

Diffusion Models for Image Restoration and Super Resolution: Methods and Evaluation

Rejina P V

Assistant professor, Co-Operative Arts And Science College, Madayi , Pazhayangadi, Kannur, India

Article information

Received: 22nd May 2026

Received in revised form: 24th June 2026

Accepted: 27th July 2026

Available online: 12th September 2026

Volume: 1

Issue: 3

DOI: <https://doi.org/10.5281/zenodo.22701392>

Abstract

Image restoration asks for the clean picture behind a blurred, noisy, downsampled or partly missing one. That question has no single answer, and the honest output is a distribution rather than a point estimate. Diffusion models produce exactly that, which explains why they moved into restoration so quickly. This review covers thirty three works and separates the field into two routes. One trains a conditional model on paired data. The other reuses an unconditional prior and enforces agreement with the measurement at every sampling step, which needs no retraining but does need a forward model. We describe how each route handles blur, noise, downsampling and missing pixels, and we set out the sampling cost that both share. On evaluation we take a firm position: distortion metrics and perceptual metrics disagree by construction, so a paper reporting only one of them has not reported a result. We end with four open problems, of which blind degradation and hallucinated detail are the ones that matter most for scientific and medical use.

Keywords:- Diffusion Model, Image Restoration, Super Resolution, Inverse Problem, Score Matching, Perceptual Quality, Sampling Cost.

I. INTRODUCTION

Restoration problems are underdetermined. Downsample an image by four and each output pixel stands for sixteen inputs, so many clean images explain the same low resolution observation equally well. Early deep learning treated this as regression. SRCNN mapped low resolution to high resolution with three convolutional layers and a squared error loss [3]. The reconstructions scored well and looked soft, because a squared error loss returns the average of all plausible answers and the average of many sharp images is blurry.

Adversarial training was the first serious answer to that. SRGAN traded some fidelity for texture that looked real [4], ESRGAN refined the recipe [6], and Real-ESRGAN pushed it towards degradations found outside the laboratory [19]. Attention based backbones such as RCAN [7] and SwinIR [18] improved the regression side in parallel. Adversarial models are hard to train, though, and mode collapse is a real risk [13].

Diffusion models arrived from a different direction. Sohl-Dickstein and colleagues framed generation as the reversal of a gradual noising process [2]; Song and Ermon learned the score of the data distribution instead [11]; Ho and colleagues simplified the training objective into noise prediction and made the approach practical [12]. Within two years these models were beating adversarial ones on image synthesis [17]. Restoration followed. This review is about what happened next, what works, and what the reported numbers do and do not support.

II. HOW A DIFFUSION MODEL WORKS

The forward process is fixed, not learned. It takes a clean image and adds Gaussian noise in small steps until nothing recognisable is left. Because each step is Gaussian, the image at any step can be sampled directly from the clean image without walking the chain. The reverse process is the learned part: a network sees a noisy image and its noise level, and predicts the noise that was added. Subtract, repeat, and a sample emerges. Figure 1 shows the two directions.

Several refinements matter in practice. Nichol and Dhariwal learned the reverse variance and changed the noise schedule, which improved likelihood without hurting sample quality [16]. Song and colleagues showed that the discrete chain is one instance of a stochastic differential equation, and that the same trained model supports a deterministic probability flow sampler [15]. Karras and colleagues separated the design choices that papers had been bundling together, then tuned each one [21]. Rombach and colleagues moved the whole process into the latent space of an autoencoder, which cut the cost of high resolution work by a large factor [20].

III. TWO ROUTES TO RESTORATION

A. Train a conditional model

The direct approach conditions the reverse process on the degraded image. SR3 does this for super resolution: the low resolution input is concatenated to the noisy state at every step, and the model learns iterative refinement towards a plausible high resolution result [26]. Palette generalises the idea to colourisation, inpainting, uncropping and JPEG artefact removal, using one architecture and one loss across the four tasks [22].

This route is simple and it works. The cost is paired training data for each degradation you care about, plus a retraining run whenever the degradation changes. A model trained on bicubic downsampling will not handle sensor noise well.

B. Reuse an unconditional prior

The second route keeps a single unconditional diffusion model and injects the measurement during sampling. RePaint was an early example: for inpainting, replace the known region with a correctly noised version of the observation at each step, and resample to fix the seams [24]. DDRM formalises the idea for linear degradations by working in the singular value basis of the known degradation operator [23]. DDNM decomposes the solution into a range space fixed by the measurement and a null space filled in by the prior, which gives zero shot restoration across several tasks from one pretrained model [28]. Zero shot behaviour of this kind is attractive for the same reasons it is attractive in classification [32].

DPS extends the family to noisy and nonlinear forward models by approximating the gradient of the log likelihood at each step [27]. Medical imaging has taken up the approach quickly, since the forward operator in accelerated MRI is known exactly and the reconstruction is the unknown [25]. Table 1 sets the two routes side by side.

Table 1. The two restoration routes compared on requirements and coverage, following the descriptions given in the cited papers.

Property	Route A: conditional training	Route B: prior plus consistency
Training data	Paired clean and degraded	Clean images only
Retraining per task	Yes	No
Forward operator	Learned implicitly	Must be known or estimated
Degradations covered	Those seen in training	Linear [23], [28]; nonlinear [27]
Representative works	SR3 [26], Palette [22]	RePaint [24], DDRM [23], DDNM [28], DPS [27]

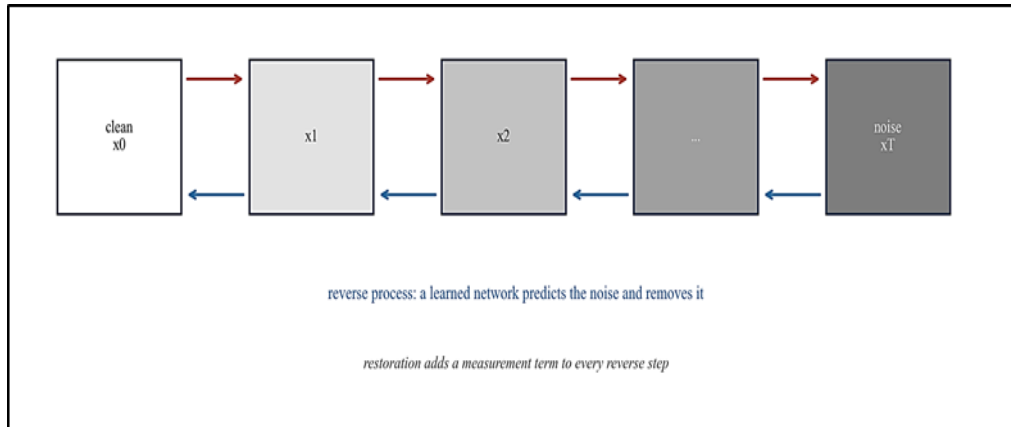


Figure. 1: The forward chain adds noise on a fixed schedule; the reverse chain removes it with a learned network. Restoration methods add a measurement term to the reverse step [23], [27], [28].

IV. SAMPLING COST

Diffusion sampling is slow. A thousand network evaluations per image was normal in the early work, against a single forward pass for a regression model. Three lines of work attack this. DDIM reformulates sampling as a non-Markovian process whose deterministic limit produces good images in tens of steps rather than hundreds [14]. Higher order and better tuned solvers for the probability flow equation cut the count further [21]. Consistency models go further still by training a network to map any point on the trajectory directly to its origin, which supports single step generation [29].

Cost matters more for restoration than for synthesis. A generative art tool can wait ten seconds. A radiology console cannot, and neither can a camera pipeline. Where the model must run near the sensor, the deployment question becomes an edge computing question rather than a modelling one [33]. Latent space diffusion helps here as well, because the expensive steps happen at reduced resolution [20].

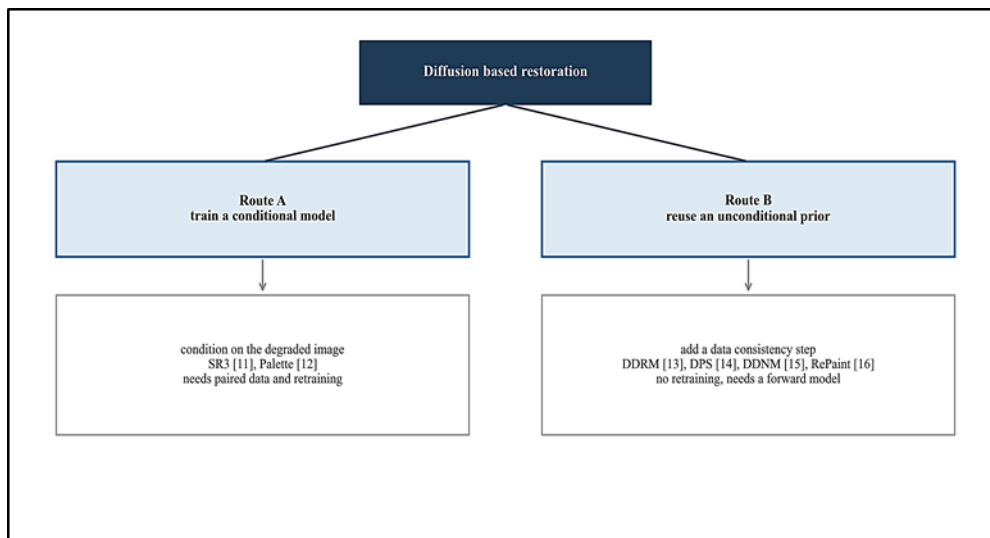


Figure. 2: The two routes used to organise this review, with the primary works assigned to each.

V. MEASURING RESTORATION QUALITY

Peak signal to noise ratio and SSIM [1] measure distortion: how close the output is to the reference. LPIPS [9] and FID [5] measure perceptual quality: whether the output looks like a real image. These two families are not merely different. Blau and Michaeli proved that they are in tension, and that improving perceptual quality past a point must cost distortion [10]. A diffusion sample sits on the perceptual side of that curve by design.

The practical consequence is blunt. A paper that reports PSNR alone will make diffusion look worse than regression. A paper that reports FID alone will make it look better than it is. Both numbers, on the same split, are the minimum. The NTIRE challenge protocol is one of the few settings where splits and degradations are pinned down well enough for cross paper comparison [8]. Table 2 lists what each metric actually measures and where it misleads.

Table 2. Metrics used across the reviewed papers, with the property each one tracks and the situation in which it is known to mislead.

Metric	Tracks	Known failure mode	Source
PSNR	Pixel level distortion	Rewards blur over texture	[10]
SSIM	Local structure and contrast	Insensitive to fine texture loss	[1]
LPIPS	Perceptual distance in feature space	Depends on the backbone network	[9]
FID	Distance between image distributions	Unstable on small sample sets	[5]

One more caution. Diffusion restoration is stochastic, so two runs give two different outputs from the same input. Reporting a single sample hides that. Where the restored image feeds a downstream measurement, the spread across samples is the number a reader needs, and very few papers report it.

VI. OPEN PROBLEMS

Blind degradation comes first. Route B assumes the forward operator is known, which is true for MRI undersampling and false for a photograph of unknown provenance. Estimating the operator and the image together remains awkward.

Hallucination comes second, and it is the one that should worry practitioners most. A diffusion model fills the null space with detail drawn from its prior. That detail is plausible and it is invented. In a photograph nobody is harmed. In a scan read for diagnosis, or in forensic imagery, an invented texture is a fabricated finding. The surveys covering this literature raise the point [30], [31], but no standard test for it has been agreed.

Third, evaluation on real degradations rather than synthetic ones. Fourth, the cost question again: single step distillation looks promising [29], yet its behaviour on restoration tasks under strict data consistency is not well characterised.

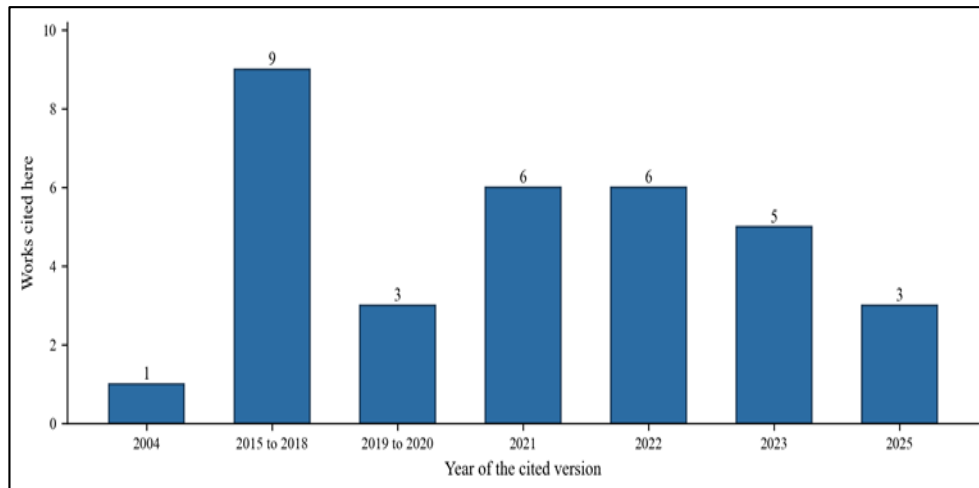


Figure. 3. Year of the cited version for the thirty three works in our reference list. The chart counts our own bibliography and reports no experimental result.

VII. CONCLUSION

Diffusion models fit restoration because restoration is a sampling problem wearing the costume of a regression problem. The plug and play route, where a pretrained unconditional prior meets a known forward operator, is the more useful of the two for most readers: one model covers many tasks, and no paired data are needed. Conditional training still wins when the degradation is fixed and paired data are plentiful.

Two recommendations follow from the reading. Report a distortion metric and a perceptual metric together, always. And treat generated detail as a hypothesis rather than a measurement, particularly in medical and scientific imaging, until a test for hallucinated structure exists.

REFERENCES

- [1] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.
- [2] J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2015, pp. 2256–2265.
- [3] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, 2016.
- [4] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 105–114.
- [5] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2017, pp. 6626–6637.
- [6] X. Wang *et al.*, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. Workshops (ECCVW)*, 2018, pp. 63–79.
- [7] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 294–310.
- [8] R. Timofte, S. Gu, J. Wu, and L. Van Gool, "NTIRE 2018 challenge on single image super-resolution: Methods and results," in *Proc. CVPR Workshops*, 2018, pp. 852–863.
- [9] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. CVPR*, 2018, pp. 586–595.
- [10] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *Proc. CVPR*, 2018, pp. 6228–6237.
- [11] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2019, pp. 11895–11907.
- [12] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. NeurIPS*, 2020.
- [13] I. Goodfellow *et al.*, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [14] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2021.
- [15] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," in *Proc. ICLR*, 2021.
- [16] A. Q. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," in *Proc. ICML*, 2021, pp. 8162–8171.
- [17] P. Dhariwal and A. Q. Nichol, "Diffusion models beat GANs on image synthesis," in *Proc. NeurIPS*, 2021, pp. 8780–8794.
- [18] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using Swin transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, 2021, pp. 1833–1844.
- [19] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data," in *Proc. ICCVW*, 2021, pp. 1905–1914.
- [20] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and D. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proc. CVPR*, 2022, pp. 10674–10685.
- [21] T. Karras, M. Aittala, A. Aila, and S. Laine, "Elucidating the design space of diffusion-based generative models," in *Proc. NeurIPS*, 2022.
- [22] C. Saharia *et al.*, "Palette: Image-to-image diffusion models," in *Proc. ACM SIGGRAPH*, 2022, pp. 15:1–15:10.
- [23] B. Kawar, M. Elad, S. Ermon, and J. Song, "Denoising diffusion restoration models," in *Proc. NeurIPS*, 2022, pp. 23593–23606.
- [24] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. Van Gool, "RePaint: Inpainting using denoising diffusion probabilistic models," in *Proc. CVPR*, 2022, pp. 11451–11461.
- [25] H. Chung and J. C. Ye, "Score-based diffusion models for accelerated MRI," *Med. Image Anal.*, vol. 80, 2022, Art. no. 102479.
- [26] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, pp. 4713–4726, 2023.
- [27] H. Chung, J. Kim, M. T. McCann, M. L. Klasky, and J. C. Ye, "Diffusion posterior sampling for general noisy inverse problems," in *Proc. ICLR*, 2023.
- [28] Y. Wang, J. Yu, and J. Zhang, "Zero-shot image restoration using denoising diffusion null-space model," in *Proc. ICLR*, 2023.
- [29] Y. Song, P. Dhariwal, M. Chen, and I. Sutskever, "Consistency models," in *Proc. ICML*, 2023, pp. 32211–32252.
- [30] F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah, "Diffusion models in vision: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 9, pp. 10850–10869, 2023.
- [31] X. Li *et al.*, "Diffusion models for image restoration and enhancement: A comprehensive survey," *Int. J. Comput. Vis.*, vol. 133, no. 11, pp. 8078–8108, 2025.
- [32] R. P. V., "Zero-shot learning: Enhancing models to recognize unseen classes with minimal data," *Int. J. Inf. Technol. Res. Stud. (IJITRS)*, vol. 1, no. 3, p. 157, Oct. 2025, doi: 10.63090/ijitrs/3139.3209.0014.
- [33] S. Charly, "Edge computing in networks: Reducing latency using AI-driven edge computing strategies," *Int. J. Inf. Technol. Res. Stud. (IJITRS)*, vol. 1, no. 2, p. 106, Jul. 2025, doi: 10.63090/ijitrs/3139.3209.0009.