FastDiff: A Lightweight and Accelerated Diffusion Framework for Efficient Image-to-Image Generation

Jincheng Rao, Zihan Xiao

Project Overview - Purpose of FastDiff

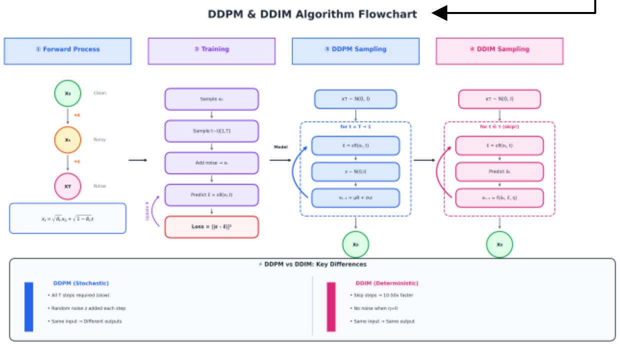
Diffusion models have become a leading approach for high-quality image generation by learning to reverse a gradual noising process. In this project, we propose FastDiff, a lightweight and accelerated diffusion framework tailored for efficient image-to-image generation. By improving both model architecture and sampling strategy, FastDiff aims to deliver strong visual quality with significantly reduced computational overhead.

Key Points:

- **Lightweight U-Net** for reduced computational cost
- **Mixed-precision training (AMP)** for 1.5–2× faster training
- **Accelerated sampling** via DDIM with fewer reverse steps
- **Efficiency-focused design** enabling faster image-to-image generation

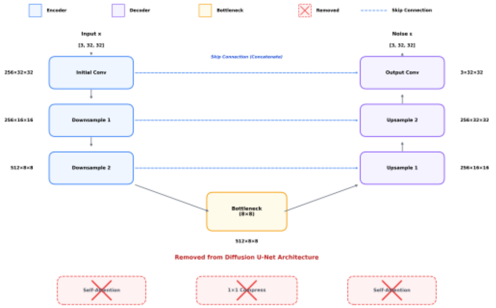
Project Objectives

- Build an FastDiff Training Pipeline
- AnalyzeQuality–Efficiency Trade-offs
- DDPM vs. DDIM
- Explore extensions



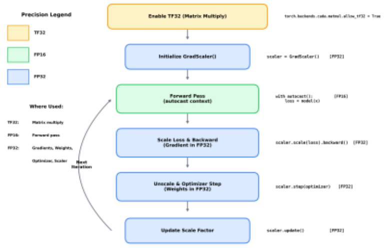
Improved Structure

Lightweight U-Net Architecture



Self-attention introduces $O(n^2)$computational cost, making it the primary bottleneck in diffusion U-Nets. Removing attention significantly reduces **FLOPs(Floating Point Operations)** and accelerates both training and sampling.

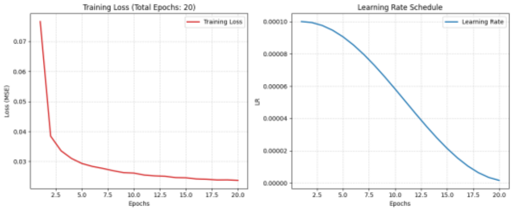
AMP (Automatic Mixed Precision) Training Flow



AMP accelerates training by combining **TF32 (TensorFloat-32)** for fast matrix operations, **FP16 (Half Precision 16-bit Floating Point)** for the forward pass, and **FP32 (Full Precision 32-bit Floating Point)** for gradient computation and optimization, enabling efficient yet numerically stable mixed-precision learning.

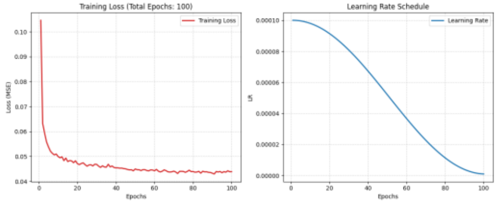
Training Behavior and Diffusion Reconstruction on MNIST and CIFAR-10

Training Loss & Learning Rate Schedule on MNIST (20 Epochs)



Forward and Reverse Diffusion Process on MNIST: The top row illustrates the gradual addition of noise to a clean digit, while the bottom row shows the DDPM reverse denoising steps recovering the original structure.

Training Loss & Learning Rate Schedule on CIFAR-10 (100 Epochs)



Forward and Reverse Diffusion Process on CIFAR-10: The top row shows the progressive corruption of a CIFAR-10 horse image, while the bottom row illustrates the reverse denoising steps reconstructing the original visual structure.

References

Ho, J., Jain, A., & Abbeel, P. (2020). "Denoising Diffusion Probabilistic Models". In: NeurIPS 33.

Nichol, A. Q., & Dhariwal, P. (2021). "Improved Denoising Diffusion Probabilistic Models". In:ICML 2021.

Dhariwal, P., & Nichol, A. (2021). Diffusion Models Beat GANs on Image Synthesis. In:NeurIPS 34.