{\xi}}(\mathbf{x})$
  $\log p_{\mathcal{Y}\times\mathcal{Z}}(\mathbf{y}, \mathbf{z}) \approx -\frac{1}{2\sigma^2}\|\mathbf{y} - \mathcal{D}(\mathbf{x})\|_2^2 - \frac{1}{2}\|\mathbf{z}\|_2^2$
  Take gradient ascent step on
$$\nabla_{\boldsymbol{\xi}} \left\{ \log p_{\mathcal{Y}\times\mathcal{Z}}(\mathbf{y}, \mathbf{z}) + \log \left| \det \frac{\partial[\mathbf{y}, \mathbf{z}]}{\partial \mathbf{x}} \right| \right\}$$
until converged
return $\boldsymbol{\xi}$

Algorithm 2 Refinement via diffusion

Input: Data, $p(\mathbf{x})$, trained flow $\mathbf{f}_{\boldsymbol{\xi}}$
Parameters: Initial DDPM, $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}$, $\{\beta_t\}_{t=1}^T$
Output: Learned $\boldsymbol{\theta}$

repeat
  $[\mathbf{y}, \mathbf{z}] \sim (\mathbf{f}_{\boldsymbol{\xi}})_{\#}p$
  $t \sim \text{Unif}(\{1, \ldots, T\})$, $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  $\mathbf{z}_t = \sqrt{\bar{\alpha}_t}\mathbf{z} + \sqrt{1 - \bar{\alpha}_t}\boldsymbol{\epsilon}$
  Take gradient descent step on
$$\nabla_{\boldsymbol{\theta}} \|\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\mathbf{z}_t, \mathbf{y}, t) - \boldsymbol{\epsilon}\|_2^2$$
until converged
return $\boldsymbol{\theta}$

4.1 Flow-Based Generative Pseudoinverse

Here we provide a construction of generalized inverses based on the coupling architecture, as discussed in Section 2. Informally, for $\mathcal{D} : \mathcal{X} \to \mathcal{Y}$, we are interested in learning a bijection such that $\mathcal{X} \leftrightarrow \mathcal{Y} \times \mathcal{Z}$, where $\mathcal{Z}$ describes the generalized kernel, which extends from the linear case.

Definition 3 (Generalized Kernel)**.** *The kernel of $\mathcal{D} : \mathcal{X} \to \mathcal{Y}$ is the set $\{\mathbf{x}' - \mathbf{x} : \mathcal{D}(\mathbf{x}') = \mathcal{D}(\mathbf{x})\}$.*

Intuitively, the kernel is the set of perturbations of solutions to our SR problem such that consistency is preserved. However, it is generally not disentangled from the measurements $\mathbf{y}$. To be precise, we learn $\mathcal{X} \leftrightarrow \mathcal{Y} \times \mathcal{Z}$ such that $\mathcal{Z}$ space is regularized to be approximately independent from $\mathbf{y}$.

Such transformations may be learned by application of the change of variables formula: $p_{\mathcal{X}}(\mathbf{x}) = p_{\mathcal{Y}}(\mathbf{y}) p_{\mathcal{Z}|\mathcal{Y}}(\mathbf{z}|\mathbf{y}) \left| \det \frac{\partial[\mathbf{y}, \mathbf{z}]}{\partial \mathbf{x}} \right|$. We use this to model $p_{\mathcal{Y}}(\mathbf{y}) = \mathbb{E}_{\mathbf{x}\sim p_{\mathcal{X}}} \mathcal{N}(\mathbf{y}; \mathcal{D}(\mathbf{x}), \sigma^2\mathbf{I})$ with $\sigma \approx 0$ and $p_{\mathcal{Z}|\mathcal{Y}}(\mathbf{z}|\mathbf{y}) = \mathcal{N}(\mathbf{z}; \mathbf{0}, \mathbf{I})$, according to Algorithm 1. Specifically, for $p_{\mathcal{Y}}(\mathbf{y})$, we model the relationship $p_{\mathcal{Y}|\mathcal{X}}(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}; \mathcal{D}(\mathbf{x}), \sigma^2\mathbf{I})$[1] with Monte Carlo simulations over $\mathbf{x}$. We summarize our method by highlighting the key properties of the flow-based construction below.

Remark 2. *Any flow such that $\mathcal{X} \leftrightarrow \mathcal{Y} \times \mathcal{Z}$ defines a degradation process, $\mathcal{D} : \mathcal{X} \to \mathcal{Y}$, along with its generalized reflexive inverses. In particular, elements in $\mathcal{Z}$ characterize the associated kernel.*

Remark 3. *Suppose the flow learns $p_{\mathcal{Y}|\mathcal{X}}(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}; \mathcal{D}(\mathbf{x}), \sigma^2\mathbf{I})$. Let $q(\mathbf{y}) := \mathbb{E}_{\mathbf{x}\sim p}\delta_{\mathcal{D}(\mathbf{x})-\mathbf{y}}$. Sampling with $p_{\mathcal{Y}}(\mathbf{y}) = q(\mathbf{y})$ is asymptotically consistent, i.e., $\mathbb{E}_{\mathbf{y}\sim q, \mathbf{x}\sim p_{\mathcal{X}|\mathcal{Y}}(\cdot|\mathbf{y})}\|\mathcal{D}(\mathbf{x}) - \mathbf{y}\|_2^2 \propto \sigma^2$.*

4.2 Refinement of the Kernel via Diffusion

Supposing the flow fits the degradation process in $\mathcal{Y}$, any further processing in $\mathcal{Z}$ preserves consistency. To overcome expressivity limitations of normalizing flows, we propose a refinement via DDPMs, shown in Algorithm 2. As this happens strictly in $\mathcal{Z}$ space, the DDPM is theoretically unconstrained. Moreover, by designing $\mathcal{Z}$ a priori, to approximate independent and identically distributed Gaussians, we intuitively simplify the reverse diffusion process and facilitate convergence. The complete sampling process with our method is given in Algorithm 3.

[1] In practice, this distribution is realized by dequantization, extending pixel values to a continuous range.

Algorithm 3 SR with generative pseudoinverse

Input: Low-resolution samples $\mathbf{y} \sim q$
Parameters: Trained $\mathbf{f}_{\boldsymbol{\xi}}$, $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}$, $\{\beta_t, \sigma_t\}_{t=1}^{T}$
Output: $\mathbf{x} \sim p_{\mathcal{X}|\mathcal{Y}}(\cdot|\mathbf{y})$

1: $\mathbf{z}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T, \ldots, 1$ **do**
3: $\quad \boldsymbol{\mu}_{t-1|t} = \frac{1}{\sqrt{\alpha_t}} \left[\mathbf{z}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\mathbf{z}_t, \mathbf{y}, t) \right]$
4: $\quad \mathbf{z}_{t-1} \sim \mathcal{N}(\boldsymbol{\mu}_{t-1|t}, \sigma_t^2 \mathbf{I})$
5: **end for**
6: **return** $\mathbf{f}_{\boldsymbol{\xi}}^{-1}(\mathbf{y}, \mathbf{z}_0)$

Table 1: $8\times$ SR methods on CelebA-HQ.

Method	PSNR ↑	SSIM ↑	Consistency ↓
PULSE	16.88	0.44	161.1
FSRGAN	23.01	0.62	33.8
SR3-R	23.96	0.69	2.71
SR3-D	23.04	0.65	2.68
IDM	24.01	**0.71**	2.14
Our method			
7+1 NFEs	**24.09**	**0.71**	0.31
100+1 NFEs	23.16	0.67	**0.06**

5 Experiments

We train our system on bicubic degradations, for $16 \times 16 \rightarrow 128 \times 128$ face SR. We use the FFHQ dataset (Karras et al., 2018b) during development and the CelebA-HQ dataset (Karras et al., 2018a) for evaluation. PSNR, SSIM and consistency errors are used to quantify performance. Consistency is calculated as the MSE ($\times 10^{-5}$) between the measurements and the downsampled SR images.

For modeling the flow, we use the affine coupling architecture from Xiao et al. (2022). In particular, a single-scale flow is chosen as this greatly simplifies our framework. It is adapted via nearest upsampling, which is trivially invertible, matching lower-resolution measurements to the $\mathcal{Y}$ space. Our DDPM is based on the architecture from Nichol & Dhariwal (2021), with a linear noise schedule that may be subsampled during evaluation for faster inference.

All experiments are conducted on a Linux cluster with NVIDIA L40S GPUs. The flow is trained for 500k iterations and our DDPM for 1M iterations with batch sizes 16 and 256 respectively. In total, our system has 110M parameters, with 31M allocated to the flow and 79M to the DDPM.

5.1 Results

We present results for face SR in Figures 2, 3 and Table 1. We compare our method with PULSE (Menon et al., 2020), FSRGAN (Chen et al., 2018), SR3 (Saharia et al., 2022), IDM (Gao et al., 2023) and DiWa (Moser et al., 2024). Our algorithm shows strong performance, surpassing IDM in terms of PSNR and consistency with only seven Neural Function Evaluations (NFEs). However, focusing on the last two rows of Table 1, we observe a trade-off between consistency and the other objective metrics. Still, when the NFEs are increased to 100, matching SR3 diffusion, we maintain competitive scores while achieving a further reduction in consistency errors by a factor of five.

6 Discussion

Having presented consistency from the perspective of trustworthiness, we expect this research to be valuable in fields such as medical, forensic imaging, remote sensing and autonomous systems, particularly in cases where noise is negligible or has been effectively mitigated. In these scenarios, maintaining the integrity of image reconstructions is paramount as inconsistencies can lead to misinterpretations with significant real-world consequences.

However, there is also an inherent trade-off between minimizing hallucinations and maximizing quality (Cohen et al., 2024). Therefore, depending on the application, strict consistency may not be ideal and alternatives that instead loosely follow the measurements may be preferred.

We conclude by noting that, although we specifically studied SR, our methods could potentially be generalized beyond inverse problems to arbitrary conditioning mechanisms, enabling deep learning systems to reliably follow guiding signals that may extend to alternative modalities.

REFERENCES

Brian D.O. Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982. ISSN 0304-4149.

Lynton Ardizzone, Jakob Kruse, Carsten Rother, and Ullrich Köthe. Analyzing inverse problems with invertible neural networks. In *International Conference on Learning Representations*, 2019.

Yu* Chen, Ying* Tai, Xiaoming Liu, Chunhua Shen, and Jian Yang. Fsrnet: End-to-end learning face super-resolution with facial priors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

Regev Cohen, Idan Kligvasser, Ehud Rivlin, and Daniel Freedman. Looks too good to be true: An information-theoretic analysis of hallucinations in generative restoration models. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

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

Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011. ISSN 01621459.

Sicheng Gao, Xuhui Liu, Bohan Zeng, Sheng Xu, Yanjing Li, Xiaoyan Luo, Jianzhuang Liu, Xiantong Zhen, and Baochang Zhang. Implicit diffusion models for continuous super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10021–10030, June 2023.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851. Curran Associates, Inc., 2020.

Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of GANs for improved quality, stability, and variation. In *International Conference on Learning Representations*, 2018a.

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4396–4405, 2018b.

Jingyun Liang, Andreas Lugmayr, Kai Zhang, Martin Danelljan, Luc Van Gool, and Radu Timofte. Hierarchical conditional flow: A unified framework for image super-resolution and image rescaling. In *IEEE International Conference on Computer Vision*, 2021.

Sachit Menon, Alexandru Damian, Shijia Hu, Nikhil Ravi, and Cynthia Rudin. Pulse: Self-supervised photo upsampling via latent space exploration of generative models. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2434–2442, 2020.

Brian B Moser, Stanislav Frolov, Federico Raue, Sebastian Palacio, and Andreas Dengel. Waving goodbye to low-res: A diffusion-wavelet approach for image super-resolution. In *2024 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. IEEE, 2024.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8162–8171. PMLR, 18–24 Jul 2021.

Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP:1–14, 09 2022.

Mingqing Xiao, Shuxin Zheng, Chang Liu, Zhouchen Lin, and Tie-Yan Liu. Invertible rescaling network and its extensions. *Int. J. Comput. Vision*, 131(1):134–159, October 2022. ISSN 0920-5691.