Reordered Noise Processing in Super-Resolution

Ryan Kwon
University of Michigan
Ann Arbor, MI, USA
ryankwon@umich.edu

Josh Wu
University of Michigan
Ann Arbor, MI, USA
joshwu@umich.edu

Euijeong Cho
University of Michigan
Ann Arbor, MI, USA
euijeong@umich.edu

Abstract

The DDPM Paper[5] effectively demonstrates the Diffusion Model's capability in denoising images with standard noise. However, denoising a single image with numerous layers of random noises and augmentations still remains as a problem. In cases where random noise such as downsampling is introduced, resulting in a degradation of both image quality and resolution, completely recovering the distorted image back to the original image remains as a challenging task. In this project, we propose a novel approach for addressing this gap by rearranging the inverse-operation (denoising) order, by first generating a super-resolution of the "multi-noised images" to encapsulate the details of the low-resolution-noised images. This approach leverages the principles of denoising diffusion models while integrating additional mechanisms to handle the complexities introduced by multi-noise scenarios.

1. Introduction

Recently, denoising diffusion models have exhibited high-quality image generation by demonstrating their capacity to remove random noises from images. However, the challenge of outputting clean images after applying combinations of different noises and augmentations (which we will call as "multi-noise" for brevity) remains even for state-of-the-art denoising models. For the simplicity of our demonstration, we assume that "downsampling" is always applied in our multi-noise.

This paper presents a novel approach by first upscaling the low-resolution multi-noised image to a 4x Super Resolution, to not only resize the image back to its original size but to also proliferate the distorted image from the multi-noise. The Super Resolution emphasizes the features of the image and the noises. We show that the Super Resolution of multi-noise images enhances the overall quality of the denoising step for the "downsampling" step, as well as the final output image. By using this two-step method, we aim to improve the resemblance of the processed images to their original forms.

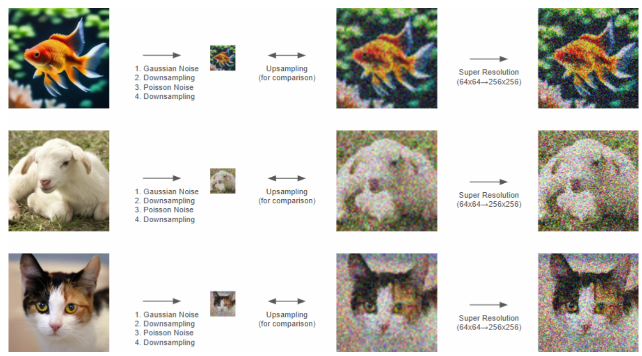

Figure 1. Comparison between the results of the resized image and Super Resolution using images with 4 layers of noises applied.

As a whole, we demonstrate that the reordered noise processing, by first using Super Resolutions to elevate the image quality before denoising other noises from the stacked layers of noises, enhances the overall quality of the images when inversely processing images.

2. Related Work

In the past years, numerous deep-learning methods for super-resolution have been discovered. Previous methods for generating Super Resolutions include SRGAN[7], ESRGAN[9], ESRGAN+[4]. These Super Resolution models are built upon the framework of generative adversarial networks (GANs), whereas our approach leverages diffusion, a state-of-the-art technique for image generation.

The Super-Resolution of noise has been researched in previous works but was not on the scale of numerous layers of noise. The present work aims to take a step further to effectively reversing the random image operations applied to the image while preserving the quality.

3. Methods

$$\text{for } k = 1, 2, \ldots, M : \quad I_k = T_k(I_{k-1}). \tag{1}$$

$$z_T \sim \mathcal{N}(0, I),$$

$$\text{for } t = T, T-1, \ldots, 1: \quad z_{t-1} = p_\theta(z_{t-1} \mid z_t, I_M). \quad (2)$$

$$\hat{I} = F(\hat{I}_{SR}) = F\big(\text{DiffusionSR}(I_M; \theta)\big), \quad (3)$$

$$\text{where } I_M = (T_M \circ T_{M-1} \circ \cdots \circ T_1)(I).$$

Our paper presents a two-step method using a super-resolution model, followed by a denoising model. We evaluate two prominent models, StabilityAI[2] and OpenAI[8]'s Diffusion Probabilistic Models (DDPM)[5], as our choices for the super-resolution models. StabilityAI's stable diffusion upscaler model produces significantly improved quality images than DDPM, but also removes all of the noise that was applied onto the input image. This was problematic as there was no use case for the secondary step which would be to input the super-resolved image into the denoising model. Therefore, we opted to use DDPM, producing super-resolution images while preserving the underlying noise applied.

3.1. Applying Noise

In this paper, we sequentially add noise and perform downscaling to generate the multi-noised images. We apply the Gaussian Noise by convoluting the Gaussian matrix and the Poisson noise by using the Wand library[3], with downscaling handled by `Image.Resampling.LANCZOS` from the Python library `Pillow`.

3.2. OpenAI Guided Diffusion

OpenAI[8]'s Guided Diffusion Upscaler model features UNet architecture for diffusion models, and improvements upon previous models, including BigGAN residual block usage, and an increase in residual connections. We utilize this model to perform an upscaled super-resolution of the multi-noised images.

4. Experiments

The experiment was conducted through hyperparameter searching from the pre-trained model by OpenAI[8] that was trained with the ImageNet[6] dataset. We set the model to have 1000 diffusion steps and 250 timestep respacing to obtain the image that best enhances the details of the image. We evaluate our Super Resolution model by comparing the Fréchet Inception Distance (FID) and Peak Signal-to-Noise Ratio (PSNR) metrics with the resized original image. process.

5. Results

Although there exist denoising models that are trained with natural noises such as the SIDD dataset[1], the models were not able to denoise images for our particular scenario.

Image\Metrics	**FID (↓)**	**PSNR (↑)**
Resized (256x256)	84.38	32.92
Ours (Super Resolution)	61.43	30.41

Table 1. Comparison of metrics for Upscaling and Super Resolution methods.

Figure 2. Final image generated through Nero AI's online denoiser

As shown in **Figure 2**, noises are not completely denoised, however, we anticipate this phenomenon to alleviate with a model trained specifically with our Super Resolution noises.

6. Future Work

For this paper, for simplicity, we only demonstrated 4 steps of image augmentations (noise1 + downsampling + noise2 + downsampling). However, ideally, given any combinations of noises and augmentations, the model should be able to give a Super Resolution, and effectively restore it back to the original image. Given the results of our PSNR scores, there is still space for more improvements through hyperparameter searching. In this paper, we employed an open-source commercial model to denoise our Super Resolution Noised image. However, future work will focus on developing a custom model specifically to address and remove such noise effectively.

7. Conclusion

We demonstrate that modifying the order when reversing image operations with multiple layers of noise results in enhanced quality. This study potentially could be reproduced with a variety of different image augmentations and noises, following with other super-resolution, and denoising models. Although our results did not match the original image's quality due to utilizing an online denoiser that was not specifically trained with our settings, we successfully demonstrated the feasibility of applying such tools for rapid image processing and identified areas for future optimization.

References

[1] Abdulla Abdelhamed, Stephen Lin, and Michael S. Brown. Toward a benchmark for smartphone camera image denoising. *IEEE Transactions on Image Processing*, 28(11):5596–5610, 2019.

[2] Stability AI. Stable diffusion x4 upscaler, 2024. Accessed: 2024-12-10.

[3] Emmanuel Arias and contributors. *Wand: Python bindings for ImageMagick*, 2024. Version 0.6.12.

[4] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *arXiv preprint arXiv:2001.08073*, 2020.

[5] Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *arXiv preprint arXiv:2006.11239*, 2020.

[6] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255. IEEE, 2009.

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *arXiv preprint arXiv:1609.04802*, 2016.

[8] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Diffusion models beat gans on image synthesis. *arXiv preprint arXiv:2105.05233*, 2021.

[9] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Tensor2tensor for neural machine translation. *arXiv preprint arXiv:1809.00219*, 2018.