Super-Resolution Using A Unified Latent Diffusion Model For Medical Imaging

Ava kouhana[1], Mojtaba Jafaritadi[1], Garry Chinn[1], Craig S. Levin[1,2]

[1]Stanford University, Radiology Department

[2] Stanford University, Physics, Electrical Engineering, and Bioengineering Department

{akouhana, mojjaf, gchinn, cslevin}@stanford.edu

MIPS Molecular Imaging Program at Stanford

Purpose

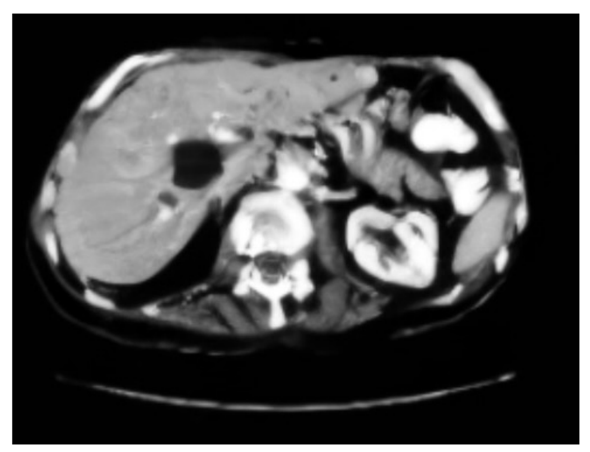
- **Motivation:** Super-resolution (SR) techniques reconstruct high-resolution images from low-resolution data. In medicine, high-quality imaging is costly but crucial for diagnosis. Deep learning-based SR methods have been developed to address this. While diffusion models show promise in generating high-resolution images, their potential for medical image enhancement remains unexplored.

- **Purpose:** To investigate the potential of a conditional latent diffusion model (LDM) architectures as a viable alternative to GANs for SR in medical imaging.

New Metrics

- We propose Laplacian variance and FFT as alternative metrics to PSNR, SSIM, and NRMSE for better capturing perceptual differences in SR image quality.

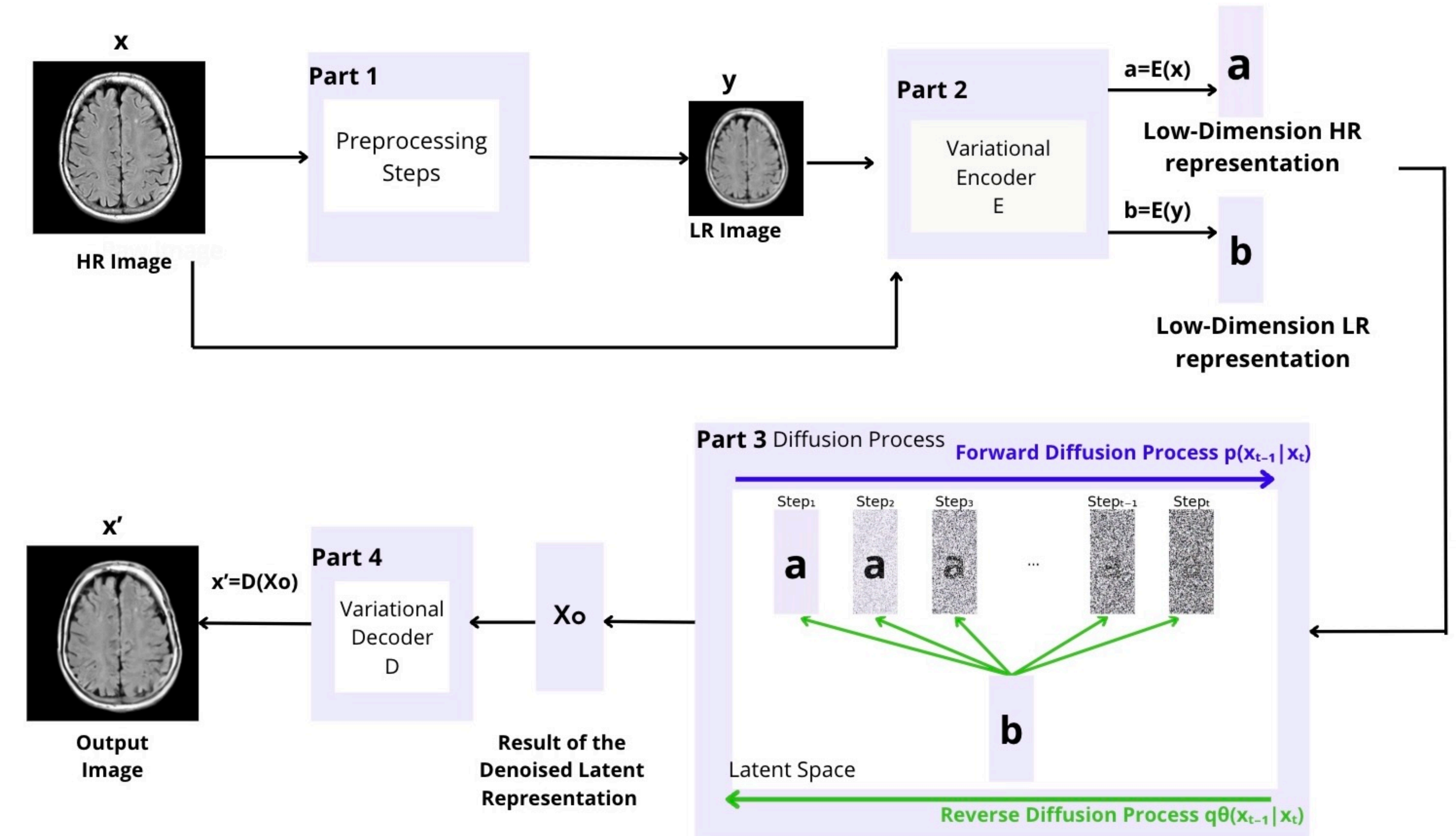
NRMSE: 0.064
PSNR: 24.33
MS-SSIM: 0.976



NRMSE: 0.061
PSNR: 23.96
MS-SSIM: 0.979

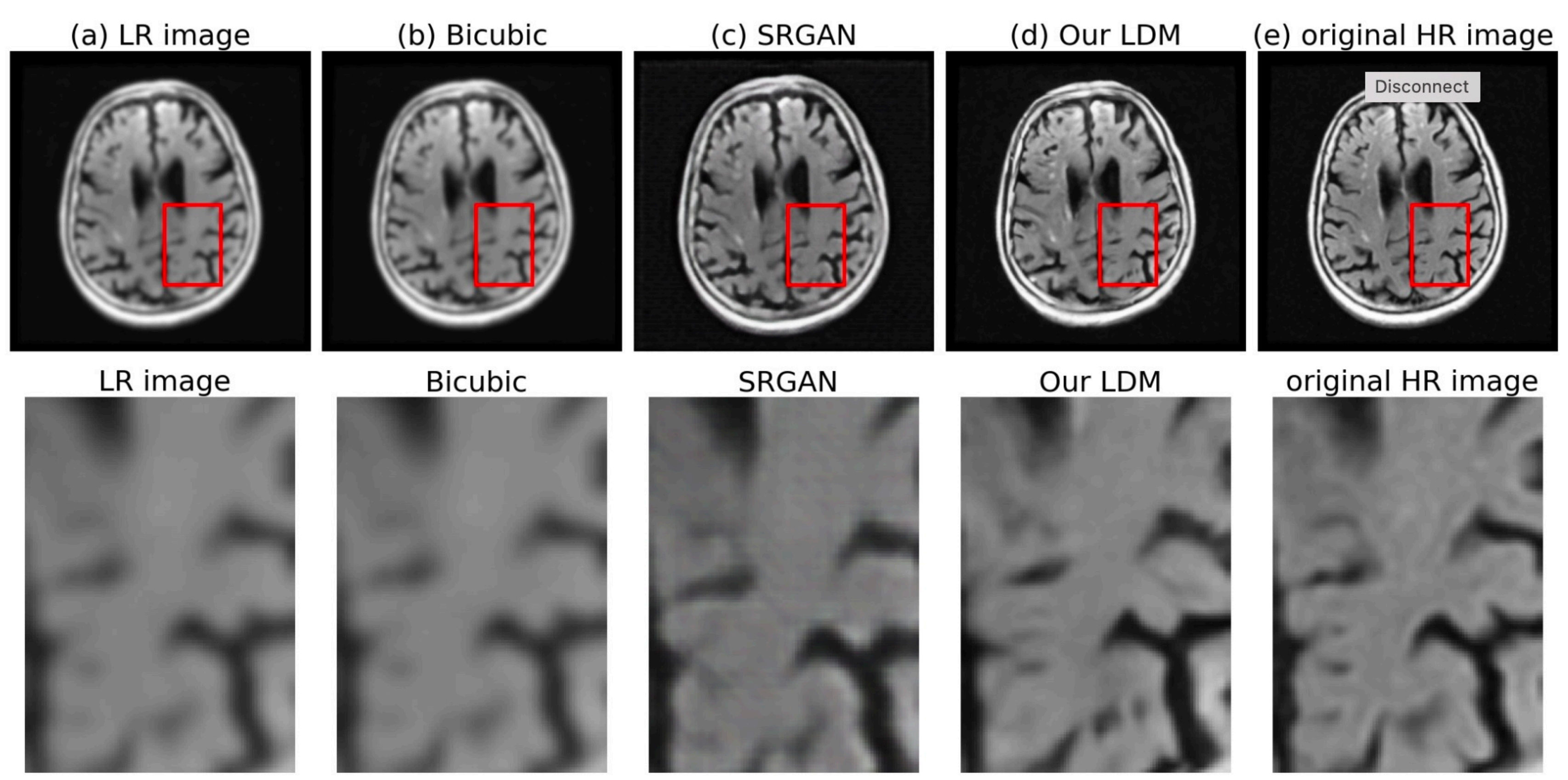
Methods

- **Approach:**



- The conditional LDM architecture, conditioned on LR image, relies on a U-Net-like structure with 4 convolutional layers. A pretrained Auto-encoder was used to encode and decode LR and HR images in the latent space.

Results



Metrics	Laplacian Variance	FFT
LR image	26.83 ± 10.05	-58.08 ± 9.95
Bicubic	48.9 ± 25.9	-39.39 ± 7.90
SRGAN	292.0 ± 125.2	-1.16 ± 6.72
Our LDM	549.80 ± 198.0	3.20 ± 6.89
HR original image	554.68 ± 224.1	4.45 ± 8.02

- Laplacian variance measures image sharpness by quantifying edge frequency. Higher variance indicates more edges and greater sharpness. It's calculated as: $\sigma L^2 = (1/N)\ \Sigma(Li - \bar{L})^2$ where N is pixel count, Li is Laplacian of ith pixel, and $\bar{L}$ is mean Laplacian value.

- FFT decomposes an image into frequency components. Blur reduces high frequencies more than low frequencies. Lower μ (mean value) indicates more blur; higher μ suggests a sharper image with $F(u,v) = \Sigma\ \Sigma\ I(x,y)\ e^{(-2\pi i(ux+vy)/N)}$.

- The average Laplacian variance for the LR, SRGAN, predictions, and HR original images were 27, 292, 550, and 555 respectively. Similarly, the average FFT values were −58, −1, 3, and 4, respectively.

Conclusions

- The results indicate that the synthetic SR images generated by the conditional LDM exhibit enhanced sharpness and reduced blurriness compared to state-of-the-art SRGAN, highlighting LDM's potential for advancing SR techniques in radiology.

- Diffusion model-based SR could enhance medical image quality, potentially improving diagnoses without costlier HR imaging and limitations risks for patients health.

- **Limitations:** Diffusion models are computationally intensive and resource-demanding at sampling time compared to other generative models like GANs.

- Also, in imaging, diffusion models may generate non-existent features, potentially leading in the context of medical imaging to potential misdiagnosis.