Exploring Diffusion Models: A Survey in Super-Resolution and In-Painting Applications

Stelios Zarifis [1]

[1]National Technical University of Athens

Fundamentals of Diffusion Models

Deep Unsupervised Learning using Nonequilibrium Thermodynamics

The core concept of this study [1] draws inspiration from non-equilibrium thermodynamics. The central objective is to methodically and gradually disrupt the inherent structure present in a data distribution by employing an iterative forward diffusion process. Following this process, a model can acquiring insights into a reverse diffusion process designed to systematically reintroduce structure into the data.

Figure 1. The analogy between a diffusion process and the dispersal of color in water.

In microscopic contexts such as Brownian motion, where the positional updates follow small Gaussian distributions, reversing the diffusion process is feasible.

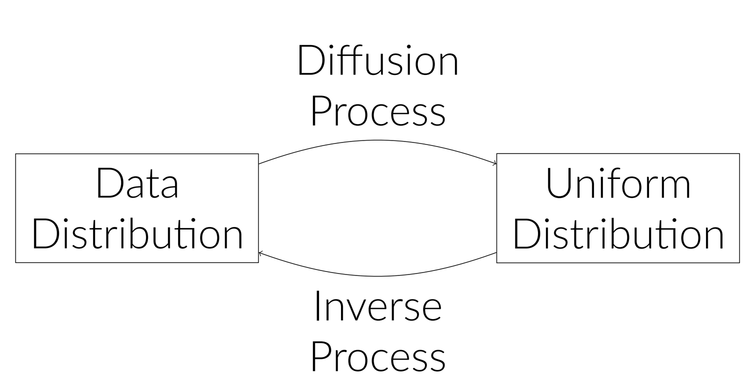

Figure 2. Illustration of the diffusion and inverse processes.

Denoising Diffusion Probabilistic Models

DDPMs [2] are generative models leveraging denoising processes, bounded on a Markov chain, to capture complex data distributions. A forward process gradually destroys the input with small Gaussian perturbations $q(x_t \mid x_{t-1})$. Mean while, a neural network learns the underlying patterns of the data. The reverse process uses the same Markov chain, in reverse. A sample from the prior distribution passes through the learned kernels $p_\theta(x_{t-1} \mid x_t)$ to form an image similar to the ones of the dataset.

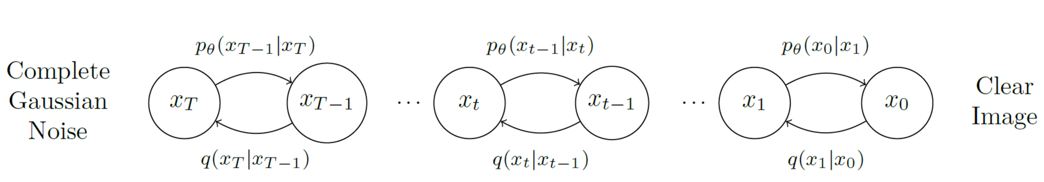

Figure 3. Forward and Reverse Diffusion Process Markov Chain.

Score-Based Generative Models with Stochastic Differential Equations

Score-Based Generative Modeling through Stochastic Differential Equations provides an equivalent perspective to DDPMs. The method involves perturbing data using continuous SDEs, enabling generation through both forward and reverse diffusion processes. The SDE is defined as:

$$d\vec{x} = \underbrace{\vec{f}(\vec{x},t)dt}_{\text{Deterministic Drift}} + \underbrace{g(t)d\vec{w}}_{\text{Stochastic Diffusion}}, \quad \vec{w}: \text{Brownian Motion}$$

Forward Diffusion Process as Stochastic Differential Equation

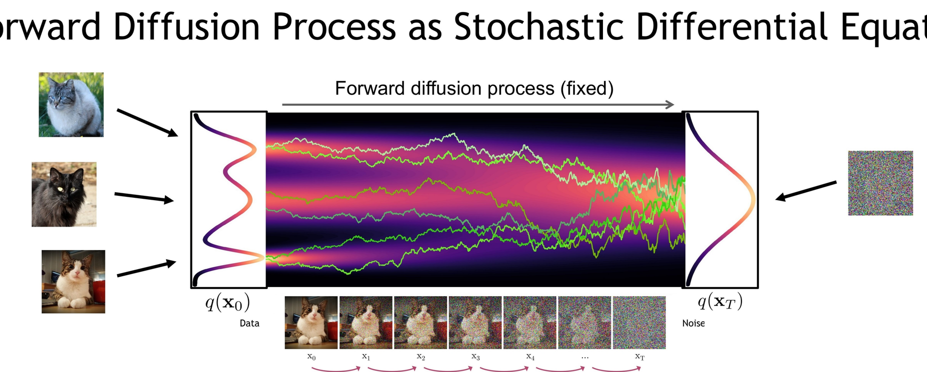

Figure 4. Visualization of Diffusion Process as SDE

References

[1] Sohl-Dickstein et al., *Deep Unsupervised Learning using Nonequilibrium Thermodynamics*, arXiv preprint arXiv:1503.03585 (2015).

[2] Ho et al., *Denoising Diffusion Probabilistic Models*, arXiv preprint arXiv:2006.11239 (2020).

[3] Song et al., *Score-Based Generative Modeling through Stochastic Differential Equations*, arXiv preprint arXiv:2011.13456 (2021).

[4] Yang et al., *Diffusion Models: A Comprehensive Survey of Methods and Applications*, ACM Transactions on Computing, Vol. 1, No. 1, pp. 1-49, October 2023.

Inpainting Applications

RePaint: DDPM-based Inpainting

RePaint [5] is an inpainting method that conditions the generation of a DDPM, using a reference image. The two main components of the algorithm are:

- **Conditioning on Known Region:** RePaint integrates known and unknown pixels during the diffusion process, using pre-trained DDPMs to guide the reconstruction. This robust approach is independent of the mask type, distinguishing itself from other methods that are applicable only on certain masks.
- **Resampling:** A strategic resampling process enhances semantic correctness in the inpainted image. By diffusing the output back to the input, RePaint improves the harmony and quality of the results.

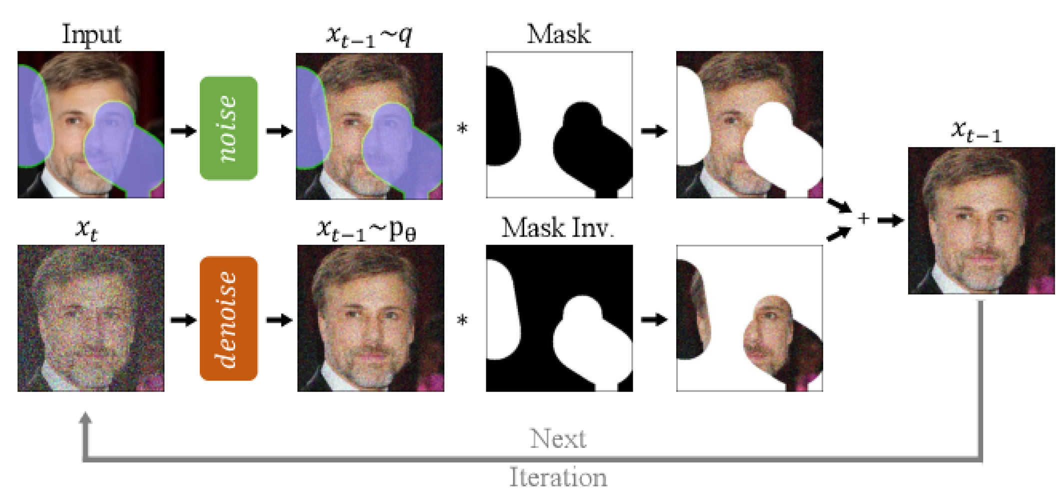

Figure 5. At each step, RePaint samples from the input the known region (top), while the inpainted section is sampled from the output of the DDPM (bottom).

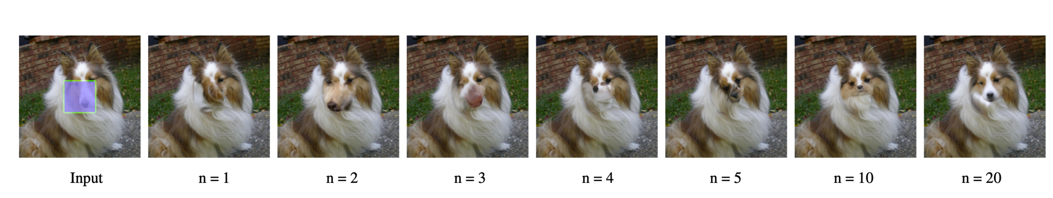

Figure 6. More resampling steps in the RePaint method improve image harmony.

Figure 7. RePaint Method Results.

CoPaint: Coherent Image Inpainting

CoPAINT [6] is an advanced image inpainting algorithm that addresses the incoherence challenges faced by inpainting models, such as RePaint. With an additional condition for a perfect match between the inpainted image and the revealed part of the reference, CoPAINT utilizes a denoising successive correction approach within a Bayesian framework. Overall, the method formulates a practical method for effective image inpainting, achieving coherence without violating constraints and without modifying pre-trained models.

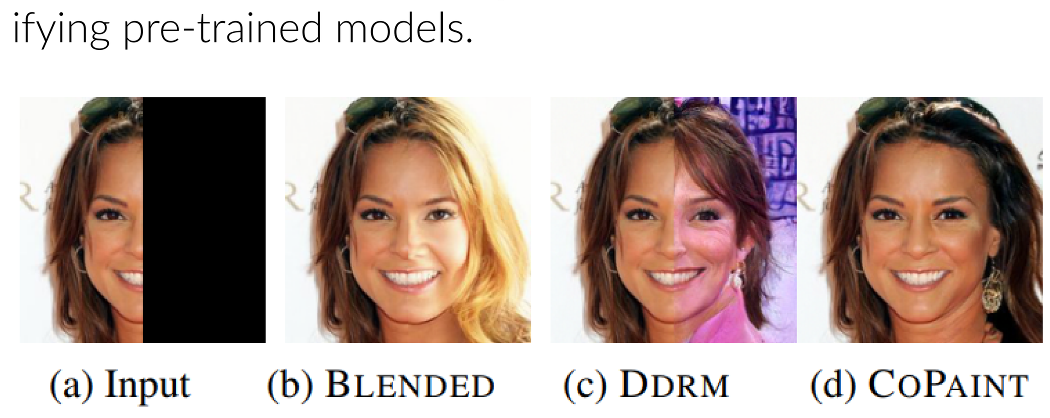

Figure 8. CoPAINT algorithm results comparison with other Inpainting models.

Furthermore, an improvement of the algorithm incorporates the "Time Travel" feature, revisiting previous denoising steps to enhance internal consistency.

References

[5] Lugmayr et al., *RePaint: Inpainting using Denoising Diffusion Probabilistic Models*, arXiv preprint arXiv:2201.09865 (2022).

[6] Zhang et al., *Towards Coherent Image Inpainting Using Denoising Diffusion Implicit Models*, arXiv preprint arXiv:2304.03322 (2023).

Super-Resolution Applications

Latent Diffusion Models: Efficient High-Resolution Synthesis

High-resolution image synthesis using AutoRegressive Transformers faces excessive energy consumption due to complex architectures. Latent Diffusion Models [7] offer a computationally efficient solution without compromising synthesis quality for high-resolution images.

- Perceptual Compression: An Autoencoder learns an efficient latent space for the images.
- Generation: A diffusion model generates images, sampling from the latent space.

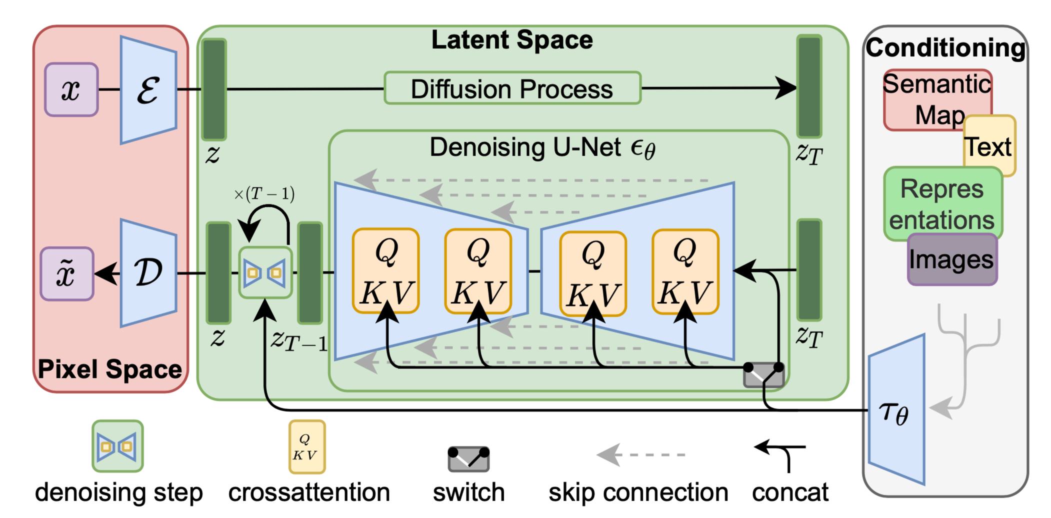

Figure 9. Latent Diffusion Model Architecture.

Partial Diffusion Models: Efficient Computational Handling

Image super-resolution models suffer from high computational demands [8].

- Assumption: Assumes convergence of latent states between low and high-resolution images, reducing computational cost.
- Efficient Denoising: Utilizing an intermediate state from the diffusion of a low-resolution image as a proxy for the high-resolution version, it reduces computational costs without sacrificing quality.
- Latent Alignment: The model gradually aligns disparities between LR and HR latent states, addressing approximation errors.

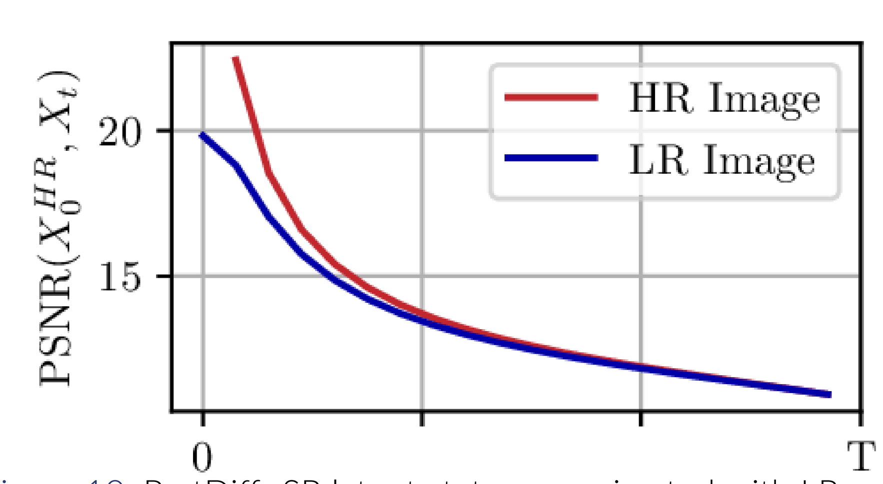

Figure 10. PartDiff: SR latent state approximated with LR proxy.

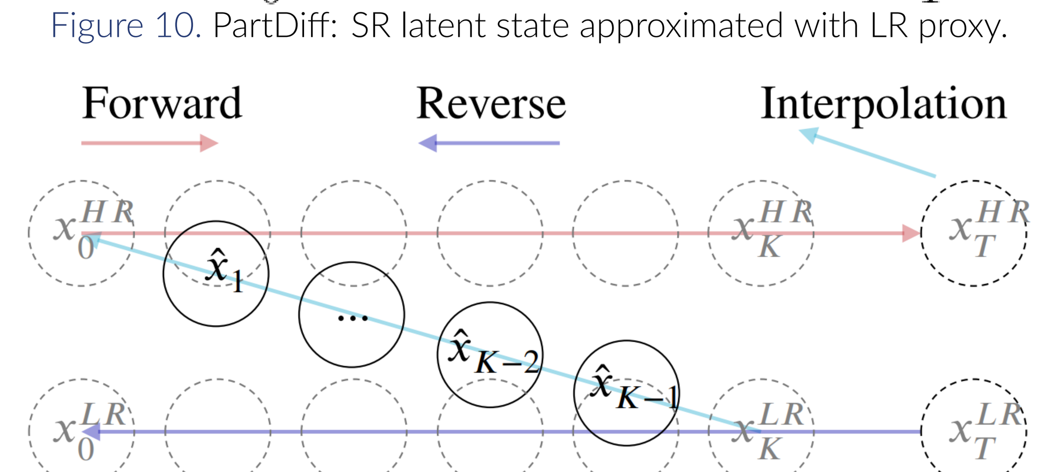

Figure 11. PartDiff: Latent alignment used for mitigation of the approximation error.

Diffusion Models for Plug-and-Play Applications: Unified Image Enhancement

DiffPIR [9] is a unified approach for denoising that addresses the task-specific nature of existing methods. Utilizing Bayes' Theorem the posterior is divided into the data and prior term, the model recruits a pre-trained unconditional diffusion model and a classifier to solve image restoration problems, conditioned on any task.

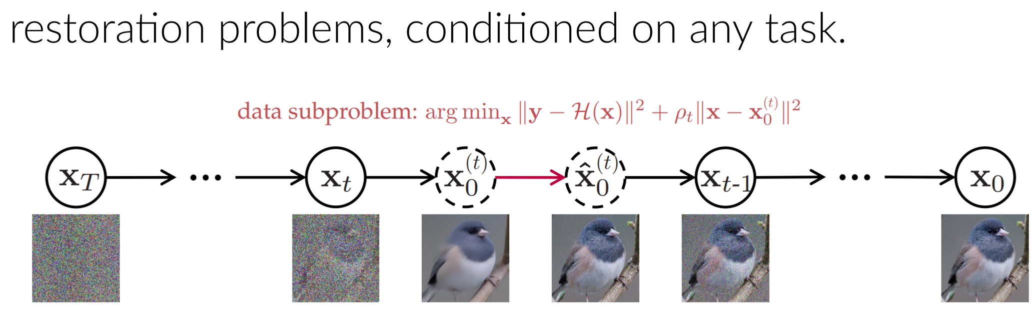

Figure 12. DiffPIR Sampling.

References

[7] Rombach et al., *High-Resolution Image Synthesis with Latent Diffusion Models*, arXiv preprint arXiv:2112.10752 (2022).

[8] Zhao et al., *PartDiff: Image Super-resolution with Partial Diffusion Models*, arXiv preprint arXiv:2307.11926 (2023).

[9] Zhu et al., *Denoising Diffusion Models for Plug-and-Play Image Restoration*, arXiv preprint arXiv:2305.08995 (2023).

el20435@mail.ntua.gr