DENOISING DIFFUSION IMPLICIT MODELS (DDIMs)

Author : Jiaming Song, Chenlin Meng, Stefano Ermon
Prof. Dr. Tushar Sandhan, Indian Institute of Technonlogy, Kanpur

Abstract

Denoising Diffusion Implicit Models (DDIM) introduce a novel, non-Markovian diffusion process for generative modeling that efficiently produces high-quality samples from complex data distributions. Unlike traditional Denoising Diffusion Probabilistic Models (DDPM), which require hundreds or thousands of diffusion steps, DDIM achieves comparable or better sample quality with fewer steps by using a deterministic sampling technique. This approach preserves the simplicity and robustness of the original diffusion framework while substantially reducing computational costs. DDIM leverages a deterministic inversion of the diffusion process, allowing control over the sampling trajectory and enabling faster image generation. The technique also offers a flexibility that facilitates more diverse applications, including image inpainting, super-resolution, and conditional generation. Overall, DDIM represents a significant advancement in generative modeling, improving both efficiency and flexibility for practical applications.

Introduction

Generative models have gained significant attention for their ability to synthesize realistic images, audio, and other forms of high-dimensional data. Among these, diffusion models, such as Denoising Diffusion Probabilistic Models (DDPM), have emerged as powerful generative methods due to their simplicity and ability to capture intricate data distributions. However, traditional DDPMs rely on a time-consuming, high-step denoising process to gradually transform random noise into realistic data, which can be computationally expensive.

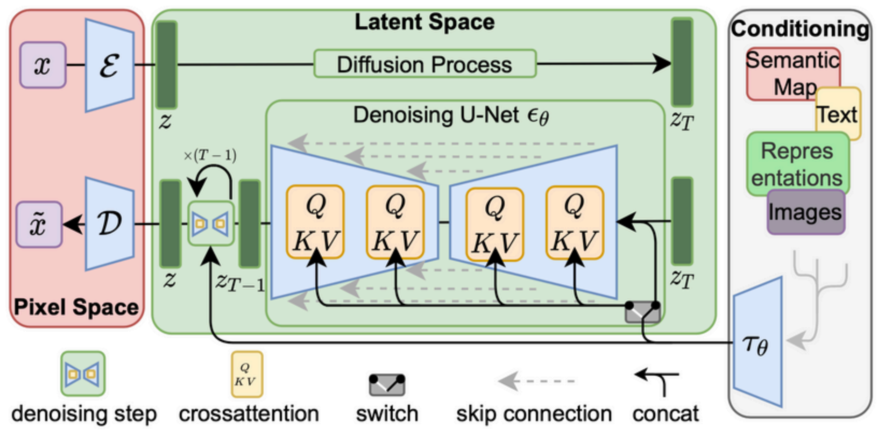
Denoising Diffusion Implicit Models (DDIM) address this limitation by proposing a deterministic, non-Markovian diffusion process that enables faster sampling without sacrificing quality. Instead of the stochastic noise used in DDPMs, DDIM utilizes a modified denoising process to efficiently sample data in a fraction of the steps required by conventional diffusion models. This faster sampling is achieved by allowing direct control over the diffusion trajectory, making DDIM a practical choice for applications requiring rapid generation, such as image synthesis, inpainting, and super-resolution.

DDIM's approach to sampling also brings new flexibility, enabling creative applications and adaptations within generative modeling. By reducing computational requirements while maintaining quality, DDIM contributes to the field of generative models by expanding the usability of diffusion-based methods in both research and real-world settings.

Methodology

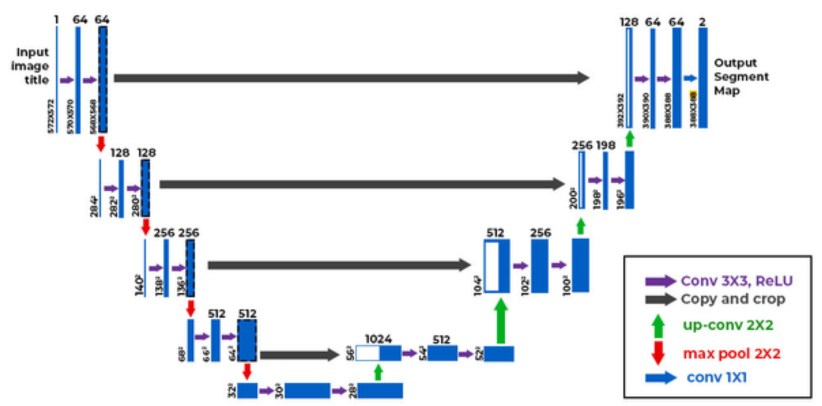
Diffusion Model Framework:

DDIM builds on the diffusion model framework, which gradually transforms a data sample into noise through a forward process and then reconstructs it using a reverse denoising process. The forward diffusion process adds noise in multiple steps, making the data distribution increasingly Gaussian. The reverse diffusion then removes noise step-by-step to generate new data samples from noise, preserving the data distribution's complexity.



UNet Architecture:

At the core of DDIM's denoising mechanism is a UNet architecture, a convolutional neural network known for its strong performance in image-to-image tasks. The UNet's encoder-decoder structure with skip connections allows it to capture both global and local features, essential for high-quality image synthesis. In DDIM, the UNet is conditioned on time step embeddings, enabling it to effectively denoise inputs based on their noise level at each step.



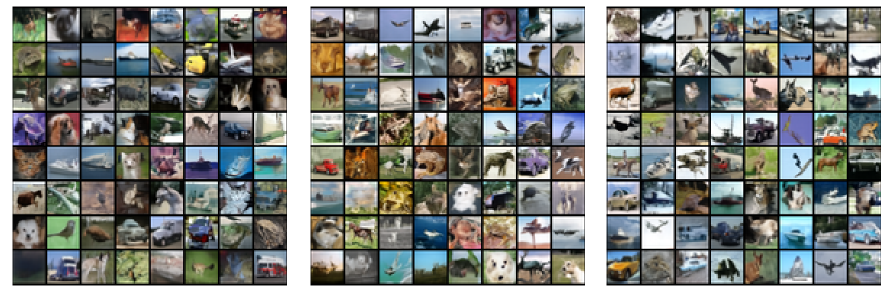
Deterministic Sampling and Adaptive Step Size:

DDIM introduces a deterministic sampling method, diverging from the stochastic noise approach used in traditional diffusion models. By controlling the denoising trajectory deterministically, DDIM enables faster sampling, achieving high-quality outputs with fewer steps. This method includes an **adaptive step size** strategy, where the model adjusts the step size dynamically based on the noise schedule, further improving efficiency and control over the generated outputs. This adaptive approach allows DDIM to achieve results with minimal steps compared to standard diffusion models.

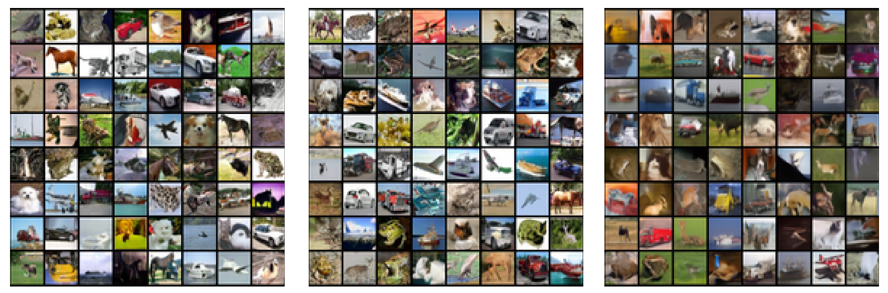
Results

DDIM Sampling Images for Different Datasets

Cifar10 64-dimension



Cifar10 128-dimesnsion



CelebA HQ

Result Metrics

Dataset	FID score	Reconstruction loss
Cifar10 64-dimension	11.81	0.03-0.04
Cifar10 128-dimension	8.31	0.02-0.03
CelebA-HQ	11.97	0.03-0.04

Inference

- Good image quality and acceptable loss with 64-dimension embedding
- Improved perceptual quality and accuracy with higher dimension
- Comparable FID and reconstruction loss to CIFAR10 (64-dim), demonstrating model generalizability across complex datasets.

Conclusion

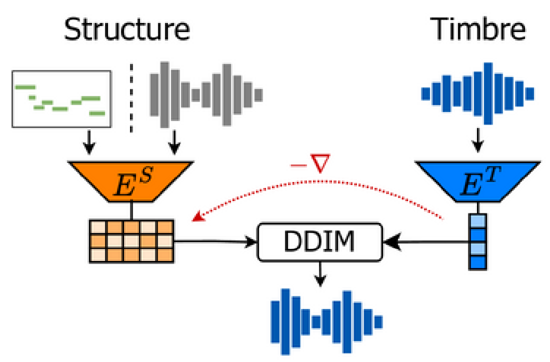
Denoising Diffusion Implicit Models offer a more revolutionary way in which there is generative modeling capability-the ability to achieve highquality sampling at fewer steps than does diffusion. DDIM realizes fast synthesis of very high-details images with fidelity using deterministic sampling and reduced numbers of needed denoising steps. A UNet-based architecture allows DDIM to capture good global and local features, so the quality of the output can adapt to varying noise ranges and even intricate patterns in the data.

This efficiency not only accelerates the generative process but also opens doors to applications in fields requiring real-time performance, such as interactive image synthesis, video frame interpolation, and image inpainting. Moreover, deterministic sampling with DDIM can be applied to conditional generation and controlled image editing where users guide the generation process according to specific requirements.

Summarily, DDIM stands for a major innovation that guarantees diffusion-based generative models with a robust, efficient framework that balances the computational demands by perfecting output quality. Absolutely makes diffusion models more accessible and practical for real-world applications-really an important step forward in the baroque AI-driven creativity and image formation.

Applications

Image Generation: DDIM can be used for generating high-quality images, such as faces, objects, and scenes.
Audio Generation: DDIM can also be used for generating high-quality audio, such as music and speech.
Image Editing: DDIM can be used for image editing tasks like image interpolation and manipulation.



Combining audio control and style transfer using latent diffusion