Cold Diffusion: Inverting Arbitrary Image Transforms Without Noise

https://github.com/ramzd18/Noiseless-Transform

Akshay Aravind, Abhi Vetukuri, Ram Peddu, Nathan Palamuttam
CS 4782

May 12, 2025

1. Introduction

We reproduce the method proposed in Cold Diffusion: Inverting Arbitrary Image Transforms Without Noise by Bansal et al. [1], which proposes methods to perform diffusion without the use of stochastic noise during the degradation process. The authors demonstrate that diffusion can be performed using deterministic degradations, such as blurring, pixelation, and snowing. Despite these degradations being deterministic, it is shown that generative and reconstruction capabilities of typical diffusion models can be reproduced.

2. Chosen Result

We focus on reproducing Figure 3 in Cold Diffusion: Inverting Arbitrary Image Transforms Without Noise by Bansal et al. [1], which shows the accuracy improvement using cold diffusion on MNIST, CIFAR-10, and CelebA datasets. The paper demonstrates the use of an alternative sampling algorithm that takes advantage of deterministic degradation. The algorithm, denoted as Algorithm 2 in the paper, represents an improved sampling process for cold diffusion, which significantly outperforms the typical sampling algorithm of diffusion. This was chosen as it best illustrates the reconstructive capabilities of the algorithm in comparison to direct diffusion methods.

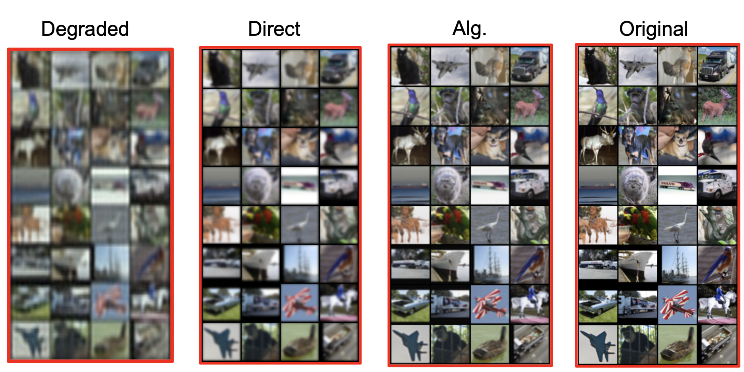

Figure 1: Reproduced version of Figure 3 from [1].

3. Methodology

We re-implemented the cold diffusion framework proposed by Bansal et al. [1], adapting it to work within limited compute resources while preserving the core structure. Our model is a time-conditioned UNet, where each timestep is embedded using sinusoidal positional encodings followed by a two-layer MLP. The encoder and decoder consist of ConvNeXt-style residual blocks, with linear attention modules to capture global dependencies. Skip connections allow information flow between corresponding encoder and decoder layers, and the final output is produced via a convolutional head.

We trained the model on the STL10 dataset, using Gaussian blur as the deterministic degradation operator. At each training step, a blur intensity corresponding to a randomly sampled timestep was applied to the image. The model was trained to reconstruct the original image using an L1 loss function and optimized with Adam. We evaluated performance using FID to measure the quality of image restoration, and we also qualitatively compared reconstructions to ground truth samples.

To manage memory and training time, we reduced the network depth and channel width from the original architecture and limited our experiments to a single degradation type. These changes allowed us to train effectively on modest hardware while still capturing the key behaviors of cold diffusion. Our implementation focused on validating the approach's effectiveness in deterministic image restoration, especially under constrained settings.

4. Results & Analysis

Table 1	**Degraded**	**Sampled**
Ours-STL10	268.05	42.75
Ours-CIFAR10	218.45	79.64
Paper-CIFAR10	298.60	80.08

Table 1: FID Scores for our Model Compared to the Original

We evaluated our cold diffusion model trained on STL10 using Gaussian blur and compared its performance to the original results reported in the paper on CIFAR-10. Our model achieved an FID of 42.75 for sampled reconstructions and 268.05 for blurred inputs, demonstrating successful recovery of image quality. Compared to the original paper's CIFAR-10 scores (FID 80.08 for sampled images, 298.60 for degraded), we see our model's FID score for CIFAR-10 is 218.45 after degradation, this makes sense due to our modified degradation methodology. Due to limited compute, we were only able to test a single degradation operator and a reduced model architecture. Furthermore, we only performed 100 blurring degradation steps compared to the paper's 300. This change in degradation steps is also part of the reason for why our model outperforms the paper's model on CIFAR-10 with a sampled FID score of 79.64. It is important to note that we used the same model, which was trained on the STL10 dataset, for the sampling (de-blurring) process for both the STL10 and CIFAR10 datasets, thus the FID score for CIFAR10 demonstrates our model's ability to generalize to other data distributions. While the quantitative results are promising, direct comparison to the original model is imperfect due to dataset differences. Nonetheless, our qualitative results also showed clear de-blurring and content recovery, reinforcing that cold diffusion's deterministic sampling strategy remains effective even under constrained settings.

5. Reflections

In re-implementing the cold diffusion framework, we quickly discovered that trying to exactly reproduce the paper's full architecture for the U-net was infeasible on our available hardware and naively scaling the network down led to poor convergence. By iteratively slimming the model, reducing depth and hidden dimensions, and then reintroducing capacity through skip connections, we were able to restore reasonable de-blurring performance on Gaussian blur, but at the cost of extended training times and limiting our experiments to a single degradation operator. This process showed us the importance of aligning project scope with compute resources and helped us build a better understanding of u-net architecture and how tweaks to it can drastically effect its restoration ability.

Looking ahead, we see two main ways to deepen this work with similar resource constraints to ours: first, starting from a publicly released diffusion model and then fine-tuning for each deterministic kernel would lower training overhead and allow parallel exploration of pixelation, swirl, and other transforms. Secondly, designing a unified, kernel-conditioned network would share capacity across multiple degradations, avoiding per-transform retraining. However, we might need a larger or more complex network to compensate. Apart from these, we could also open the framework to other domains such as audio or video data which could potentially exhibit more interesting applications of noise-free diffusion.

References

[1] A. Bansal, E. Borgnia, H-M. Chu, J. S. Li, H. Kazemi, F. Huang, M. Goldblum, J. Geiping, and T. Goldstein. Cold diffusion: Inverting arbitrary image transforms without noise. 2023. URL `https://arxiv.org/abs/2208.09392`. arXiv:2302.12254.