Cold Diffusion: Inverting Arbitrary Image Transforms Without Noise

Christopher Cheng, Marc Davila, Ryan Lee, Benjamin Tang, Oscar Wang

Github: https://github.com/MarcDavila1022/DL_Project

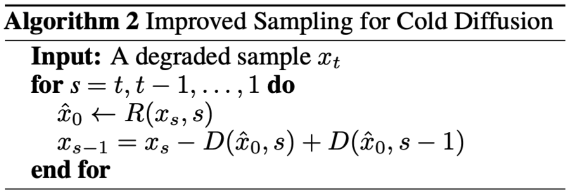
1. Introduction

Diffusion models are powerful generative models that can produce high-quality images. They work by progressively adding Gaussian noise to images, then training a neural network to reverse this process.

But what if noise wasn't necessary? In “Cold Diffusion: Inverting Arbitrary Image Transforms Without Noise”, Bansal et al. propose replacing random noise with deterministic transformations like pixelation, blurring, or masking. Just as in hot diffusion, restoration networks learn to invert these transformations, challenging the assumption that randomness is essential.

2. Chosen Result

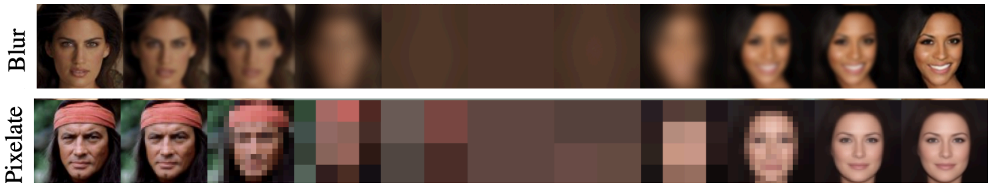
Because of the removal of Gaussian noise, the naive sampling algorithm used in hot diffusion no longer works. Thus, the authors created **Algorithm 2**, which is mathematically proven to produce high-quality reconstructions even when the restoration operator *R* fails to perfectly invert the degradation operator *D*.



Algorithm 2 Improved Sampling for Cold Diffusion

```
Input: A degraded sample $x_t$
for $s = t, t-1, \dots, 1$ do
    $\hat{x}_0 \leftarrow R(x_s, s)$
    $x_{s-1} = x_s - D(\hat{x}_0, s) + D(\hat{x}_0, s-1)$
end for
```

From here, we chose to reprove the concept of *generalized diffusion with different transformations*, by showing we can recover the original image given a degraded one (for various deterministic degradations).



3. Methodology

We focused on the *blurring* and *super-resolution* models. We also chose to forego the CelebA-HQ dataset since it was too large, and focused on the smaller MNIST and CIFAR-10 datasets.

We used Google Colab for ease in collaboration and training and to get some of our data , and eventually used their A100 GPUs (after trying T4s for a very long time). We tried to adjust various hyperparameters and degradation steps to speed up training when testing, but the following results are faithful to the paper.

In general, we use Adam with learning rate 2e-5, batch size of 32, and accumulating gradients every 2 steps. The final model is an EWMA of the trained model with decay 0.995, updated every 10 steps.

- **Deblurring:** For MNIST, we apply 40 recursive blurs with an 11x11 Gaussian kernel (sigma = 7). For CIFAR-10, we use an 11x11 kernel, adjusting the sigma at each step to 0.01 * t + 0.35.
- **Super-resolution:** We apply 3 halving steps on MNIST and CIFAR-10 to reduce images to 4x4.

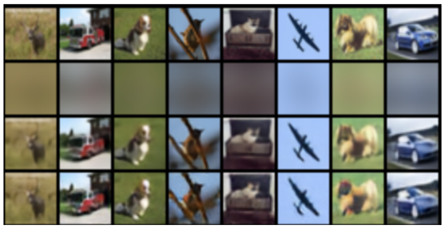
We first tried to recreate the model on our own, but we noticed that it was very hard to know how to rebuild it without knowing what they used, so we decided to reference their model architecture in their Github. We used the standard U-net with residual blocks, self-attention, and sinusoidal positional embeddings for time as a general framework, and an additional customized block specific for each degradation method.

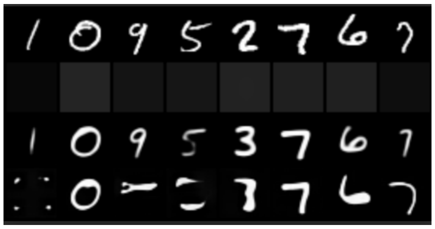
Finally, for each model, we evaluated the quality of the reconstructed images using both the direct reconstruction and improved sampling algorithm using the following metrics: Frechet Inception Distance (FID), Structural Similarity Index Measure (SSIM), and Root Mean Squared Error (RMSE).

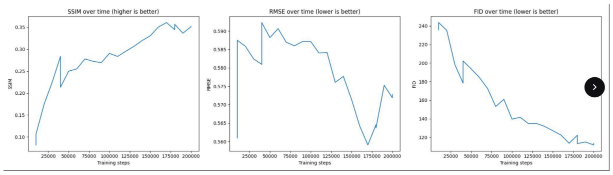
At the end, we conducted a proof of concept for *inpainting*.

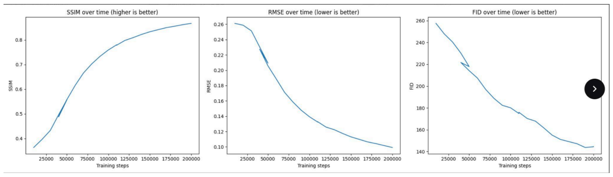
4. Results & Analysis

4a. Deblurring. Visually, even after just 200,000 steps (instead of 700,000), the resulting images look very similar to the original.



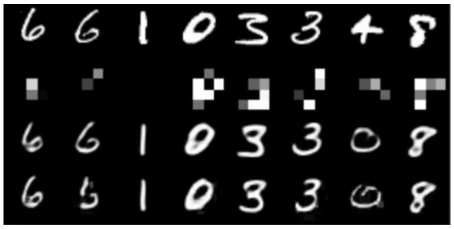


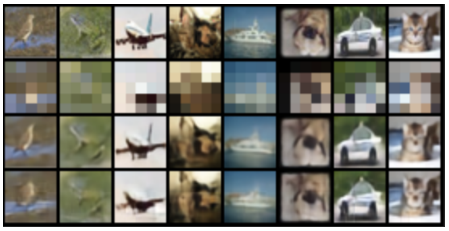




Numerically, looking at the final FID, SSIM, and RMSE scores, we have yet to achieve the scores from the paper, which was expected. However, we do see clear trends, and extrapolating them for another 500,000 steps would likely bridge the gaps. For MNIST, the scores progressed rather erratically, which we attribute to the simplicity of the task (and thus have higher variance).

4b. Super-resolution: We trained this for 200,000 steps as well (instead of 700,000).





As we can see, we are also making good progress. The trends FID, SSIM, and RMSE are similar to above, and are on track to achieving the paper's scores given the full 700,000 steps.

Our re-implemetation supports the paper's main claim. By inverting deterministic degradations like blurring and pixelating, we show that randomness is not essential for high-quality image restoration.

5. Reflections

This paper challenged the key assumptions in our understanding of randomness in diffusion models from lectures; thus, it was an extremely educational experience to dive deep into this paper and recreate the results. Working on this project further emphasized the fact that training is expensive and annoying, but very rewarding once we get results. We learned the importance of designing before coding: we tried to dive straight into the code, but got terrible results as we realized we completely oversimplified the architecture. Finally, we learned that even though images may look similar visually, their similarity scores might not be as close, which we were surprised at.

Future research directions include proving a formal framework for when cold diffusion works, particularly when it works better than hot diffusion. Hybrid models combining hot and cold diffusion is another avenue of research. Moreover, we can look into applying cold diffusion to other mediums, such as audio, video, or text as well. Cold diffusion also seems to have good properties that could be used for cryptographic purposes too.

6. Use of AI Disclaimer

We used ChatGPT-4o to help us generate code to plot graphs and display information.

7. References

We used the MNIST and CIFAR-10 datasets.

Bansal, A., Borgnia, E., Chu, H.-M., Li, J. S., Kazemi, H., Huang, F., Goldblum, M., Geiping, J., & Goldstein, T. (2022). Cold Diffusion: Inverting Arbitrary Image Transforms Without Noise. *arXiv preprint arXiv:2208.09392.* https://doi.org/10.48550/arXiv.2208.09392

Sun, Jennifer, et al. (2025). Week 7: Diffusion Models. https://www.cs.cornell.edu/courses/cs4782/2025sp/slides/pdf/week9_1_slides.pdf