

Denoising Diffusion Restoration Models: A Case Study

Paul Londres¹ and Lucas Murray¹

¹Ecole Normale Supérieure, Paris-Saclay

1. Introduction

Diffusion models have become a powerful framework for generative modeling and image restoration, achieving state-of-the-art performance in tasks such as super-resolution, inpainting, and deblurring. A notable advancement in this area is the Denoising Diffusion Restoration Model (DDRM), which has demonstrated strong capabilities in solving inverse problems without requiring specialized training for specific tasks.

In this study, we aim to evaluate this diffusion-based approach on a novel dataset and assess its applicability and limitations. X-ray imaging, a crucial tool in medical diagnostics, is often affected by blurring due to motion artifacts, misalignment, or hardware constraints. Restoring clear and diagnostically accurate images is essential for reliable medical assessments. While diffusion models have been widely studied in the context of photographic image restoration, their potential for medical imaging applications—particularly in X-ray deblurring—remains largely unexplored.

Our primary objective is to assess the performance of DDRM on an X-ray dataset, examining its ability to reconstruct sharp images from blurred inputs and analyzing both qualitative and quantitative aspects of the generated outputs. Additionally, we compare DDRM with an alternative approach based on conditional diffusion models, applying both methods to the same dataset. Through this work, we seek to determine the suitability of diffusion-based techniques for X-ray image deblurring and gain deeper insights into their potential applications in medical imaging. All the code is available in ¹

2. Denoising Diffusion Probabilistic Models

Generative models have demonstrated significant success in solving inverse problems, often surpassing conventional structural constraints such as sparsity. Among these models, diffusion models learn a probability distribution $p_\theta(\mathbf{x})$ that approximates the true data distribution based on given samples. In particular, they have shown remarkable generative capabilities in image modeling.

Diffusion models define a generative process as a Markov chain that progressively transforms a noise variable \mathbf{x}_T into a data sample \mathbf{x}_0 through a sequence of intermediate states:

$$p_\theta(\mathbf{x}_{0:T}) = p_\theta^{(T)}(\mathbf{x}_T) \prod_{t=0}^{T-1} p_\theta^{(t)}(\mathbf{x}_t | \mathbf{x}_{t+1}). \quad (1)$$

After generating $\mathbf{x}_{0:T}$, only \mathbf{x}_0 is retained as the final output of the model.

To train a diffusion model, a fixed factorized variational inference distribution is introduced:

$$q(\mathbf{x}_{1:T} | \mathbf{x}_0) = q^{(T)}(\mathbf{x}_T | \mathbf{x}_0) \prod_{t=0}^{T-1} q^{(t)}(\mathbf{x}_t | \mathbf{x}_{t+1}, \mathbf{x}_0), \quad (2)$$

Which is used to optimize the evidence lower bound (ELBO) on the maximum likelihood objective.

A key characteristic of diffusion models is that both $p_\theta^{(t)}$ and $q^{(t)}$ are parameterized as conditional Gaussian distributions for all $t < T$, while $q(\mathbf{x}_t | \mathbf{x}_0)$ also remains Gaussian with a known mean and covariance. This allows \mathbf{x}_t to be interpreted as a noisy version of \mathbf{x}_0 , obtained by adding Gaussian noise.

Consequently, the ELBO objective simplifies to a denoising autoencoder loss, formulated as:

$$\sum_{t=1}^T \gamma_t \mathbb{E}_{(\mathbf{x}_0, \mathbf{x}_t) \sim q(\mathbf{x}_0) q(\mathbf{x}_t | \mathbf{x}_0)} \left[\|\mathbf{x}_0 - f_\theta^{(t)}(\mathbf{x}_t)\|_2^2 \right], \quad (3)$$

¹<https://github.com/lucasmurray97/gm-project>

Where $f_\theta^{(t)}$ is a neural network parameterized by θ that aims to reconstruct a noiseless observation from a corrupted input \mathbf{x}_t . The weighting coefficients γ_t depend on the inference distribution $q(\mathbf{x}_{1:T}|\mathbf{x}_0)$.

2.1 Conditional Denoising Diffusion Models

Conditional denoising diffusion models extend the standard diffusion model framework by incorporating additional conditioning information, such as a low-resolution or degraded image, to guide the generative process toward a desired target image. Unlike unconditional diffusion models, which learn to generate data from pure noise, conditional models refine an initial noisy input into a structured output that aligns with given conditions.

The forward diffusion process in conditional models is exactly as defined in the previous section in equation (2). The reverse diffusion process is defined as a Markovian process that reconstructs an image by iteratively removing noise while conditioning on an auxiliary input y , such as a degraded or low-resolution image. This process is formulated as:

$$p_\theta(x_{0:T}|y) = p_\theta(x_T) \prod_{t=1}^T p_\theta(x_{t-1}|x_t, y), \quad (4)$$

where $p(x_T)$ is assumed to be an isotropic Gaussian prior, and each reverse step is parameterized by a learned conditional distribution:

$$p_\theta(x_{t-1}|x_t, y) = \mathcal{N}(x_{t-1}|\mu_\theta(y, x_t, \gamma_t), \sigma^2 I). \quad (5)$$

The mean $\mu_\theta(y, x_t, \gamma_t)$ is modeled by a neural network that estimates the denoised image at step t , incorporating the conditioning input y . The variance σ^2 is often set as a learned or fixed parameter, controlling the stochasticity of the process. Once trained, inference is performed by iteratively denoising the input using a reverse Markovian process:

$$p_\theta(x_{0:T}|y) = p_\theta(x_T) \prod_{t=1}^T p_\theta(x_{t-1}|x_t, y), \quad (6)$$

where the learned conditionals help reconstruct x_0 by progressively refining x_t . Given a noisy target image \tilde{x} , the model produces an estimate for the clean image as:

$$\hat{x}_0 = \frac{1}{\sqrt{\gamma_t}} \left(x_t - \sqrt{1 - \gamma_t} f_\theta(y, x_t, \gamma_t) \right). \quad (7)$$

By leveraging conditional information, this approach enables high-fidelity image generation and restoration, making it suitable for tasks such as super-resolution, inpainting, and deblurring.

3. Denoising Diffusion Restoration Models

Denoising Diffusion Restoration Models (DDRM) provide an effective unsupervised approach for solving linear inverse problems. Formally, a linear inverse problem can be written as follows:

$$y = Hx + z, \quad (8)$$

Where the goal is to recover the original signal $x \in \mathbb{R}^n$ from degraded observations $y \in \mathbb{R}^m$, where $H \in \mathbb{R}^{m \times n}$ is a known linear degradation operator, and $z \sim \mathcal{N}(0, \sigma_y^2 I)$ is an independent additive Gaussian noise with known σ_y .

One can represent the underlying distribution of the data $p_\theta(x)$ through a generative model, from which given y and H , the posterior over the signal would be $p_\theta(x|y) \propto p_\theta(x)p_\theta(y|x)$. The proposed

approach aims at reconstructing this posterior through a prior on the data distribution that can be learned on an unsupervised manner.

DDRM employs a pre-trained generative diffusion model, based on a Markov chain conditioned on y , defined as:

$$p_\theta(x_{0:T}|y) = p_\theta^{(T)}(x_T|y) \prod_{t=0}^{T-1} p_\theta^{(t)}(x_t|x_{t+1}, y). \quad (9)$$

Diffusion is performed in the spectral space, the motivation behind this transformation comes from the notion that doing so ties the noise corresponding to the measurements y with that of the diffusion $x_{1:T}$, improving reconstructions. To do this, DDRM leverages the singular value decomposition (SVD) of the degradation operator:

$$H = U\Sigma V^\top, \quad (10)$$

Where $U \in \mathbb{R}^{m \times m}$, $V \in \mathbb{R}^{n \times n}$ are orthogonal matrices, and $\Sigma \in \mathbb{R}^{m \times n}$ is a diagonal matrix containing singular values in decreasing order. The diffusion operations take place in this spectral space via:

$$\bar{x}_t = V^\top x_t, \quad \bar{y} = \Sigma^\dagger U^\top y, \quad (11)$$

Where Σ^\dagger is the Moore-Penrose pseudo-inverse of Σ . The variational distribution in the spectral space is defined as:

$$q^{(T)}(\bar{x}_T^{(i)}|x_0, y) = \begin{cases} \mathcal{N}(\bar{y}^{(i)}, \sigma_T^2 - \frac{\sigma_y^2}{s_i^2}), & \text{if } s_i > 0, \\ \mathcal{N}(\bar{x}_0^{(i)}, \sigma_T^2), & \text{if } s_i = 0, \end{cases} \quad (12)$$

With s_i the singular values and with successive transitions:

$$q^{(t)}(\bar{x}_t^{(i)}|x_{t+1}, x_0, y) = \begin{cases} \mathcal{N}(\bar{x}_0^{(i)} + \sqrt{1 - \eta^2} \frac{\bar{x}_{t+1}^{(i)} - \bar{x}_0^{(i)}}{\sigma_{t+1}}, \eta^2 \sigma_t^2), & \text{if } s_i = 0, \\ \mathcal{N}(\bar{x}_0^{(i)} + \sqrt{1 - \eta^2} \frac{\bar{y}^{(i)} - \bar{x}_0^{(i)}}{\sigma_y/s_i}, \eta^2 \sigma_t^2), & \text{if } \sigma_t < \frac{\sigma_y}{s_i}, \\ \mathcal{N}((1 - \eta_b)\bar{x}_0^{(i)} + \eta_b \bar{y}^{(i)}, \sigma_t^2 - \frac{\sigma_y^2}{s_i^2} \eta_b^2), & \text{if } \sigma_t \geq \frac{\sigma_y}{s_i}, \end{cases} \quad (13)$$

Where η and η_b are hyperparameters.

The conditional generative model is then defined by replacing $\bar{x}_0^{(i)}$ with its prediction $\bar{x}_{\theta,t}^{(i)} = V^\top x_{\theta,t}^{(i)}$, where this last term comes from the pre-trained model:

$$p_\theta^{(T)}(\bar{\mathbf{x}}_T^{(i)}|\mathbf{y}) = \begin{cases} \mathcal{N}(\bar{y}^{(i)}, \sigma_T^2 - \frac{\sigma_y^2}{s_i^2}) & \text{if } s_i > 0 \\ \mathcal{N}(0, \sigma_T^2) & \text{if } s_i = 0 \end{cases} \quad (14)$$

$$p_\theta^{(t)}(\bar{\mathbf{x}}_t^{(i)}|\mathbf{x}_{t+1}, \mathbf{y}) = \begin{cases} \mathcal{N}(\bar{\mathbf{x}}_{\theta,t}^{(i)} + \sqrt{1 - \eta^2} \sigma_t \frac{\bar{\mathbf{x}}_{t+1}^{(i)} - \bar{\mathbf{x}}_{\theta,t}^{(i)}}{\sigma_{t+1}}, \eta^2 \sigma_t^2) & \text{if } s_i = 0 \\ \mathcal{N}(\bar{\mathbf{x}}_{\theta,t}^{(i)} + \sqrt{1 - \eta^2} \sigma_t \frac{\bar{y}^{(i)} - \bar{\mathbf{x}}_{\theta,t}^{(i)}}{\sigma_y/s_i}, \eta^2 \sigma_t^2) & \text{if } \sigma_t < \frac{\sigma_y}{s_i} \\ \mathcal{N}((1 - \eta_b)\bar{\mathbf{x}}_{\theta,t}^{(i)} + \eta_b \bar{y}^{(i)}, \sigma_t^2 - \frac{\sigma_y^2}{s_i^2} \eta_b^2) & \text{if } \sigma_t \geq \frac{\sigma_y}{s_i} \end{cases} \quad (15)$$

This mathematical formulation allows full exploitation of pre-trained diffusion models while ensuring precise restoration faithful to the initial observation y . It's important to highlight the following fact: the generative model $p_\theta^{(t)}$ is not learned, this would defeat the purpose of generating a model that is task independent, we can thus use a general pre-trained diffusion process and can apply it to any inverse problems such as inpainting, deblurring, super-resolution etc.

4. Case Study:

As previously mentioned, our goal is to assess the performance of the proposed model on an X-ray dataset, which presents both clinical relevance and suitability for evaluating the effectiveness of diffusion-based deblurring techniques.

4.1 Dataset

The NIH Chest X-ray Dataset [13], also known as ChestX-ray14, is a comprehensive collection of 112k frontal-view X-ray images from 30,805 unique patients, compiled by the National Institutes of Health Clinical Center. The dataset encompasses images collected between 1992 and 2015, each labeled with up to 14 thoracic disease categories identified through natural language processing of corresponding radiology reports.

The dataset is partitioned into training and testing subsets: around 86k and 25k images correspondingly each. This dataset has been instrumental in advancing research in medical imaging, particularly in developing and benchmarking algorithms for automated disease detection and localization in chest radiographs. For this application, we selected a subset of the full dataset due to computational constraints. The chosen dataset consists of 10k images, divided into training, validation, and testing sets with 8k, 1k, and 1k images, respectively. We downsampled the images to 64*64 resolution because of computational limitations.

4.2 Isotropic Gaussian Deblurring in X-ray Imaging

The process of X-ray image formation inherently introduces blur due to factors such as focal spot size [11], detector scattering [3], and motion artifacts[5]. These sources of degradation are typically isotropic [9, 1]. To effectively model and reverse this degradation, we employ an isotropic Gaussian blurring model.

4.3 Isotropic Deblurring as an Inverse Problem

The blurring process in this work is modeled as an isotropic Gaussian blur, where the observed image y is related to the sharp image x via the blurring operator H :

$$y = Hx + \eta, \quad (16)$$

Where H is a combination of two identical separable 1D convolution matrices, H_x , representing horizontal and vertical blurring, respectively:

$$H = H_x H_x. \quad (17)$$

Each matrix H_x is constructed using a discrete 1D Gaussian kernel:

$$(H_x)_{mn} = \frac{1}{Z} \exp\left(-\frac{(m-n)^2}{2\sigma^2}\right), \quad (18)$$

Where Z is a normalization factor ensuring that H_x sums to 1.

Applying the blurring transformation to the image x is performed sequentially:

$$y = H_x(H_x x). \quad (19)$$

4.3.1 Singular Value Decomposition (SVD) for Efficient Inversion

To efficiently use the DDRM algorithm, we derive the Singular Value Decomposition (SVD) for this inverse problem.

Since the blur is separable and isotropic, the SVD of H_x is:

$$H_x = U_x \Sigma_x V_x^T. \quad (20)$$

Thus, the full system becomes:

$$H = (U_x \Sigma_x V_x^T)(U_x \Sigma_x V_x^T)^T. \quad (21)$$

Which simplifies to:

$$H = U_x \Sigma_x^2 U_x^T. \quad (22)$$

To obtain an estimate of the original image x , we apply the pseudoinverse of H :

$$\hat{x} = H^\dagger y. \quad (23)$$

Using the SVD decomposition, we define the pseudoinverse as:

$$H^\dagger = U_x \Sigma_x^{\dagger 2} U_x^T. \quad (24)$$

Where $\Sigma_x^{\dagger 2}$ is the diagonal matrix of inverse squared singular values:

$$(\Sigma_x^{\dagger 2})_{j,j} = \begin{cases} \frac{1}{\sigma_j^2}, & \text{if } \sigma_j^2 > \epsilon, \\ 0, & \text{otherwise.} \end{cases} \quad (25)$$

Here, $\epsilon = 3 \times 10^{-3}$ is a small threshold to prevent numerical instability from small singular values.

Thus, the final solution for deblurring is:

$$\hat{x} = U_x \Sigma_x^{\dagger 2} U_x^T y. \quad (26)$$

This formulation ensures a stable numerical inversion while suppressing artifacts caused by near-zero singular values. In our case, we set $\sigma = 2$, based on [6].

4.4 Unconditioned Diffusion Model

For the pre-trained diffusion model, we selected the approach proposed in [4], which consolidates several recent advancements at the time in diffusion models into a single framework.

A significant enhancement in this model is the parametrization of the variance schedule, replacing the traditional fixed variance approach. This adjustment improves sample quality while reducing the number of inference steps required. The model achieves this through a dual-headed architecture that simultaneously estimates both the variance $\Sigma_\theta(x_t, t)$ and the noise $\epsilon_\theta(x_t, t)$. Training is conducted using a hybrid objective function:

$$L = L_{\text{simple}} + \lambda L_{\text{vlb}}, \quad (27)$$

Where the loss function in Equation (3) is replaced by:

$$L_{\text{simple}} = \mathbb{E}_{t, x_0, \epsilon} \left[\|\epsilon - \epsilon_\theta(x_t, t)\|^2 \right], \quad (28)$$

Here L_{vlb} , as defined in [8], represents the variational lower bound:

$$L_{\text{vlb}} := -\log p_\theta(x_0|x_1) + \sum_{t=1}^{T-1} D_{KL}(q(x_{t-1}|x_t, x_0) \parallel p_\theta(x_{t-1}|x_t)) + D_{KL}(q(x_T|x_0) \parallel p(x_T)). \quad (29)$$

This formulation provides a more adaptable variance schedule, ultimately enhancing both the efficiency and quality of generated samples.

Instead of using the general sampling strategy of DDPMs, they use the one corresponding to DDIMs proposed in [12], where they turn $\epsilon_\theta(x_t, t)$ into a deterministic mapping between latents and images.

4.5 Conditioned Diffusion model

For the conditional model, we considered a U-Net architecture that takes two inputs: the noisy version of the image x_t , and the conditioning blurry image y . The model outputs the estimated noise $\epsilon_\theta(x_t, t, y)$. Following [10], we concatenated x_t and y along the channel dimension, as more complex conditioning methods were found to yield similar results.

The architecture is composed of sequential up-sampling and attention operations. More details about the architecture can be found in Section A.

As a loss function, we use the same formulation as in [10]:

$$L = \mathbb{E}_{t, x_0, \epsilon} \left[\|\epsilon - \epsilon_\theta(x_t, t, y)\|^2 \right], \quad (30)$$

Regarding the implementation, we did not rely on any existing codebase, as they required too many modifications. Instead, we implemented the model entirely from scratch.

5. Results

Results obtained for two experiments are presented in this section. First, a thorough evaluation of the DDRM method on the ChestX-Ray14 dataset is presented. Second, a short comparison against the conditional model presented in the previous subsection. To evaluate the reconstructions, we chose 4 different metrics, 2 "traditional" metrics : Peak Signal to Noise Ratio (PSNR), Structural Similarity (SSIM) and 2 "learned" metrics : Learned Perceptual Image Patch Similarity [14] (LPIPS) and Kernel Inception Distance [2] (KID).

5.1 DDRM on ChestX-Ray14

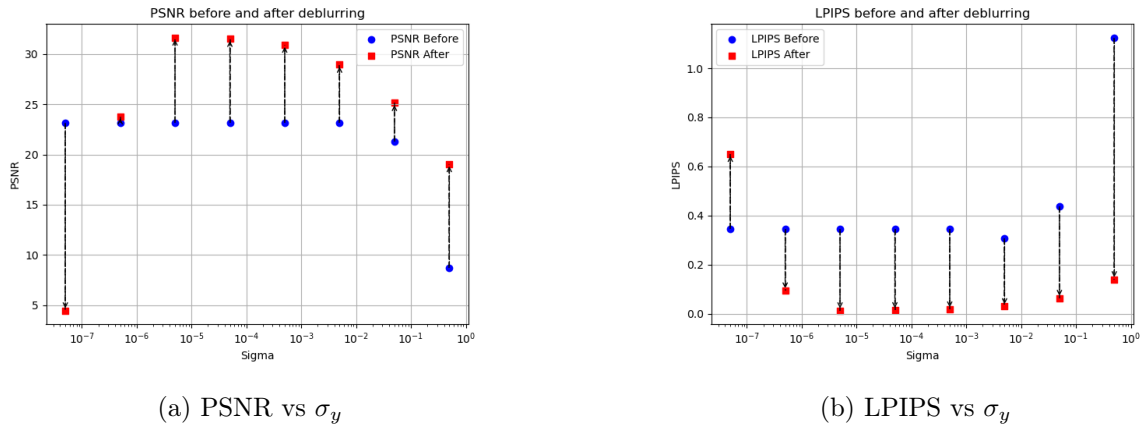


Figure 1: Values for PSNR and LPIPS before and after deblurring, depending on the selected value for σ_y . Higher PSNR indicates better reconstruction quality, while lower LPIPS values are preferred.

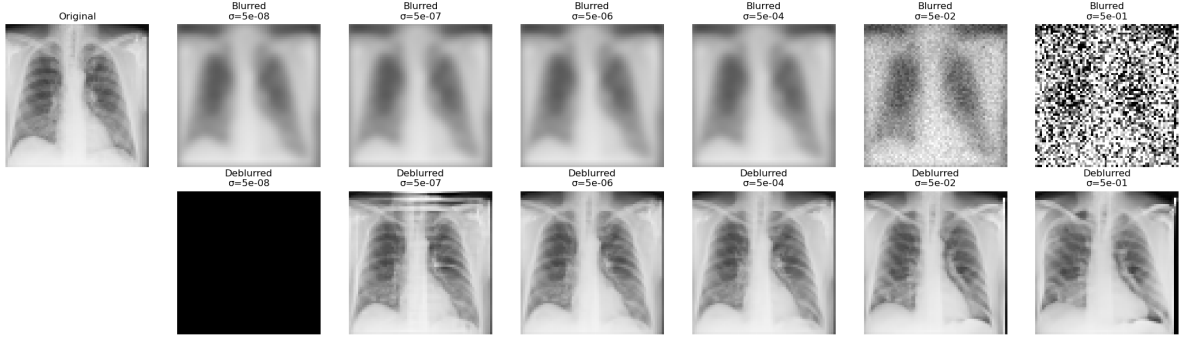


Figure 2: Reconstructions for different values of σ_y .

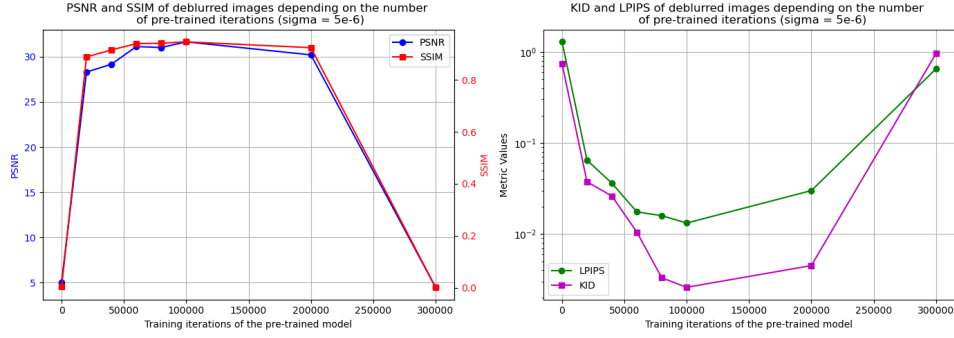


Figure 3: Reported metrics for different iterations of the training loop, with $\sigma_y = 5e^{-6}$.

5.2 Comparison with conditioned model

Metric	Unconditional ($\sigma_y = 5e^{-6}$)	Conditional
PSNR \uparrow	31.65	12.52
SSIM \uparrow	0.946	0.487
LPIPS \downarrow	0.0132	0.1089
KID \downarrow	0.00265	0.19376

Table 1: Comparison between unconditional and conditional diffusion models on image reconstruction metrics. Arrows indicate whether higher (\uparrow) or lower (\downarrow) is better.

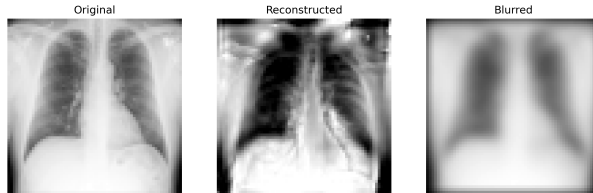


Figure 4: Reconstructions generated by the conditional diffusion model.

6. Discussion

We analyze the results obtained by the DDRM model on the deblurring task. As shown in Figure 1, the model performs strongly both in terms of quantitative metrics and visual quality. The PSNR improves significantly when compared to the blurred input image used as a baseline. It is worth noting that the PSNR of the blurred image is already relatively high, which is expected since Gaussian blurring preserves overall image structure by smoothing pixel intensities without introducing spatial distortions.

In terms of perceptual quality, the LPIPS metric—computed using a pre-trained AlexNet—decreases substantially, indicating that the reconstructions recover high-level visual features and improve human-perceived similarity. Similar improvements are observed in the SSIM and KID metrics, presented in Figure 5, reinforcing the effectiveness of the model in restoring both structural and perceptual aspects of the image.

To assess the influence of the noise associated with the linear problem defined in DDRM, we explore the effect of varying σ_y , which models the standard deviation of the Gaussian noise term. From Figure 1, we observe that the model maintains high performance over a wide range of σ_y values, particularly within $[5 \times 10^{-6}, 5 \times 10^{-3}]$. In this interval, both PSNR and LPIPS remain stable, indicating robustness to moderate noise levels in the linear problem. A significant drop in performance occurs when increasing σ_y to 5×10^{-1} and the apparition of some artifacts, likely due to the excessive degradation introduced. Nevertheless, the corresponding reconstruction in Figure 2 still resembles the original image, suggesting that the model retains some capacity for recovery even under challenging conditions. In contrast, when σ_y is decreased to 5×10^{-8} , the model fails to reconstruct the image: both PSNR and LPIPS indicate worse quality than the blurred input, and the reconstruction is nearly black. This failure is likely due to numerical instabilities in equations (13) and (15), where σ_y appears in the denominator, and may also be influenced by ill-posedness in the computation of the pseudoinverse in equation (23).

To evaluate the effectiveness of the training procedure of the pre-trained model, we analyze the evolution of key evaluation metrics of de DDRM reconstructions depending on the number of iterations of the pre-trained model. As shown in Figure 3, the model exhibits a rapid improvement in performance during the initial training phase, followed by a more gradual but steady increase until approximately 100k iterations. Beyond this point, the metrics stabilize, indicating that the model has reached an optimal performance range. However, after 200k iterations, a consistent decline in performance is observed, with a sharp drop from 300k iterations leading the model to predict plain black images. This degradation suggests potential training instability or numerical issues arising from prolonged optimization [7]. The observed trends highlight the importance of careful model selection, as excessive training can lead to a deterioration in both perceptual and structural quality.

To further extend our analysis, we compared the performance of DDRM against a conditional diffusion model applied directly to the blurred image. The results of this comparison are presented in Table 1 and Figure 4. Quantitative evaluation clearly favors DDRM, as the conditional model fails to achieve competitive results across all reported metrics. This suggests that the conditional model struggles to properly learn the underlying data distribution. However, despite its poor numerical performance, a visual inspection of the reconstructions reveals that the conditional model is still capable of producing images that resemble the original, although with significant artifacts.

These findings align with the conclusions drawn in the original DDRM paper, which demonstrated that DDRM serves as a strong alternative to conditional diffusion models, offering several advantages in solving inverse problems. Nonetheless, the validity of this comparative analysis is limited by the suboptimal performance of the conditional model. This discrepancy is likely due to the fact that a from-scratch implementation was used for the conditional model, whereas we used the official and efficient implementation for the pre-trained model and for DDRM.

References

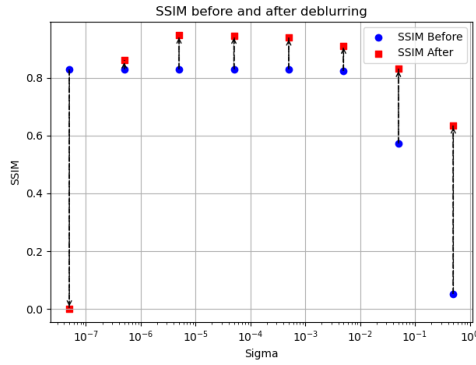
- [1] Zohair Al-Ameen, Ghazali Sulong & Md Gapar Md Johar. ‘Reducing the Gaussian blur artifact from CT medical images by employing a combination of sharpening filters and iterative deblurring algorithms’. In: *Journal of Theoretical and Applied Information Technology* 46 (Dec. 2012), pp. 31–36.
- [2] Mikołaj Bińkowski et al. *Demystifying MMD GANs*. 2021. arXiv: [1801.01401 \[stat.ML\]](https://arxiv.org/abs/1801.01401). URL: <https://arxiv.org/abs/1801.01401>.
- [3] Lutz Brügemann & Ekkehard K.E. Gerndt. ‘Detectors for X-ray diffraction and scattering: a user’s overview’. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 531.1 (2004). Proceedings of the 5th International Workshop on Radiation Imaging Detectors, pp. 292–301. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2004.06.019>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900204011842>.
- [4] Prafulla Dhariwal & Alex Nichol. ‘Diffusion Models Beat GANs on Image Synthesis’. In: *CoRR* abs/2105.05233 (2021). arXiv: [2105.05233](https://arxiv.org/abs/2105.05233). URL: <https://arxiv.org/abs/2105.05233>.
- [5] Meng Gao, Jeffrey A. Fessler & Heang-Ping Chan. ‘X-ray source motion blur modeling and deblurring with generative diffusion for digital breast tomosynthesis’. In: *Physics in Medicine and Biology* 69.11 (May 2024), p. 115003. DOI: [10.1088/1361-6560/ad40f8](https://doi.org/10.1088/1361-6560/ad40f8).
- [6] Estevao Gedraite & M. Hadad. ‘Investigation on the effect of a Gaussian Blur in image filtering and segmentation’. In: Jan. 2011, pp. 393–396. ISBN: 978-1-61284-949-2.
- [7] Zhongzhan Huang et al. *ScaleLong: Towards More Stable Training of Diffusion Model via Scaling Network Long Skip Connection*. 2023. arXiv: [2310.13545 \[cs.CV\]](https://arxiv.org/abs/2310.13545). URL: <https://arxiv.org/abs/2310.13545>.
- [8] Alex Nichol & Prafulla Dhariwal. ‘Improved Denoising Diffusion Probabilistic Models’. In: *CoRR* abs/2102.09672 (2021). arXiv: [2102.09672](https://arxiv.org/abs/2102.09672). URL: <https://arxiv.org/abs/2102.09672>.
- [9] Michael Osadebey et al. ‘Blind Blur Assessment of MRI Images Using Parallel Multi-Scale Difference of Gaussian Filters’. In: *BioMedical Engineering OnLine* 17 (June 2018), p. 76. DOI: [10.1186/s12938-018-0514-4](https://doi.org/10.1186/s12938-018-0514-4).
- [10] Chitwan Saharia et al. *Image Super-Resolution via Iterative Refinement*. 2021. arXiv: [2104.07636 \[eess.IV\]](https://arxiv.org/abs/2104.07636). URL: <https://arxiv.org/abs/2104.07636>.
- [11] s.V. Setlur Nagesh et al. ‘Focal Spot Deblurring for High Resolution Direct Conversion X-ray Detectors’. In: vol. 9783. Mar. 2016, 97833R. DOI: [10.1117/12.2216916](https://doi.org/10.1117/12.2216916).
- [12] Jiaming Song, Chenlin Meng & Stefano Ermon. *Denoising Diffusion Implicit Models*. 2022. arXiv: [2010.02502 \[cs.LG\]](https://arxiv.org/abs/2010.02502). URL: <https://arxiv.org/abs/2010.02502>.
- [13] Xiaosong Wang et al. ‘ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases’. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, July 2017, pp. 3462–3471. DOI: [10.1109/cvpr.2017.369](https://doi.org/10.1109/cvpr.2017.369). URL: <http://dx.doi.org/10.1109/CVPR.2017.369>.
- [14] Richard Zhang et al. *The Unreasonable Effectiveness of Deep Features as a Perceptual Metric*. 2018. arXiv: [1801.03924 \[cs.CV\]](https://arxiv.org/abs/1801.03924). URL: <https://arxiv.org/abs/1801.03924>.

A. Model Architectures

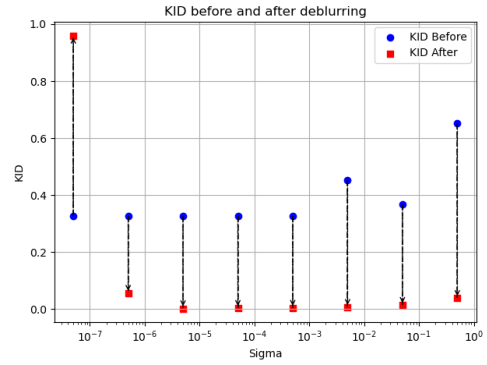
Table 2: Summary of Training Configuration and U-Net Architecture Parameters for the conditional model.

Parameter	Value
Training Configuration	
Image size	64
Train batch size	8
Eval batch size	4
Number of epochs	20
Gradient accumulation steps	1
Learning rate	2×10^{-4}
Learning rate warmup steps	500
Time steps	1000
U-Net Architecture	
Input channels	1
Output channels	1
Image resolution (sample_size)	64
Layers per block	2
Block output channels	(128, 256, 512, 512)
Down block types	DownBlock2D, AttnDownBlock2D, DownBlock2D, AttnDownBlock2D
Up block types	UpBlock2D, AttnUpBlock2D, UpBlock2D, UpBlock2D

B. Additional Metrics



(a) SSIM vs σ_y



(a) KID vs σ_y

Figure 5: Values for SSIM and KID before and after deblurring, depending on the selected value for σ_y . Higher SSIM indicates better reconstruction quality, while lower KID values are preferred.

C. Reconstructions

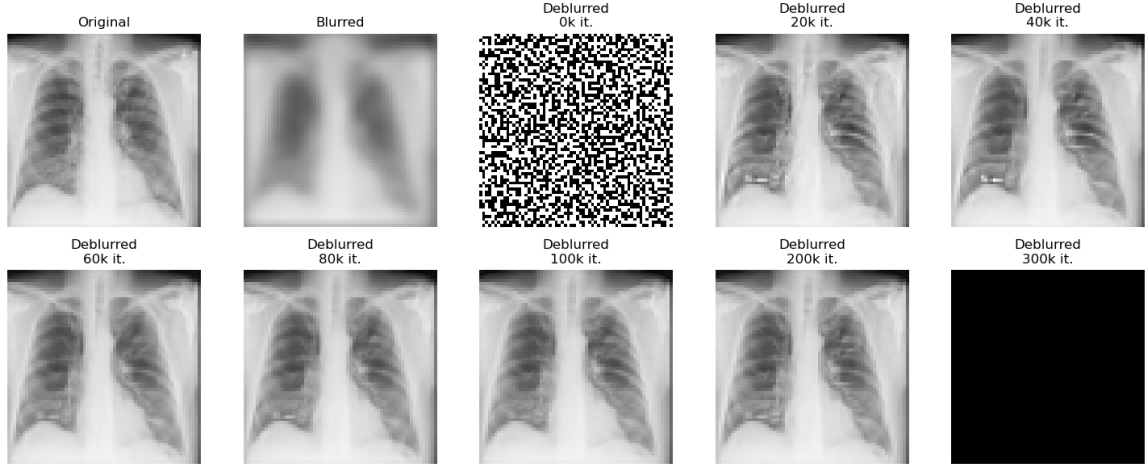


Figure 6: DDRM reconstructions varying the number of iterations of the pre-trained model

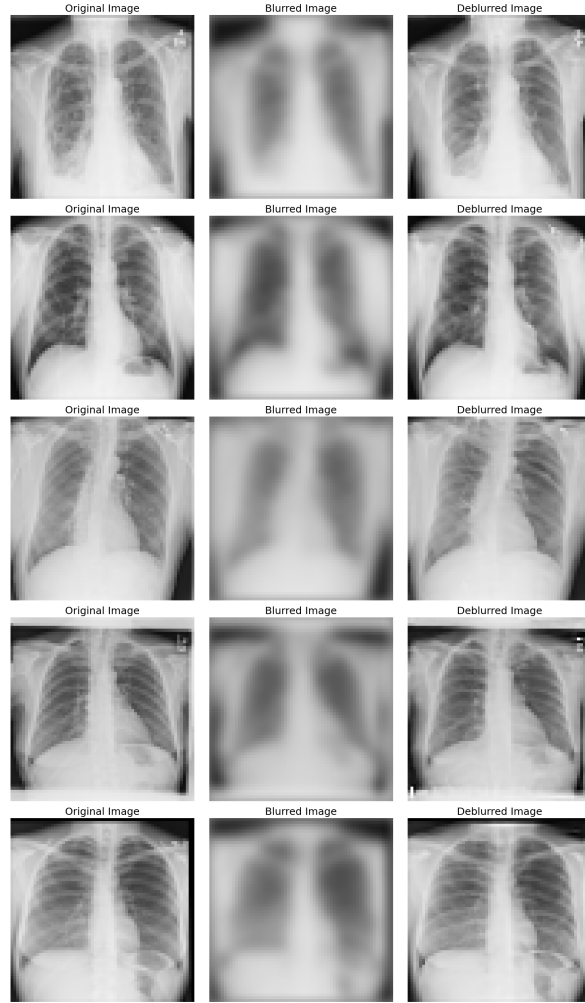


Figure 7: DDRM reconstructions for several images of the dataset with $\sigma = 5e^{-6}$ and 100k iterations on the pre-trained model