# Arbitrary-Scale Image Generation and Upsampling using Latent Diffusion Model and Implicit Neural Decoder


Jinseok Kim[1,3] Tae-Kyun Kim[1,2]

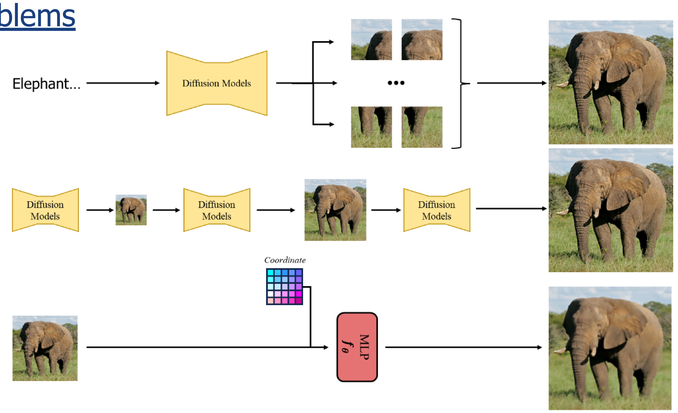
[1]KAIST [2]Imperial College London [3]AI Lab, LG Electronics



CVPR JUNE 17-21, 2024 SEATTLE, WA


## Introduction

### Problems




- Most existing diffusion-based generative models only generate images at **fixed scales** (e.g., 2x, 4x, 8x).
- To handle multiple scales often **require** significant computational resources in terms of both **memory and inference time**.
- Existing single-model super-resolution (SR) methods often produce images with **poor perceptual quality**.
- Lack of **Diversity**.
- Lack of **Scale-Consistency**.

### Goal

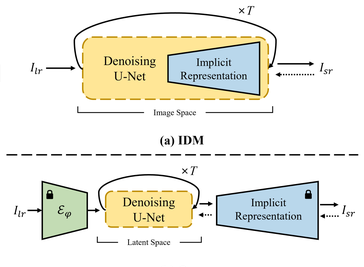
- Develop a method capable of generating and upsampling images at any **arbitrary-scale**.
- Achieve **high perceptual quality**, ensuring that the images are detailed, and free from over-smoothing and artifacts.
- Maintain **consistency** across different scales.
- Create simpler and more **efficient** architecture that reduces the complexity.

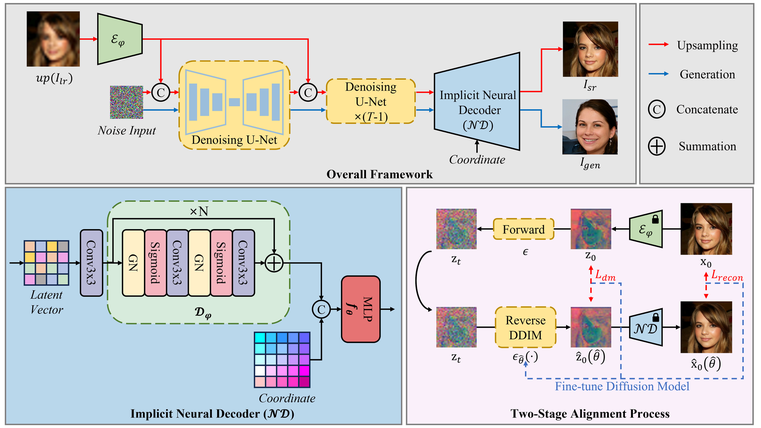
## Methods

We combined a Latent Diffusion Model (LDM) and an Implicit Neural Decoder to enable arbitrary-scale image generation and upsampling.

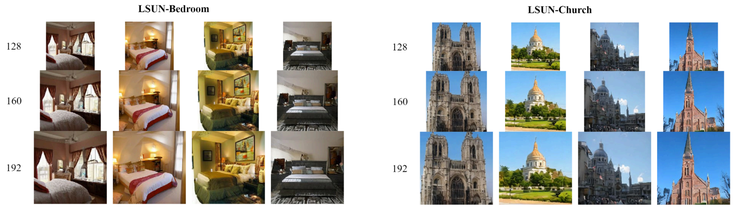
The decoder is designed symmetrically to the encoder to effectively restore the latent vector to high-resolution (HR) images. The RGB values of continuous images $I$ at coordinates $c$ are defined as

$$I(c) = \mathcal{ND}(z, c^*) = f_\theta(\mathcal{D}_\varphi(z), c^*).$$



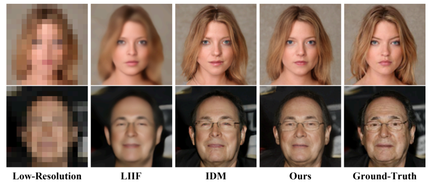

## Results

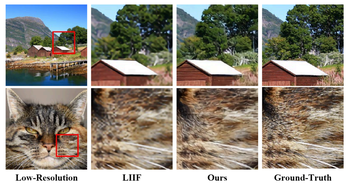



| Dataset: | CelebA-HQ | | | | |
|---|---|---|---|---|---|
| Method | 5.3× | 7× | 10× | 10.7× | 12× |
| LIIF [7] | **27.52** / 0.1207 | **25.09** / 0.1678 | 22.97 / 0.2246 | 22.39 / 0.2276 | 21.81 / 0.2332 |
| SR3 [27] | - | 21.15 / 0.1680 | 20.25 / 0.2856 | - | 19.48 / 0.3947 |
| IDM [9] | 23.34 / 0.0526 | 23.55 / 0.0736 | 23.46 / 0.1171 | 23.30 / 0.1238 | 23.06 / 0.1800 |
| Ours | 24.66 / **0.0455** | 24.13 / **0.0690** | **23.64 / 0.1110** | **23.62 / 0.1183** | **23.52 / 0.1427** |

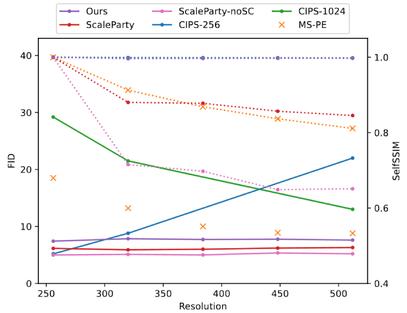
PSNR↑/LPIPS↓

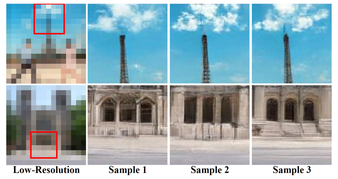
| Dataset: | Lsun Bedroom | LSUN Tower |
|---|---|---|
| Method | 16× | |
| PULSE [22] | 12.97 / 0.7131 | 13.62 / 0.7066 |
| GLEAN [6] | 19.44 / 0.3310 | 18.41 / 0.2850 |
| IDM [9] | **20.33** / 0.3290 | 19.44 / 0.2549 |
| Ours | 20.08 / **0.3269** | **21.24 / 0.1897** |




| Method | FPS↑ | | | | |
|---|---|---|---|---|---|
| | 8× | 12× | 30× | 100× | 200× |
| IDM | 0.0202 | 0.0200 | 6.54e-3 | 3.98e-4 | - |
| Ours | **0.2568** | **0.2510** | **0.2473** | **0.1833** | **0.0982** |




Diversity

Ablation Strudy