Latent Diffusion Models for Image Super-Resolution

EE 367 Project Proposal

Chih-Ying Liu
Stanford University
ying1029@stanford.edu

1 Motivation

Image super-resolution (SR) is the process of recovering high-resolution (HR) images from low-resolution (LR) images. It is a challenging and important image processing technique. It can be widely applied to real world applications and high-quality images can benefit other computer vision tasks as well. The problem is inherently ill-posed because multiple high-resolution images may map to a single low-resolution image and the conditioning distribution of output images given the input does not conform to a simple parametric distribution.

With the rapid development of deep learning techniques, various deep learning based super-resolution models have been proposed. Previous work explored techniques ranging from simple regression-based methods with feedforward convolutional network [1] to advanced generative techniques including generative adversarial nets (GAN) ([2]), variational autoencoders (VAEs) [3] and normalizing flows (NFs) [4]. However, existing techniques often suffer from various limitations: regression-based method is computationally costly for high-resolution image generation, GANs suffer instability training and mode collapse, and NFs and VAEs yield suboptimal generation quality.

Diffusion model, which iteratively refine noisy input image into a high-quality image, has been shown to have the capability to generate high quality samples [5] and achieve impressive results in various image synthesis tasks, including image super-resolution [6, 7, 8]. In this project, we will explore diffusion models for image super-resolution with a focus on Latent Diffusion Models (LDM) [7] and compare the performance and speed between different models and inference strategies.

2 Related Work

Diffusion model [9] is a probabilistic modeling that is able to convert a simple known distribution into a target distribution and is widely used in image generation. The algorithm assumes that there is a Markov chain that iteratively adds some noise to its input in the forward diffusion process, and the diffusion model is trained to reverse this chain. SR3 [6] applies the diffusion framework to image super-resolution tasks. It adopts UNet to model a stochastic iterative denoising process and iteratively refine input images by predicting the addition noise at each step. Since diffusion models typically require a large number of refinement steps during inference, and therefore they are expensive compared to other generative techniques. SR3 determines the noise schedule used in inference, allowing a trade-off between image quality and efficiency.

The training and testing of diffusion models requires repeated calculations in the high-dimensional RGB image space. Latent diffusion Model (LDM) [7] reduces the computation by applying diffusion to the latent space of a powerful pre-trained autoencoder. It also designs a general-purpose conditioning mechanism based on cross-attention, which enables multiple-modal training. It shows success in multiple image synthesis tasks, including inpainting, class-conditional image synthesis, unconditional image generation, text-to-image synthesis, and super resolution.

The above diffusion-based super-resolution models only work with a fixed magnification. That is, we require separate models for each magnification. [6] chain multiple diffusion models with different scales to achieve high-resolution synthesis, and shows that the cascaded approach requires less refinement steps compared to a single large magnification scale model. On the other hand, the

Implicit Diffusion Model (IDM) [8] achieves image super-resolution with continuous resolution by conditioning decoding with continuous-resolution representation. For this project, we will focus on the Latent Diffusion Model [7], compare it with other diffusion models, explore the cascaded inference idea [6] and other inference design with LDMs.

3 Goal

The preliminary project goal is to train and test Latent Diffusion Model at different magnification scales. If time permits, we will also train and test SR3 to understand how diffusing in latent space or pixel space differs in performance and efficiency. We plan to compare their performance qualitatively and also quantitatively with metrics such as FID, IS, PSNR, and SSIM.

The second goal is to explore cascaded high-resolution image synthesis, where we chain LDMs of different scales together to enable high-resolution synthesis. For example, we can cascade two 4x diffusion models to get a 16x image super-resolution. We want to explore the model chain and the number of refinement step designs, and compare the total required refinement steps of the cascaded method to a single large scale diffusion model.

4 Milestone

- Week 2/19
  - Search for related work
  - Submit proposal (2/23)
  - Look into LDM and SR3 codebase
- Week 2/26
  - Set up code for LDM training and inference at different scales.
  - Run training experiments.
  - If time permits, set up code for SR3 training and inference.
- Week 3/4
  - Set up code for cascaded inference.
  - Run training and inference experiments.
- Week 3/11
  - Keep running experiments.
  - Make a presentation (3/13) and write a report (3/15).

References

[1] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks, 2015.

[2] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network, 2017.

[3] Arash Vahdat and Jan Kautz. Nvae: A deep hierarchical variational autoencoder, 2021.

[4] Diederik P. Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions, 2018.

[5] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models, 2020.

[6] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J. Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement, 2021.

[7] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models, 2022.

[8] Sicheng Gao, Xuhui Liu, Bohan Zeng, Sheng Xu, Yanjing Li, Xiaoyan Luo, Jianzhuang Liu, Xiantong Zhen, and Baochang Zhang. Implicit diffusion models for continuous super-resolution, 2023.

[9] 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.