Light-weight Diffusion Prior for Multi-task Real-world Image Restoration

Yining Mao
Department of Electrical Engineering
Stanford University
Stanford, PA 94305
yiningm@stanford.edu

Haijing ZHANG
Department of Electrical Engineering
Stanford University
Stanford, PA 94305
haijing@stanford.edu

1 Motivation

The emergence of diffusion models represents a notable breakthrough in generative modeling, demonstrating exceptional ability to generate high-fidelity images from textual inputs. Meanwhile, image restoration (IR), encompassing super-resolution, deblurring, denoising, inpainting, and compression, remains a vital area within low-level vision studies. Lately, there has been a growing trend of integrating diffusion models into IR tasks, yielding outcomes that surpass those of previous methodologies. Nevertheless, the application of diffusion models in IR presents its own set of challenges, including complexities in model design and concerns regarding operational efficiency. This project draws inspiration from the innovative techniques introduced in "Exploiting Diffusion Prior for Real-World Image Super-Resolution" (StableSR) by Wang et al. [2023], which skillfully employs the generative capabilities of pre-trained text-to-image diffusion models to enhance blind super-resolution (SR) tasks. The framework of StableSR is shown in Figure 1. This research showcases how a time-aware encoder, fine-tuned in conjunction with an unmodified Stable Diffusion Rombach et al. [2022] model, leads to significant restoration improvements while maintaining the original generative framework and reducing training expenses. In this project, we aim to broaden the application of StableSR across various IR tasks and to investigate the potential for a more lightweight solution.

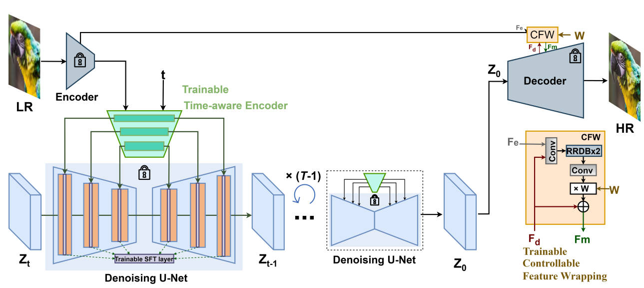

Figure 1: Framework of StableSR

2 Related Work

Image Restoration. Aiming to recover a high-quality image from its degraded observation, image restoration (IR) tasks including super-resolution (SR), deblurring, denoising, inpainting and artifacts removing, etc, has been a long-term research topic. Image restoration techniques have evolved significantly, transitioning from methods that employ predefined degradation models such as bicubic downsampling, to more advanced strategies that address real-world complexities.

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

Deep-learning Based Image Restoration. With the development of deep learning, well-designed backbones based on CNNs Krizhevsky et al. [2017] or Transformers Vaswani et al. [2017] are developed for complicated degradation. Emerging generative models in the recent decades, especially the generative adversarial network (GAN) Creswell et al. [2018], have pioneered the use of unsupervised learning and explicit degradation modeling, offering more realistic LR-HR pair synthesis and addressing blind super-resolution challenges.

Diffusion-model Based Image Restoration Diffusion models Sohl-Dickstein et al. [2015] demonstrate exceptional capability in image generation tasks. These models operate by gradually denoising a signal, making them particularly suited for image restoration tasks. Their structure and algorithmic foundations enable iterative refinement, allowing for high-fidelity image synthesis and restoration. The exploration of diffusion models for image super-resolution in StableSR Wang et al. [2023], which employs the Stable Diffusion Rombach et al. [2022] model, underscores their potential to redefine the landscape of image restoration through their iterative, noise-reduction-based processes.

3 Project Overview

Building on the foundation laid by StableSR, our project seeks to extend the exploration of diffusion models beyond super-resolution into a unified framework for comprehensive image restoration. By leveraging the versatility of diffusion models, we aim to develop a solution that not only addresses super-resolution but also dehazing, deblurring, and more image restoration tasks, within a cohesive model architecture.

Specifically, our primary objectives are as follow:

1. Adapt the StableSR architecture to facilitate multi-task learning, enabling the model to perform various image restoration tasks (deblurring, dehazing, super-resolution) directed by different prompts.
2. Curate a novel multi-task dataset by employing diverse degradation pipelines (deblurring, dehazing, super-resolution) to generate training pairs of the same image, thus enriching the model's learning environment.
3. Investigate training strategies to ascertain the efficacy of separate vs. co-training for different tasks using a shared encoder, exploring the model'sperformance across tasks.
4. Explore model efficiency optimizations to reduce the parameter using strategies such as pruning or LoRA finetuning Hu et al. [2021], aiming for a more lightweight model without compromising on restoration quality.
5. Develop an interactive, user-friendly interface for the model, facilitating easy access and usability for diverse applications and demonstrations.

4 Timeline

Week 1: Model Setup

- Review diffusion models and their image restoration applications.
- Replicate the StableSR study with the LR-HR dataset.
- Prepare multi-task dataset.

Week 2: Training and Development

- Yining explores multi-task learning adaptation.
- Haijing explores lightweight architectures.

Week 3: Evaluation and Demonstration

- Test model on various restoration tasks.
- Compile a comprehensive report of results and learnings.
- Create a user-friendly model interface or API.

References

Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin CK Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. *arXiv preprint arXiv:2305.07015*, 2023.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A Bharath. Generative adversarial networks: An overview. *IEEE signal processing magazine*, 35(1): 53–65, 2018.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pages 2256–2265. PMLR, 2015.

Edward Hu, Yelong Shen, Phil Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Lu Wang, and Weizhu Chen. LoRA: Low-Rank Adaptation of Large Language Models, 2021.