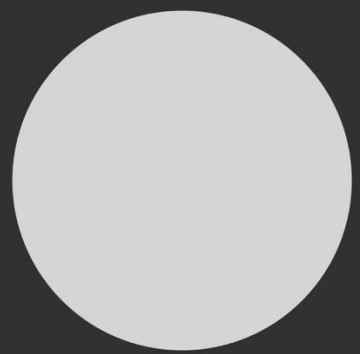


Brain Imaging Generation with Latent Diffusion Models

**Making Brains Out of
Thin Air (And lots of MRI
scans)**

A



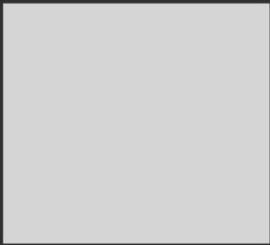
B



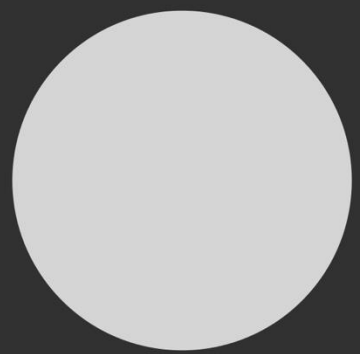
B



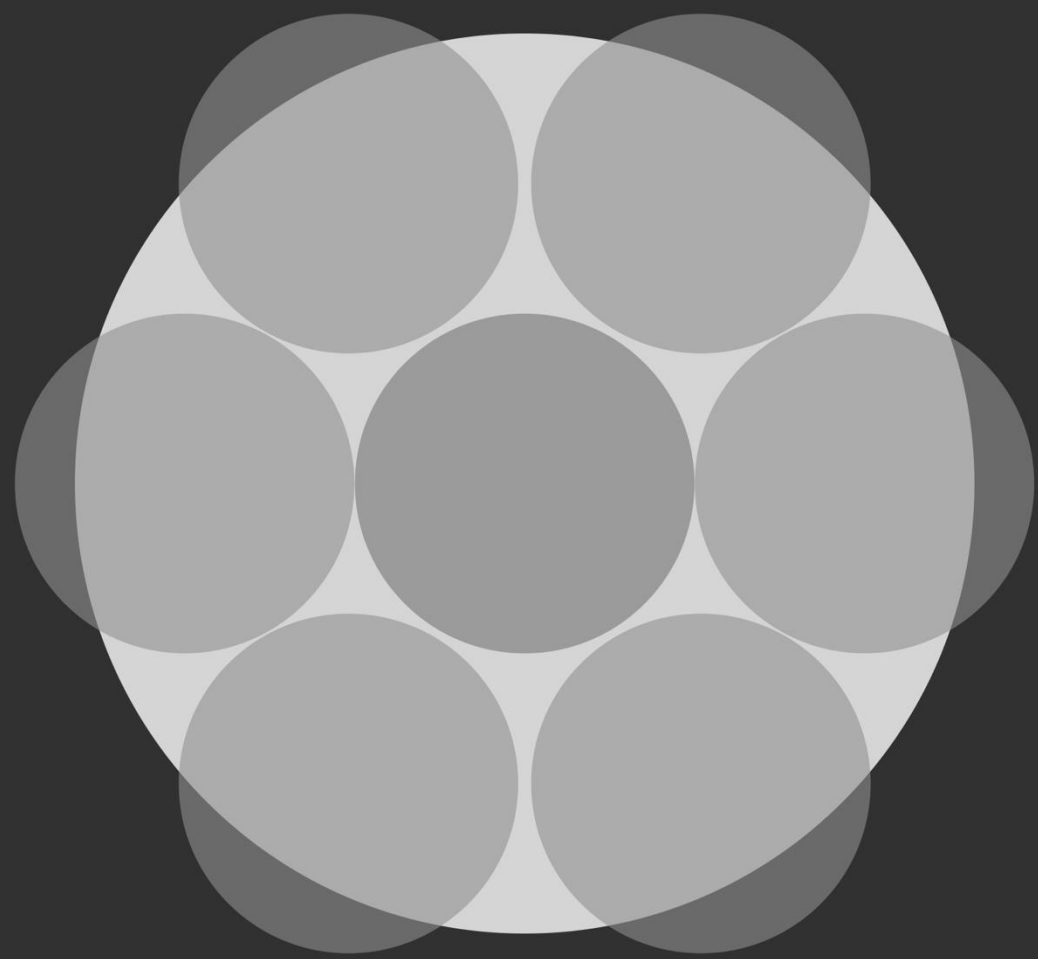
A



A



B



Area

B

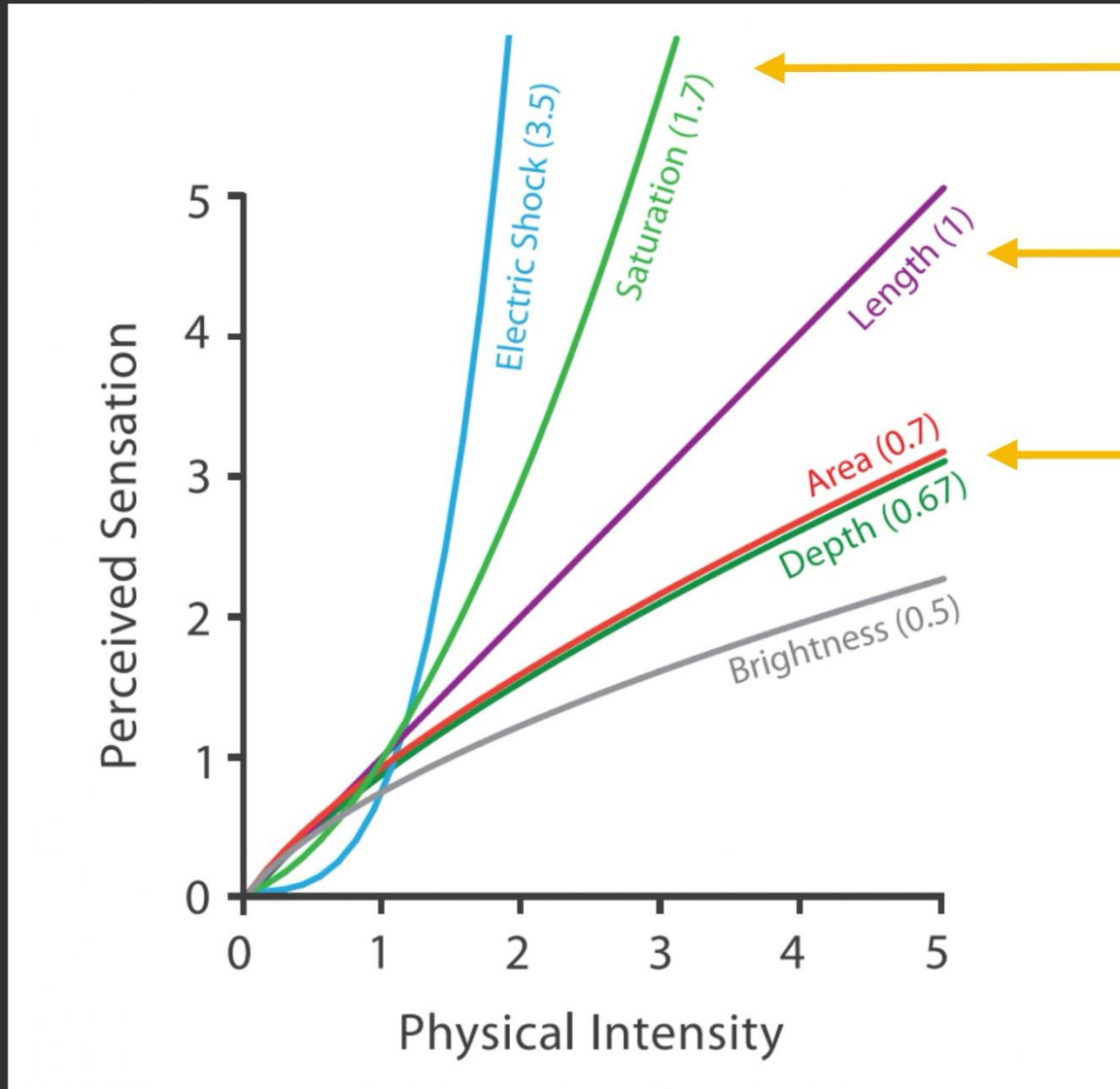


A



Length

Steven's Power Law



Overestimate

Unbiased

Underestimate

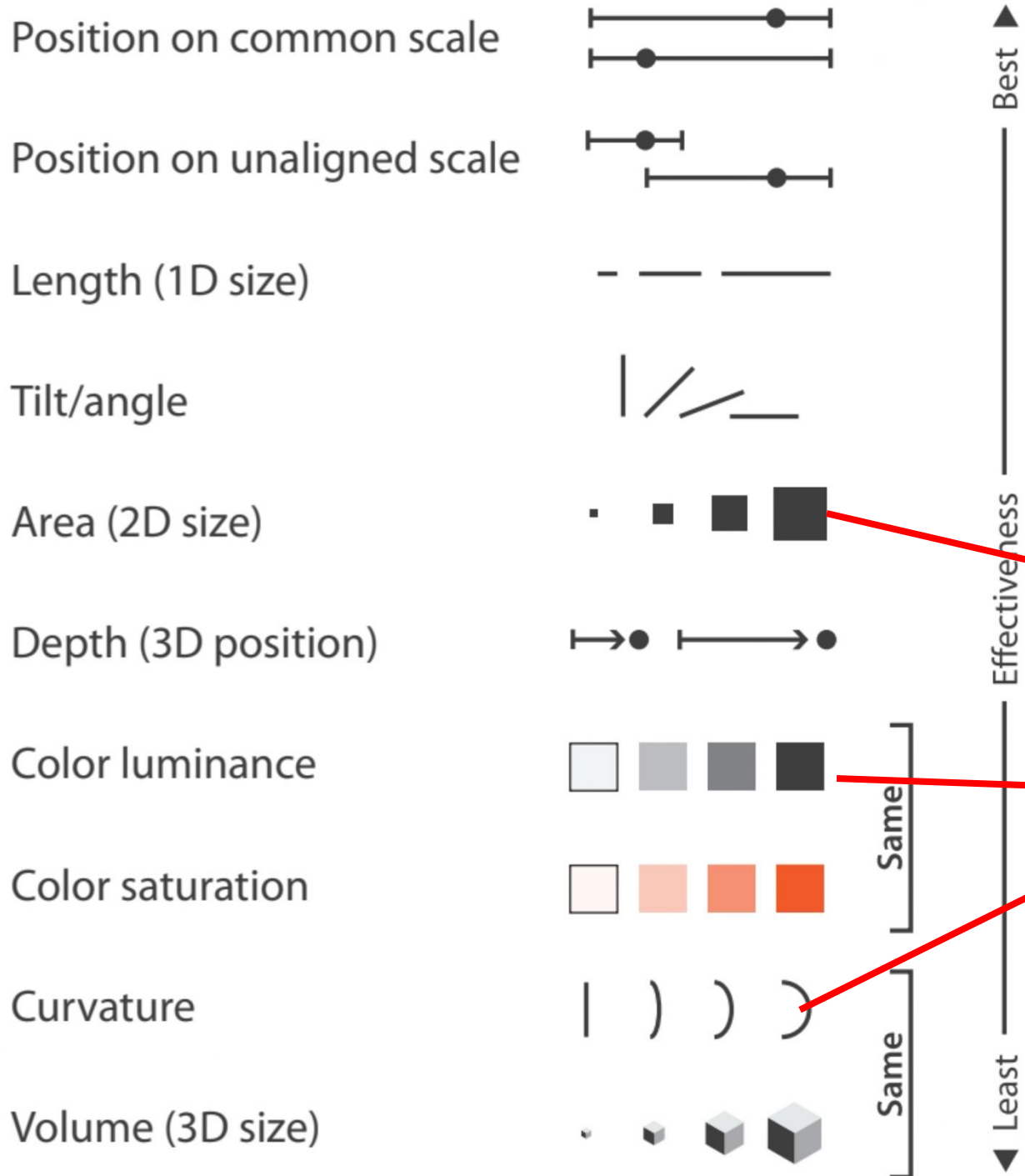
Weber's Law

Just Noticeable Difference (JND)

Perceived Change \longrightarrow $dp = k \frac{dS}{S}$ \longleftarrow Change of Intensity
 \longleftarrow Physical Intensity

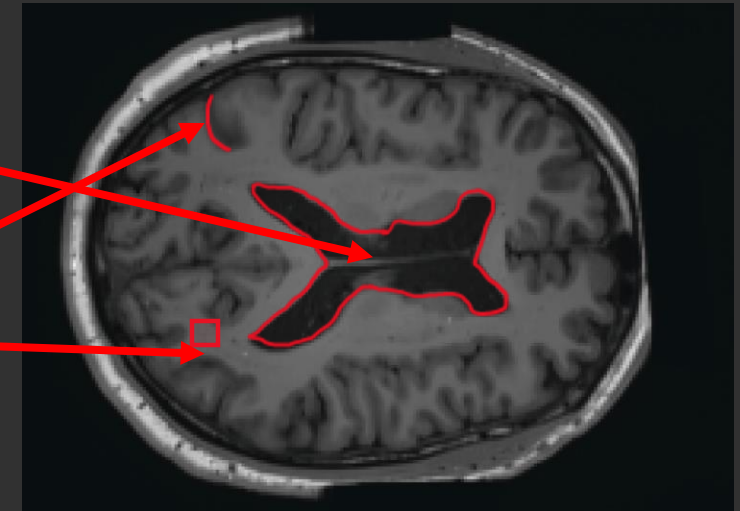
Most continuous variation in stimuli are perceived in discrete steps





Effectiveness Ranking of Visual Encoding Variables

for comparing numerical quantities

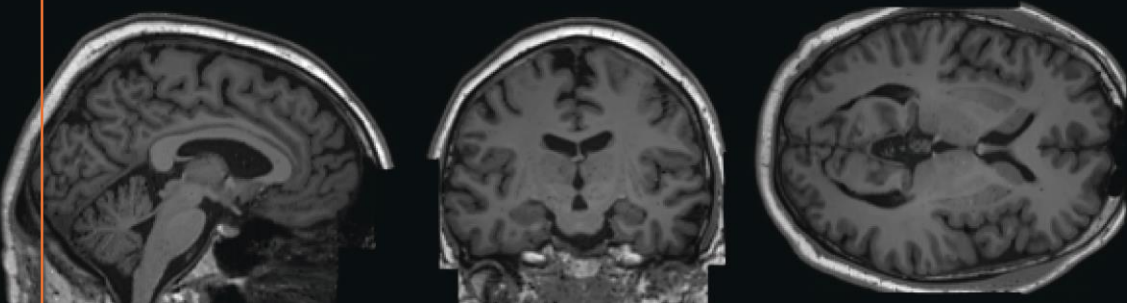
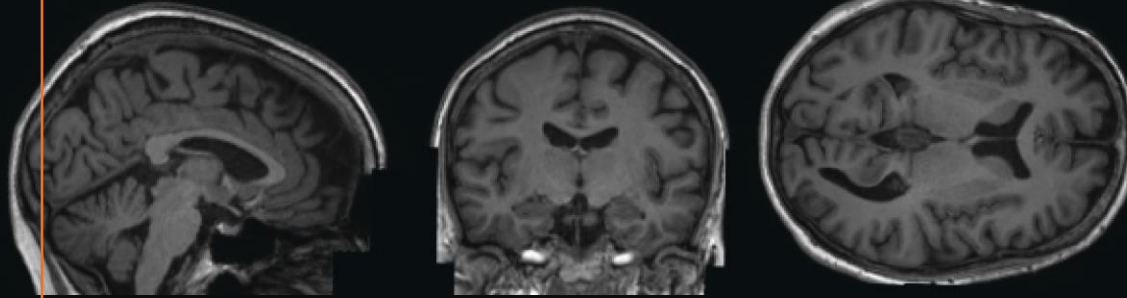
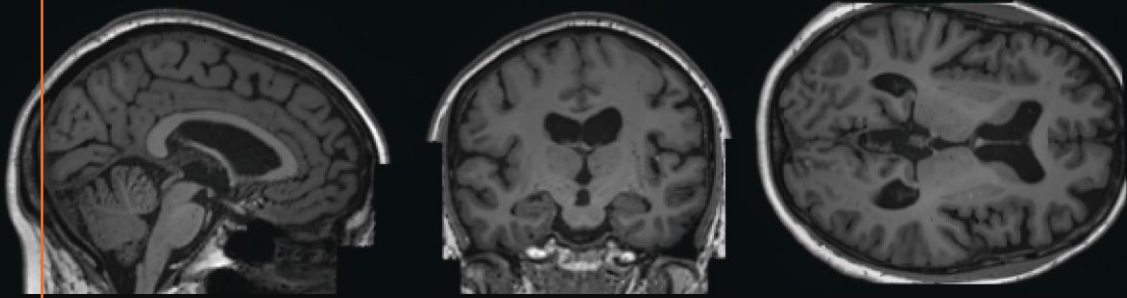


Defining MRI features

[T. Munzer 2014]

Take away messages

- Humans are biased when it comes to **magnitude perception**
- Good metrics don't necessarily mean good quality images
- Human perception **must** guide the evaluation criteria for generated scans



Brain Imaging Generation with Latent Diffusion Models

Making Brains Out of Thin Air (And lots of MRI scans)

Motivation

- Data in the medical domain is scarce
- Rare pathologies
- Privacy concerns
- **DNNs are data hungry!**



Figure from ANI South Asia, Jan 2025

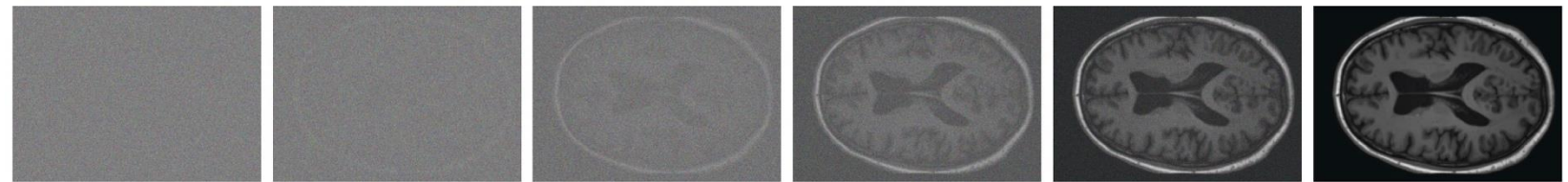
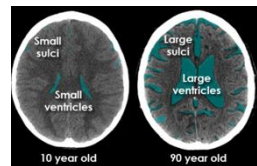


Figure from CyberDB, Jan 2025

Proposed solution

- Use Generative Networks to generate MRI scans conditioned on age, sex, brain structure volumes.

AGE



Reverse Diffusion

Challenges

- MRIs require a very high resolution
- MRI data is 3D (voxels not pixels)
- Each MRI takes up a lot of GPU memory
- **DNNs are data hungry!**

MRI data

5.7 M

Full HD RGB image

2.3 M

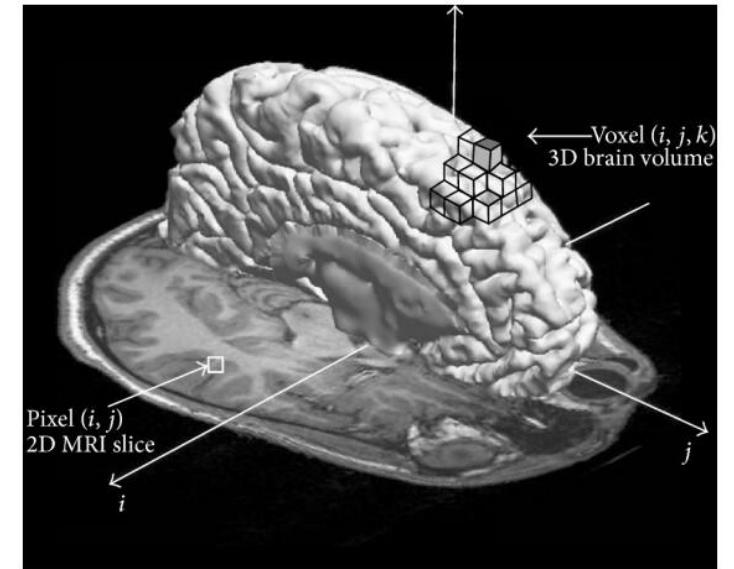


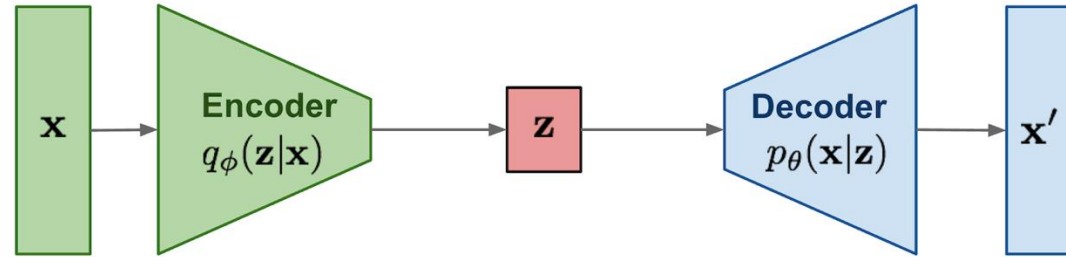
Figure from Clark's Radiography



Dalle 3 - 2024

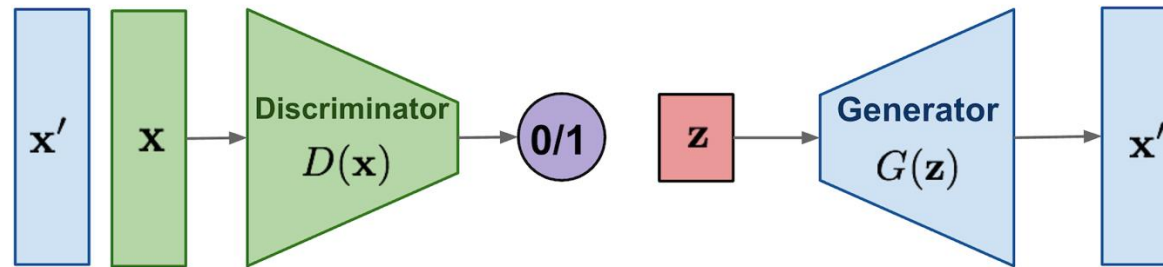
Review on Generative models

VAE: maximize variational lower bound



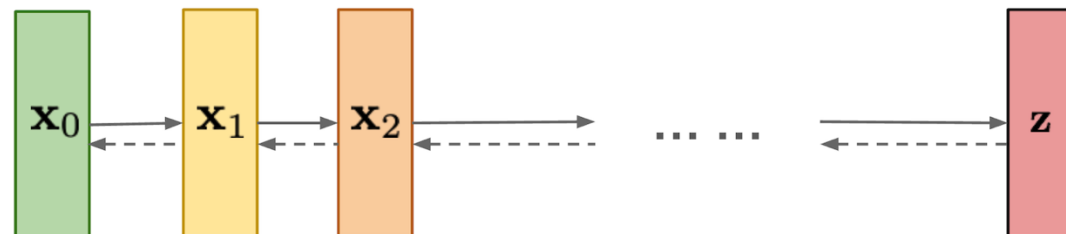
- + great diversity
- poor sampling quality (low resolution, blurry)

GAN: Adversarial training



- + quickly generate high-fidelity sample
- mode coverage is limited
- hard to train

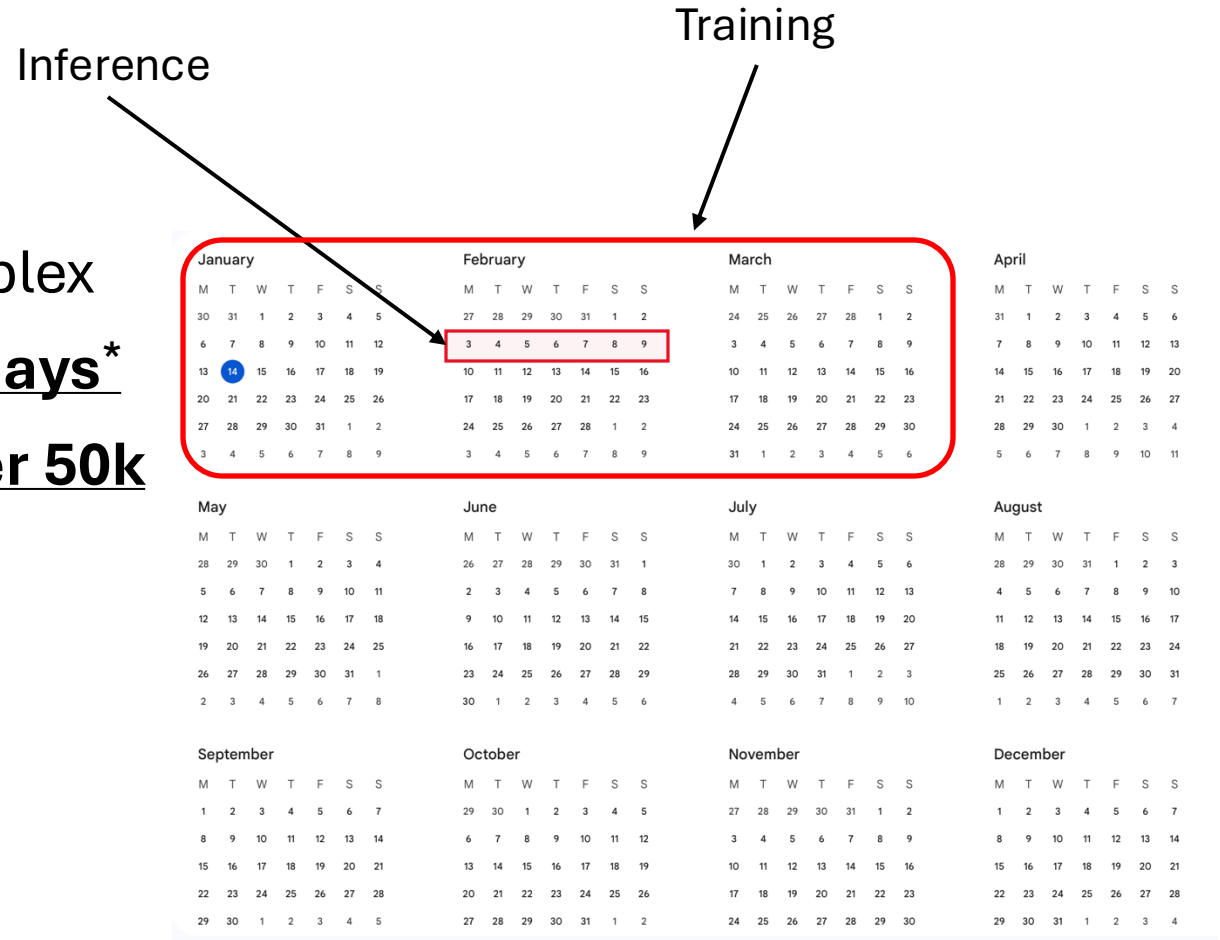
Diffusion models: Gradually add Gaussian noise and then reverse



- + good mode coverage
- + high quality sampling
- slow & expensive sampling

Challenges

- Training a Diffusion Model is complex
- Training time on **10 V100s: ≈100 days***
- Average inference time: **5 days per 50k samples*** (single A100)
- **Retraining and regeneration are common**



* training a SOTA Diffusion Model as per paper

Pixel space is overly redundant.

- **Locality and Spatial Coherence**
- **Translation Invariance**
- **Hierarchical Composition**
- **Object and Shape Bias**

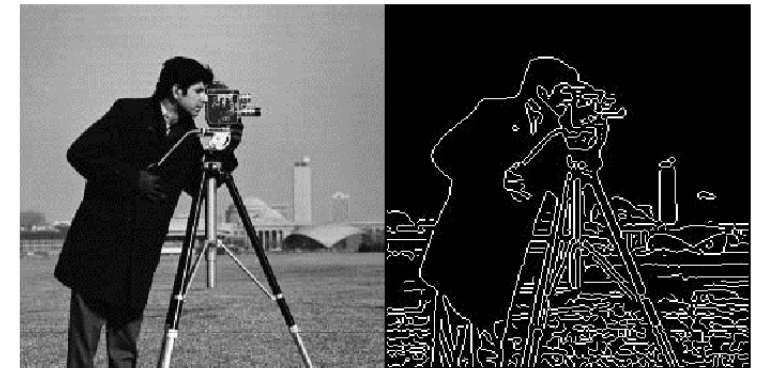
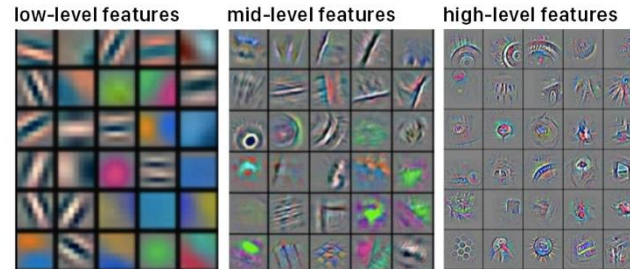
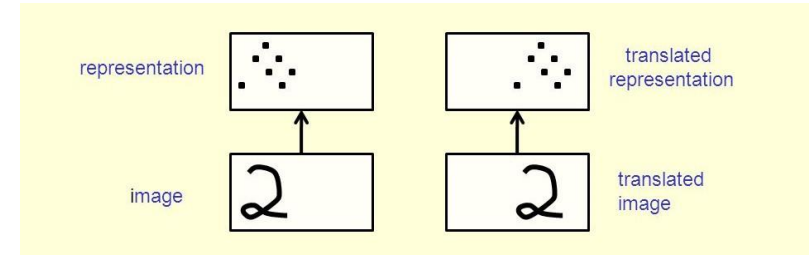
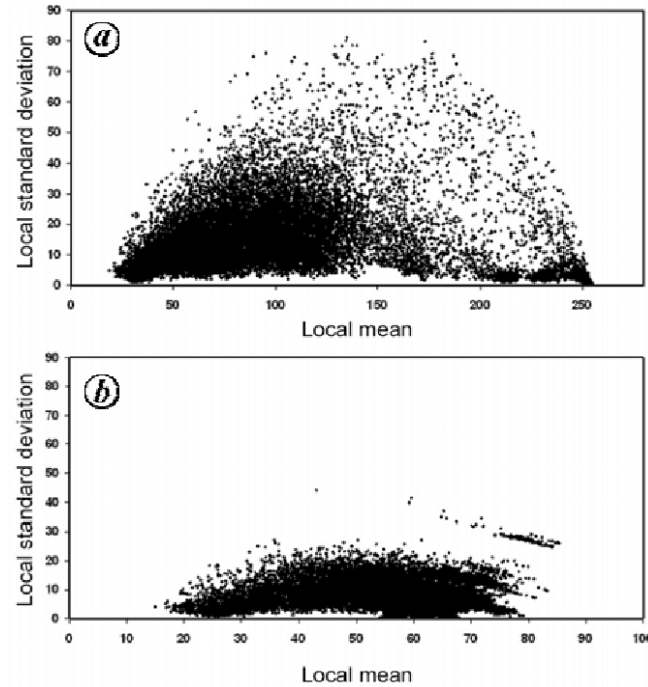


Figure 1: The cameraman image and its edges extracted

Solution: Diffusion in Latent Space

- Latent space holds a more informative lower dimensional representation of the data
- Use the **latent encoding** for the diffusion process

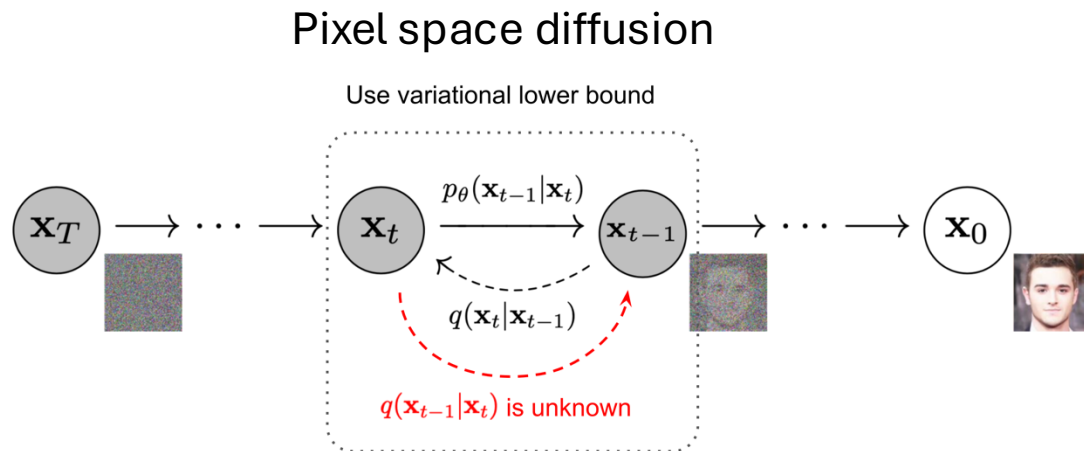


Figure from DDPM, Jonathan Ho et al.

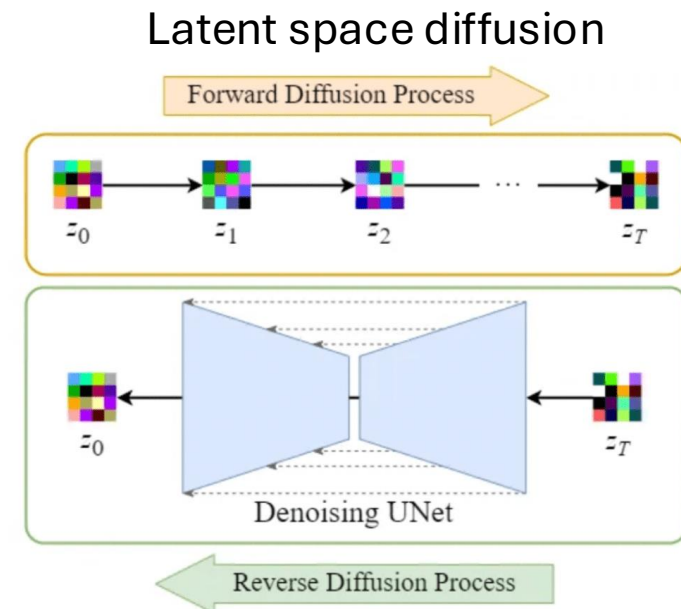


Figure excerpted from Stable Diffusion clearly explained¹⁹

High-Resolution Image Synthesis with Latent Diffusion Models

Robin Rombach¹ *

Andreas Blattmann¹ *

Dominik Lorenz¹

Patrick Esser^ℜ

Björn Ommer¹

¹Ludwig Maximilian University of Munich & IWR Heidelberg University, Germany

^ℜRunway ML

<https://github.com/CompVis/latent-diffusion>

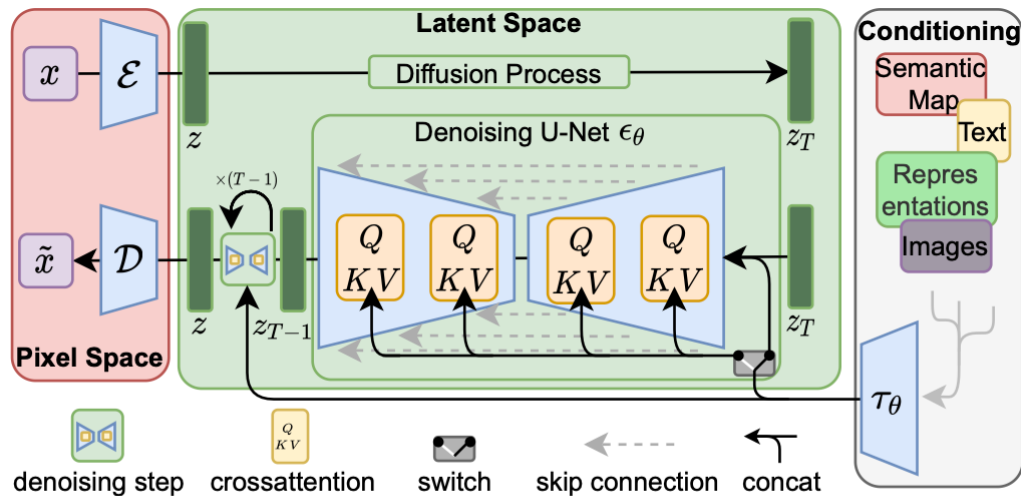


Figure 3. We condition LDMs either via concatenation or by a more general cross-attention mechanism. See Sec. 3.3



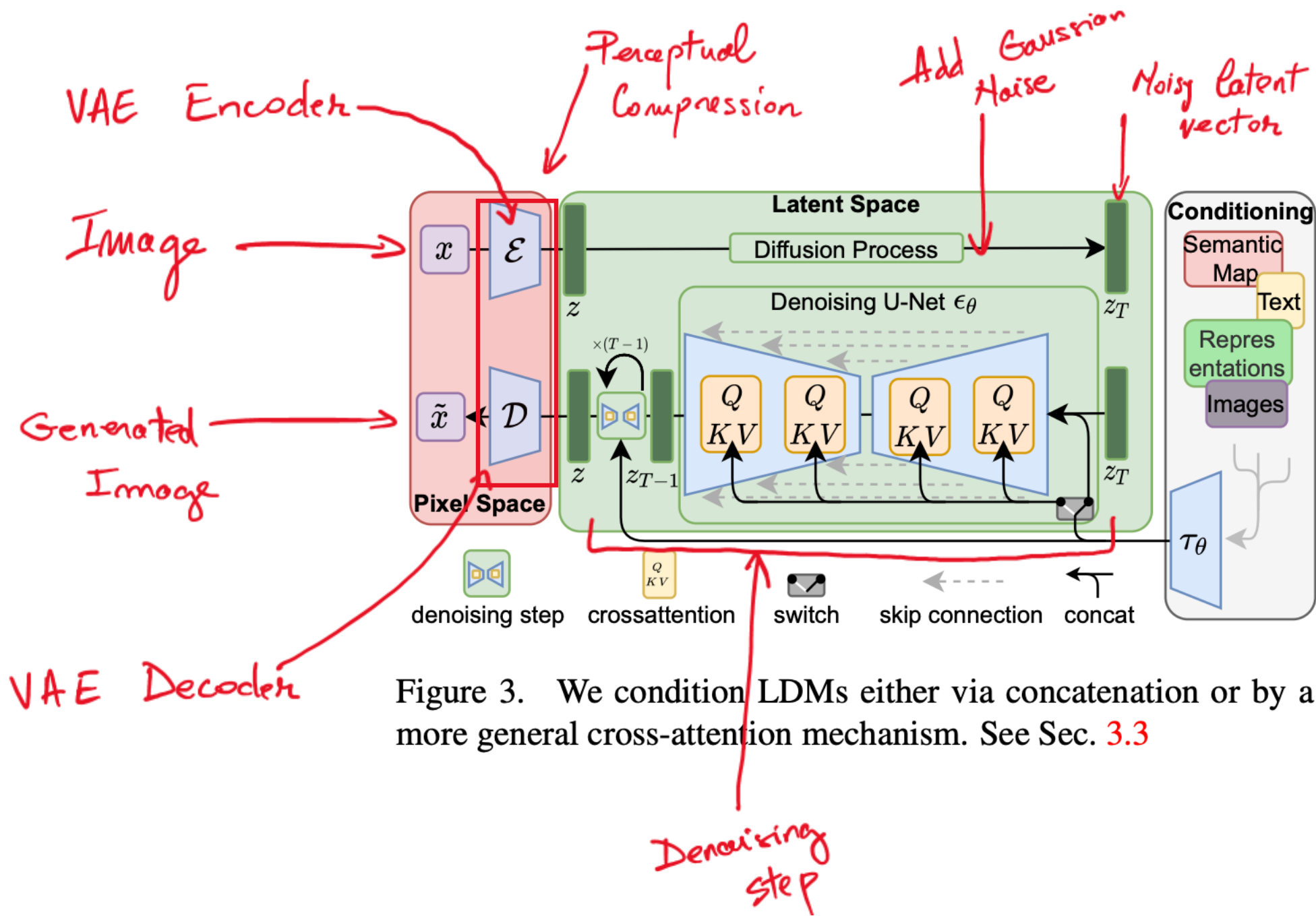


Figure 3. We condition LDMs either via concatenation or by a more general cross-attention mechanism. See Sec. 3.3

LDM Training Objective

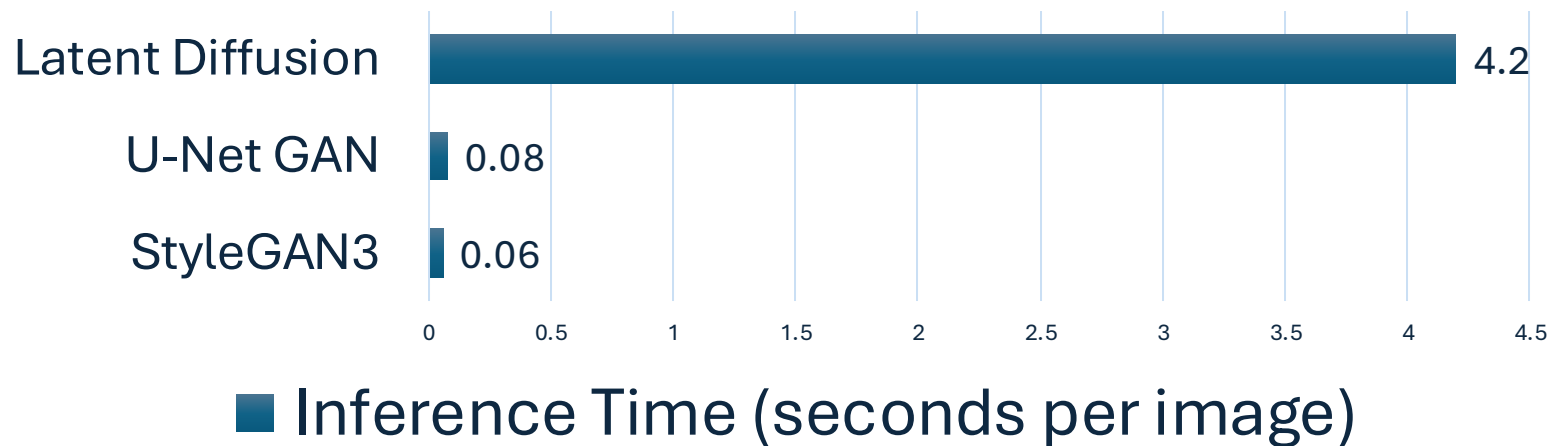
$$L_{LDM} = \mathbb{E}_{\mathcal{E}(x)\epsilon \sim \mathcal{N}(0,1), t} \left[\|\epsilon - \epsilon_{\theta}(z_t, t)\|_2^2 \right]$$

$\epsilon_{\theta}(\circ, t)$ = Time-conditional UNet

$\mathcal{E}(x)$ = Encoder output

Latent Diffusion Models

- Autoencoder can be trained separately
- Cross-attention mechanisms integrates conditioning on text description, images or other modalities
- Inference is still slower than GANs or VAEs



Data from Haotian Wang et al. 2024

Denoising Diffusion Implicit Models (DDIM)

- Replaces stochastic sampling with a non-Markovian deterministic update in the denoising process
- Allows trade-off between speed and quality without retraining

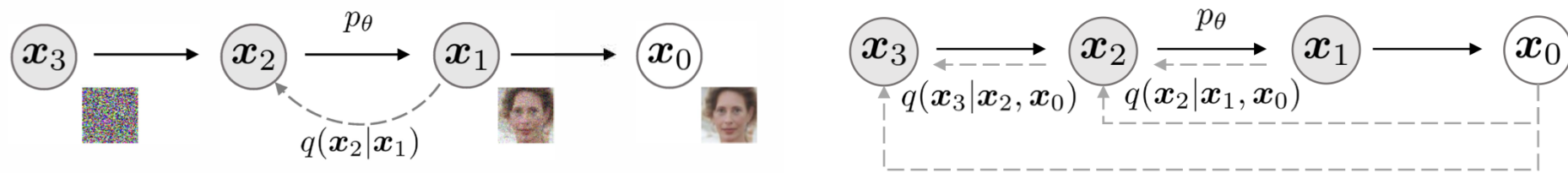
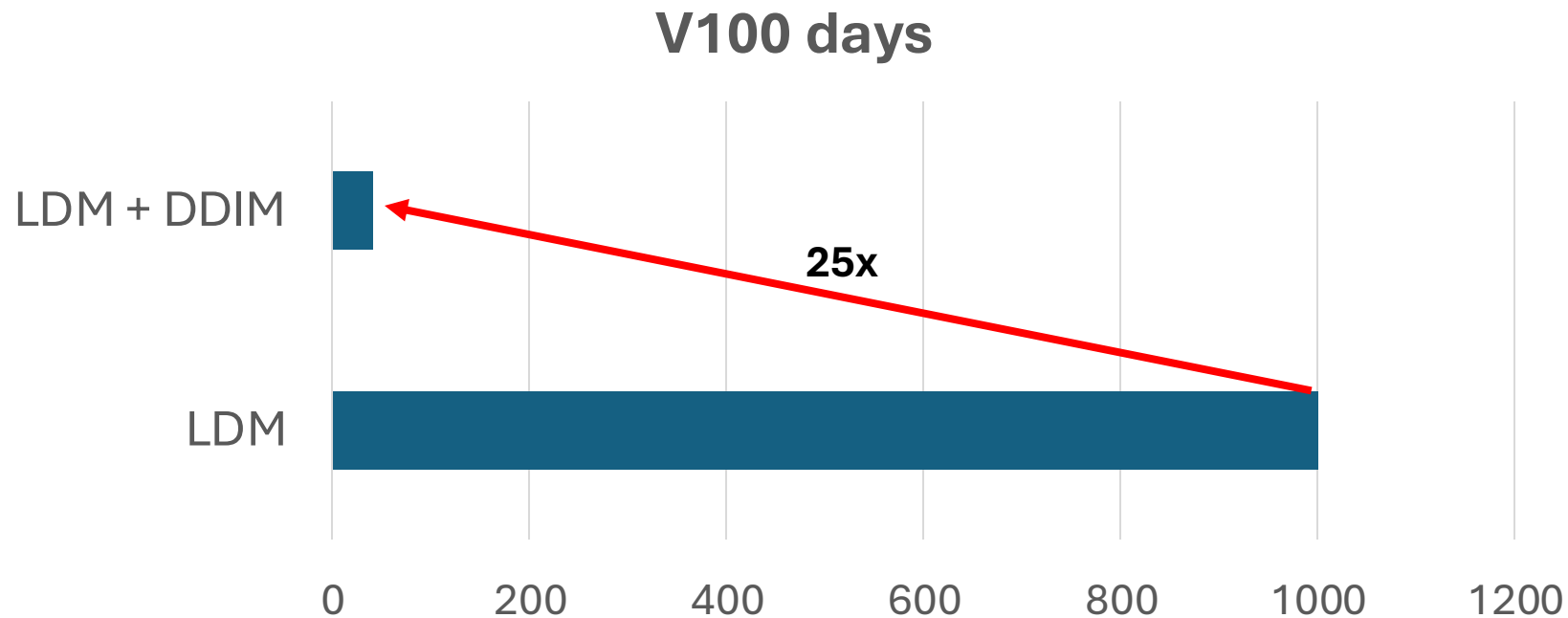


Figure 1: Graphical models for diffusion (left) and non-Markovian (right) inference models.

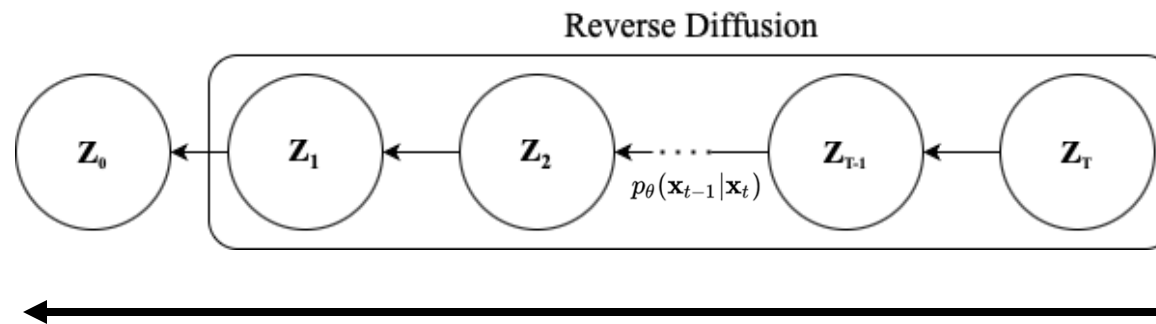
Denoising Diffusion Implicit Models (DDIM)

- Training time reduction from **1000** V100 days to **40** V100 days with SOTA metrics



Denoising Diffusion Implicit Models (DDIM)

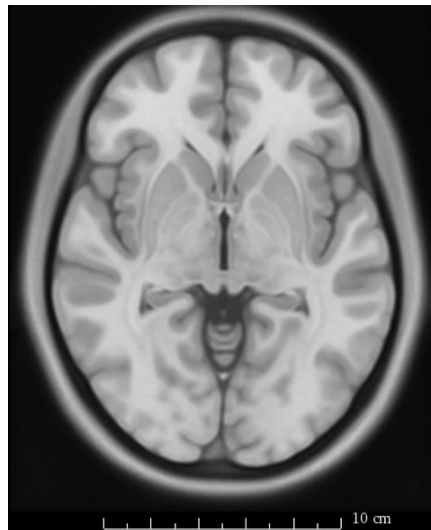
Mean sampling time per MRI (s)



LDM + DDIM **50** iterations
LDM **1000** iterations

Dataset

- T1w MRI images from the UKB dataset
- Only healthy individuals (63.6 ± 7.5 years)
- Images were rigid body registered to a common MNI space
- Final image is $160 \times 224 \times 160$ voxels, 1 mm^3 voxel size



Metrics

- Fréchet inception distance (FID) - **quality**
- Multi Scale Structural Similarity Index (MS-SSIM) - **diversity**
- Four-Grained Reference SSIM (4-G-R-SSIM) - **diversity**

Fréchet Distance Between Normal Distributions

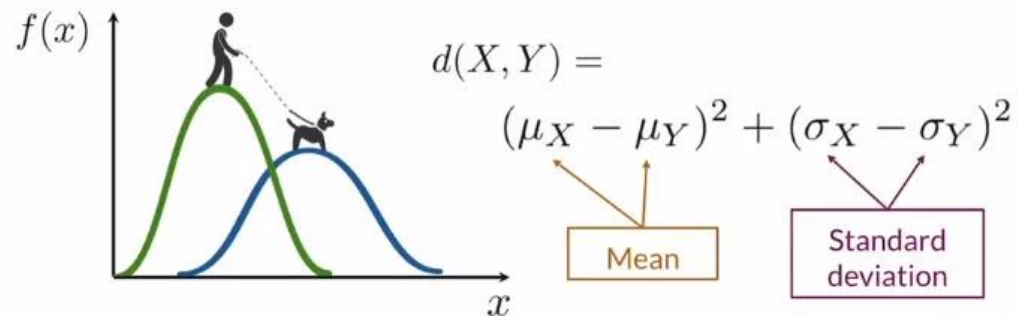


Figure by Mustafa Adel Ibrahim on Medium

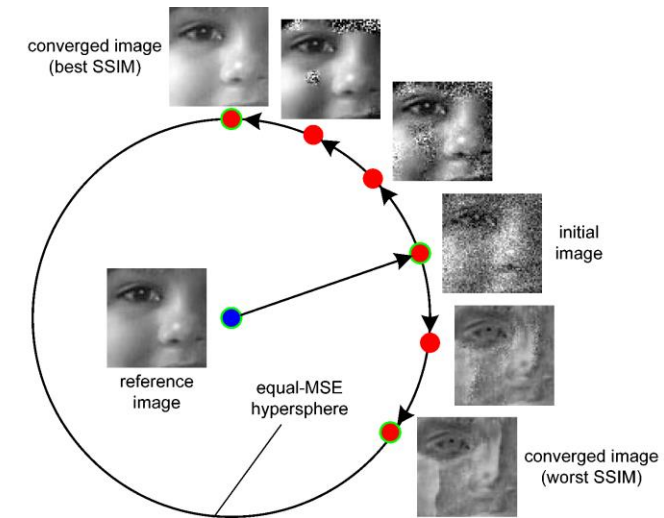
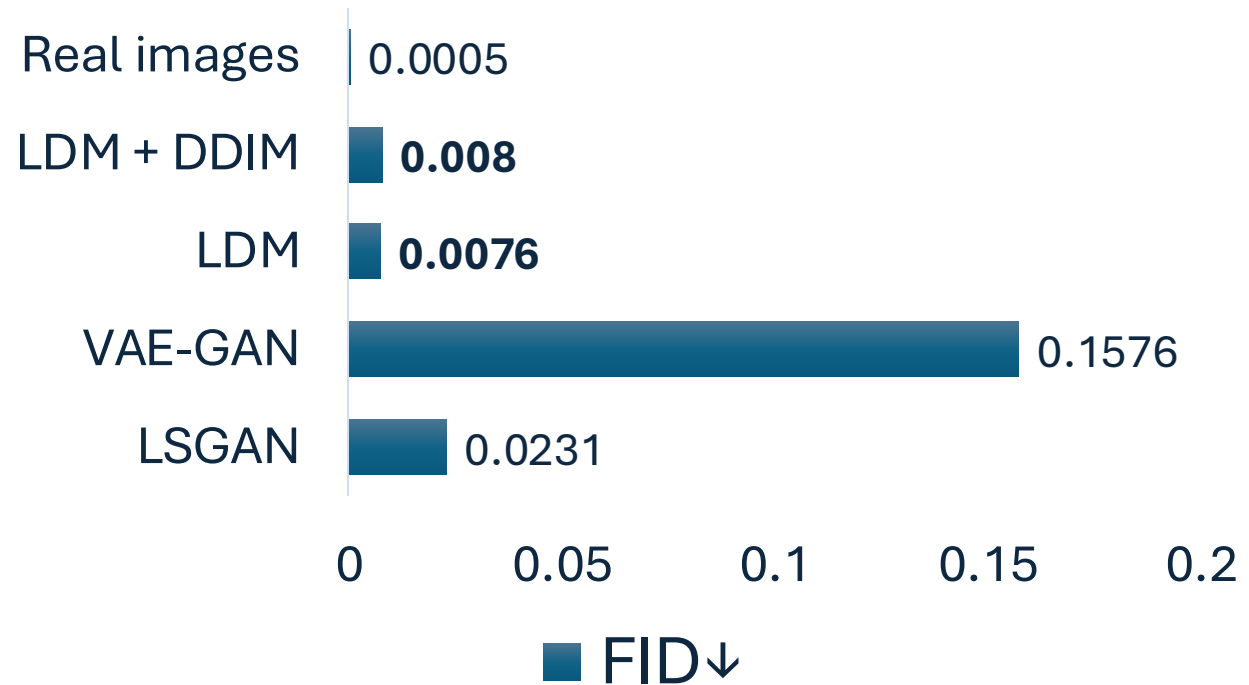
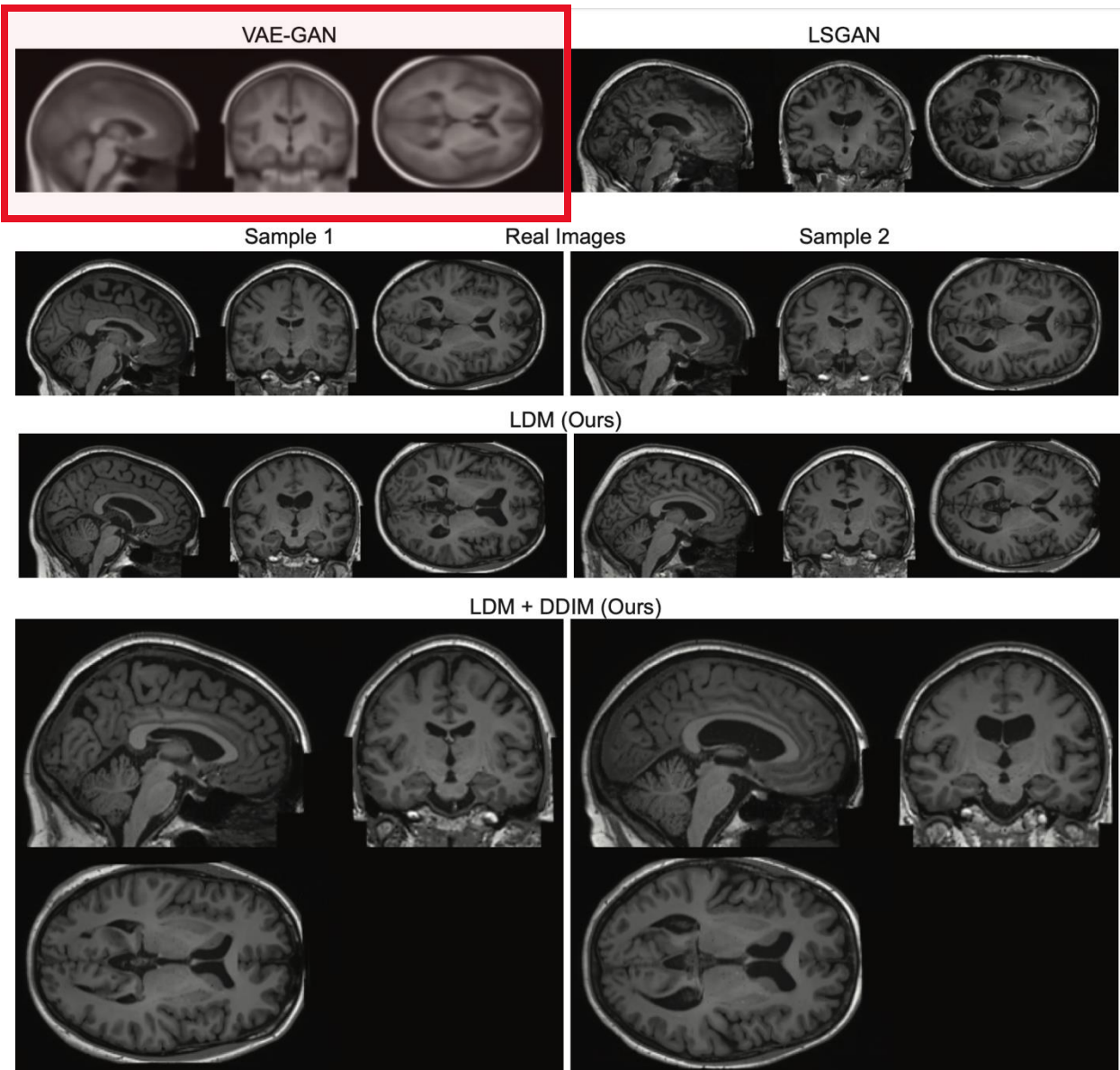
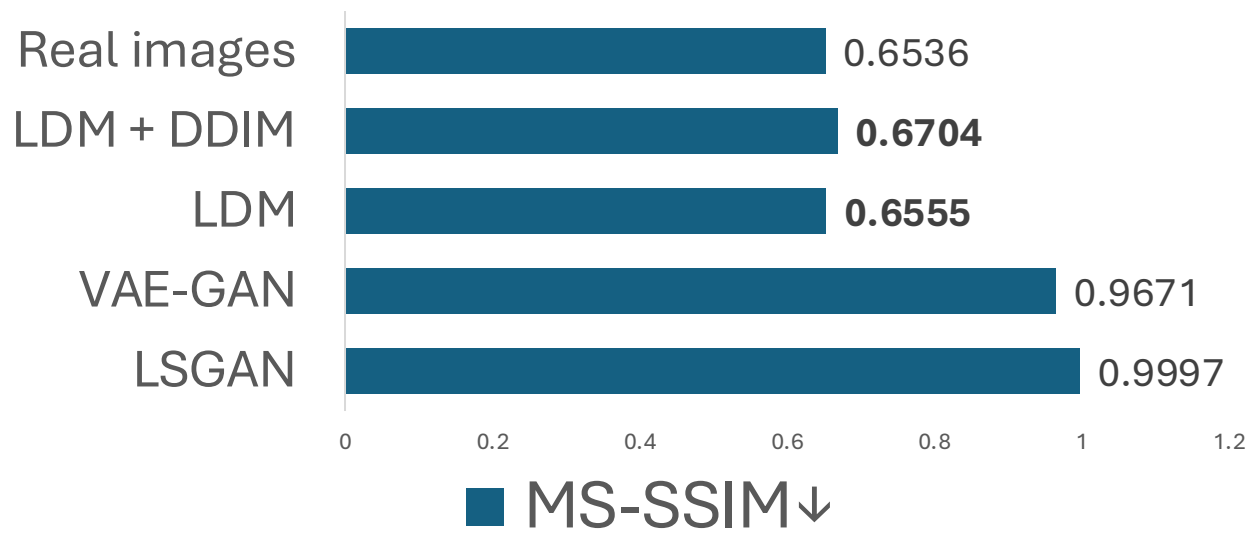
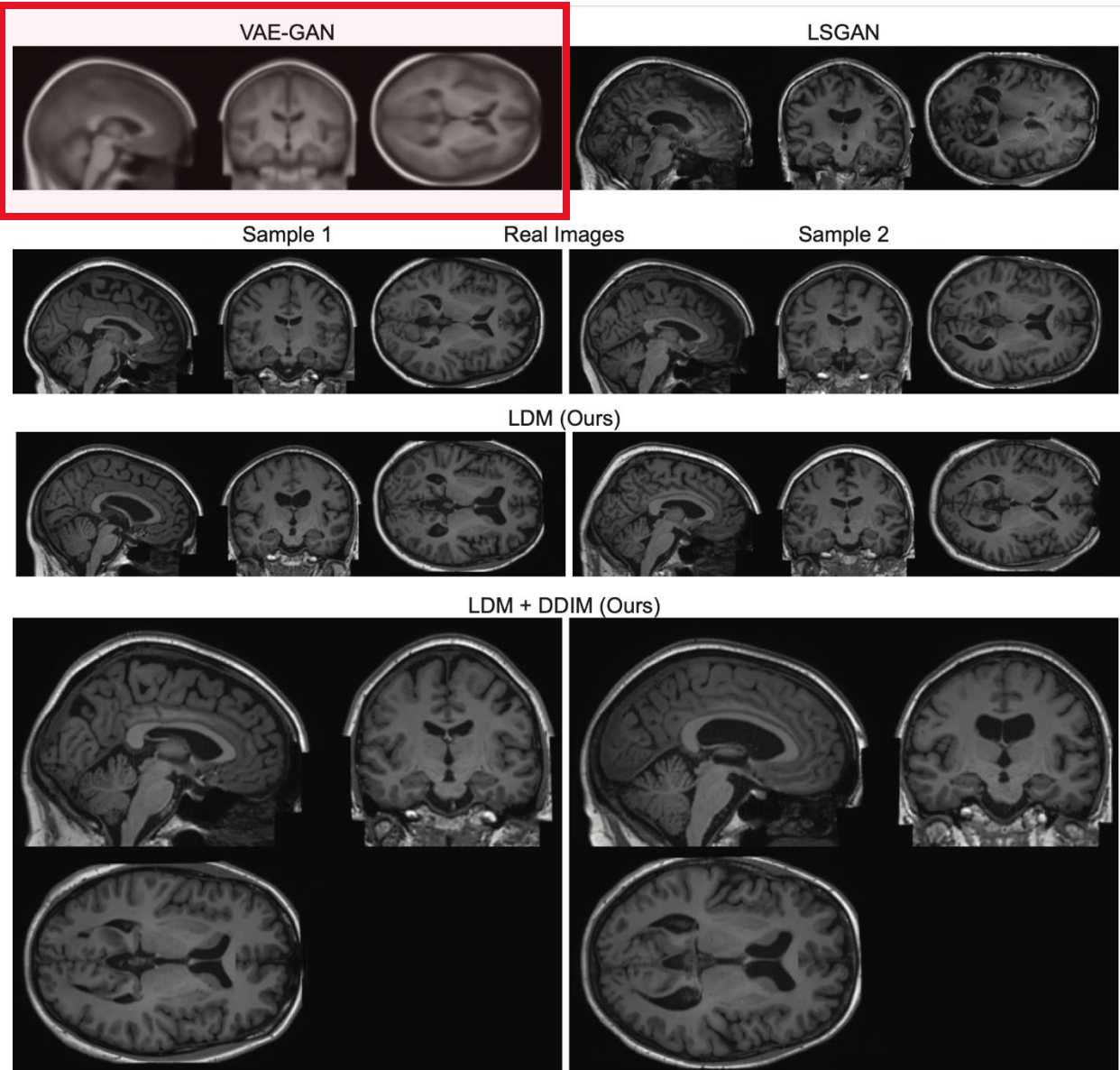


Figure by Zhou Wang, CNS NYU

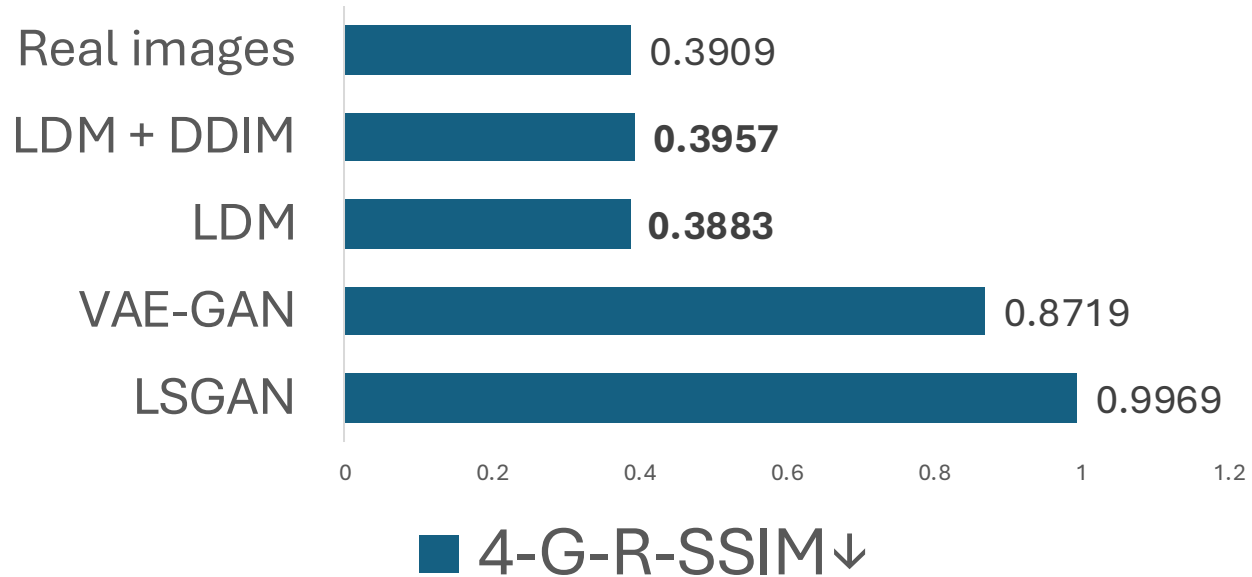
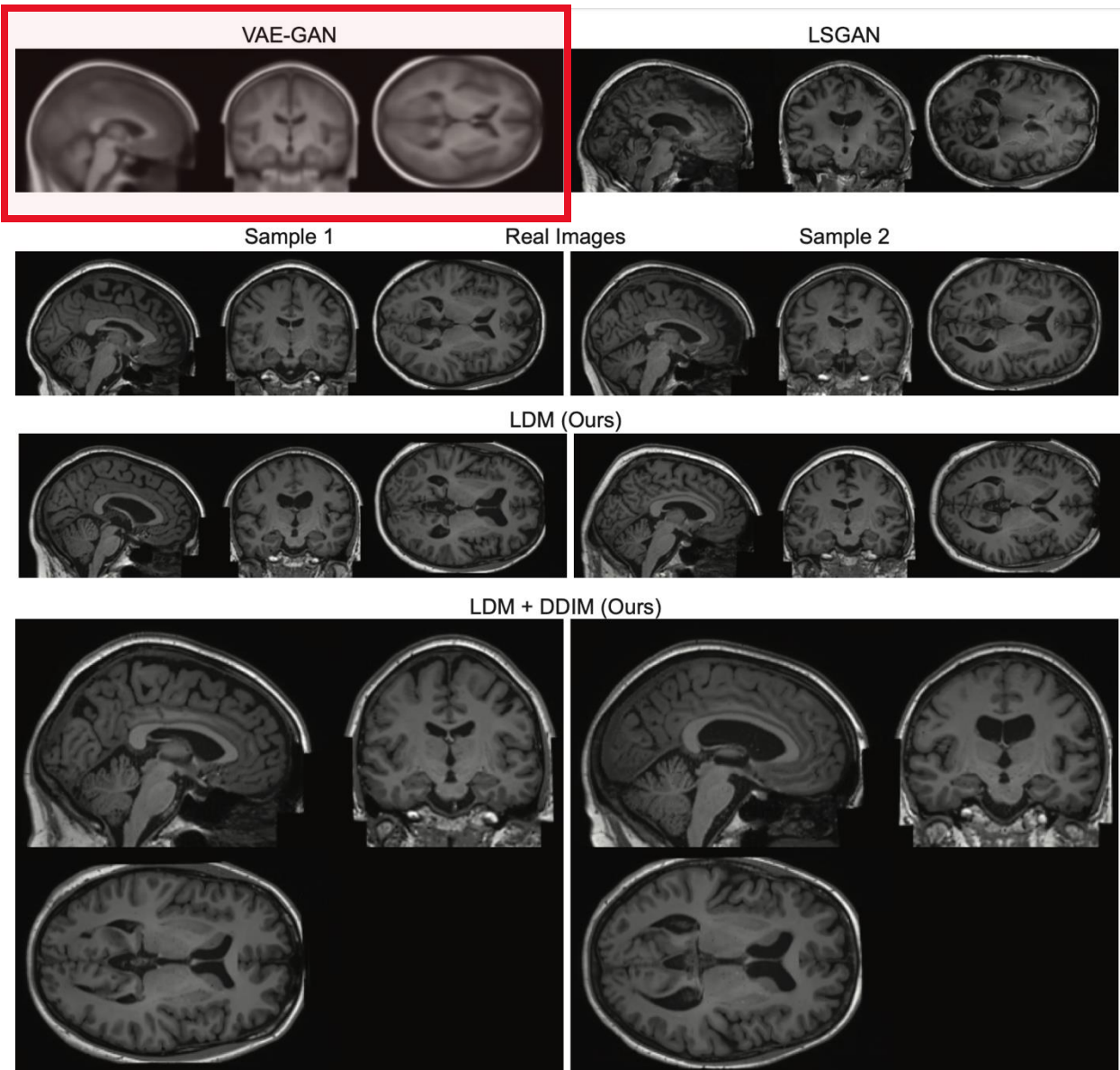
Blur output



Blur output



Blur output

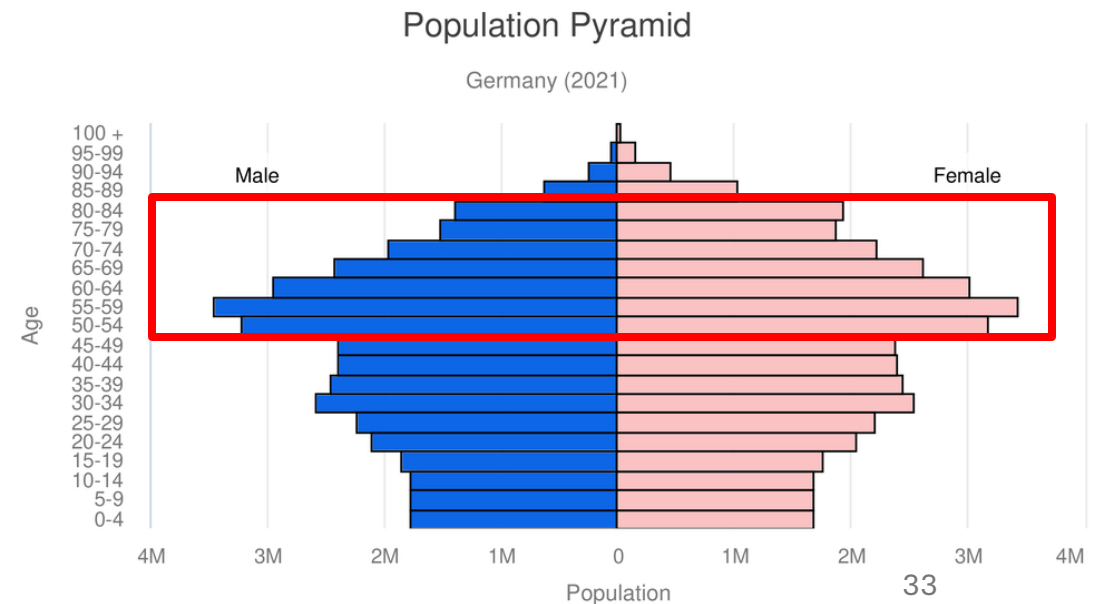


Limitations and Constraints

- Training set contains only **healthy individuals** from the UKB dataset
- Average age of the training set patients is high (**63.6** \pm 7.5 years)
- Semantic understanding of covariables is doubtful



Figure from ANI South Asia, Jan 2025



Summary

- LDMs move the Diffusion Process to latent space
- High-resolution SOTA synthetic MRIs generation using LDMs
- DDIM significantly improves inference time
- Generation conditioned on sex, age, ventricular volume
- Generated dataset released to the public

References

- <https://lilianweng.github.io/posts/2021-07-11-diffusion-models/>
- <https://medium.com/analytics-vidhya/the-world-through-the-eyes-of-cnn-5a52c034dbeb>
- <https://towardsdatascience.com/translational-invariance-vs-translational-equivariance-f9fbc8fca63a>
- <https://towardsdatascience.com/what-are-stable-diffusion-models-and-why-are-they-a-step-forward-for-image-generation-aa1182801d46>
- <https://www.volbrain.net/services/AssemblyNet-AD>
- <https://www.news-medical.net/news/20240816/Scientists-discover-key-age-related-biological-shifts-at-40-and-60.aspx>
- https://www.radiologymasterclass.co.uk/tutorials/ct/ct_acute_brain/ct_brain_ventricles_sulci
- <https://raditechies.blogspot.com/p/positioning-terminology.html>
- https://www.researchgate.net/publication/387444028_Comparative_Analysis_of_GANs_and_Diffusion_Models_in_Image_Generation
- : <https://blog.marvik.ai/2023/11/28/an-introduction-to-diffusion-models-and-stable-diffusion/>
- <https://www.lead-dbs.org/about-the-mni-spaces/>
- <https://www.aninews.in/news/health/rare-diseases-lead-to-poor-quality-of-life-in-people201712200631200001/>
- <https://medium.com/@steinsfu/stable-diffusion-clearly-explained-ed008044e07e>
- Satheesh, Sk & Moorthy, K.Krishna. (2013). AEROSAT - a space-borne sensor for continental aerosols: Evaluation of the conceptual model. *Current Science*. 104. 1008-1011.
- Song, J., Meng, C., & Ermon, S. (2021). Denoising Diffusion Implicit Models. *International Conference on Learning Representations (ICLR)*.