Imperial College London

Improving Diversity of Diffusion Models using Particle Methods

Alexander Pondaven

27th June 2023

Supervisor
Dr. Yingzhen Li

Second marker
Prof. Ruth Misener

PhD Advisor
Harrison Zhu

Are diffusion models diverse?

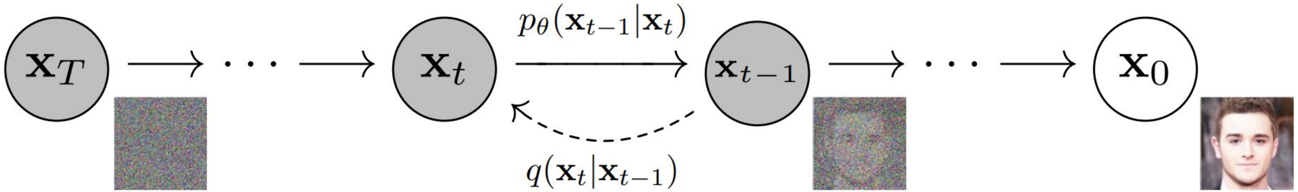


Diagram: Ho et al. (2020)

Quality

Speed

Diversity

Novel contributions

- Introduce repulsion methods
  - Increased spread of images in controlled fashion
- Diversity metrics
  - General diversity
  - Location diversity
  - Style diversity
- Evaluate repulsion methods on each metric and show improved spread

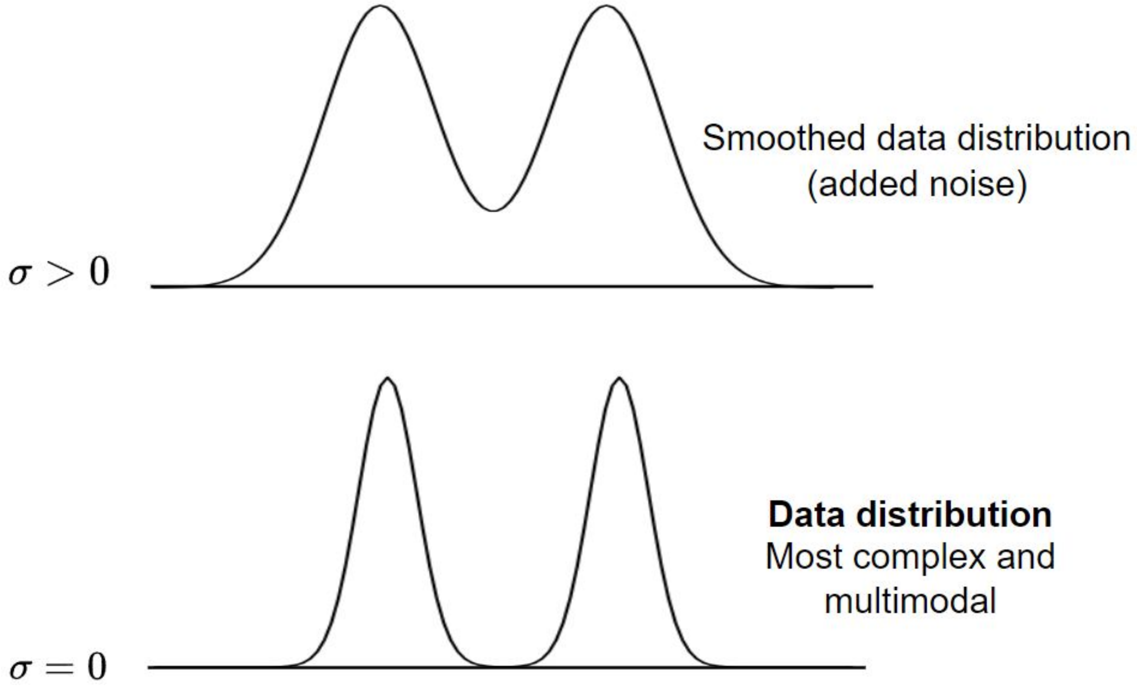
Noise levels

- Construct marginals

Data distribution
Most complex and multimodal

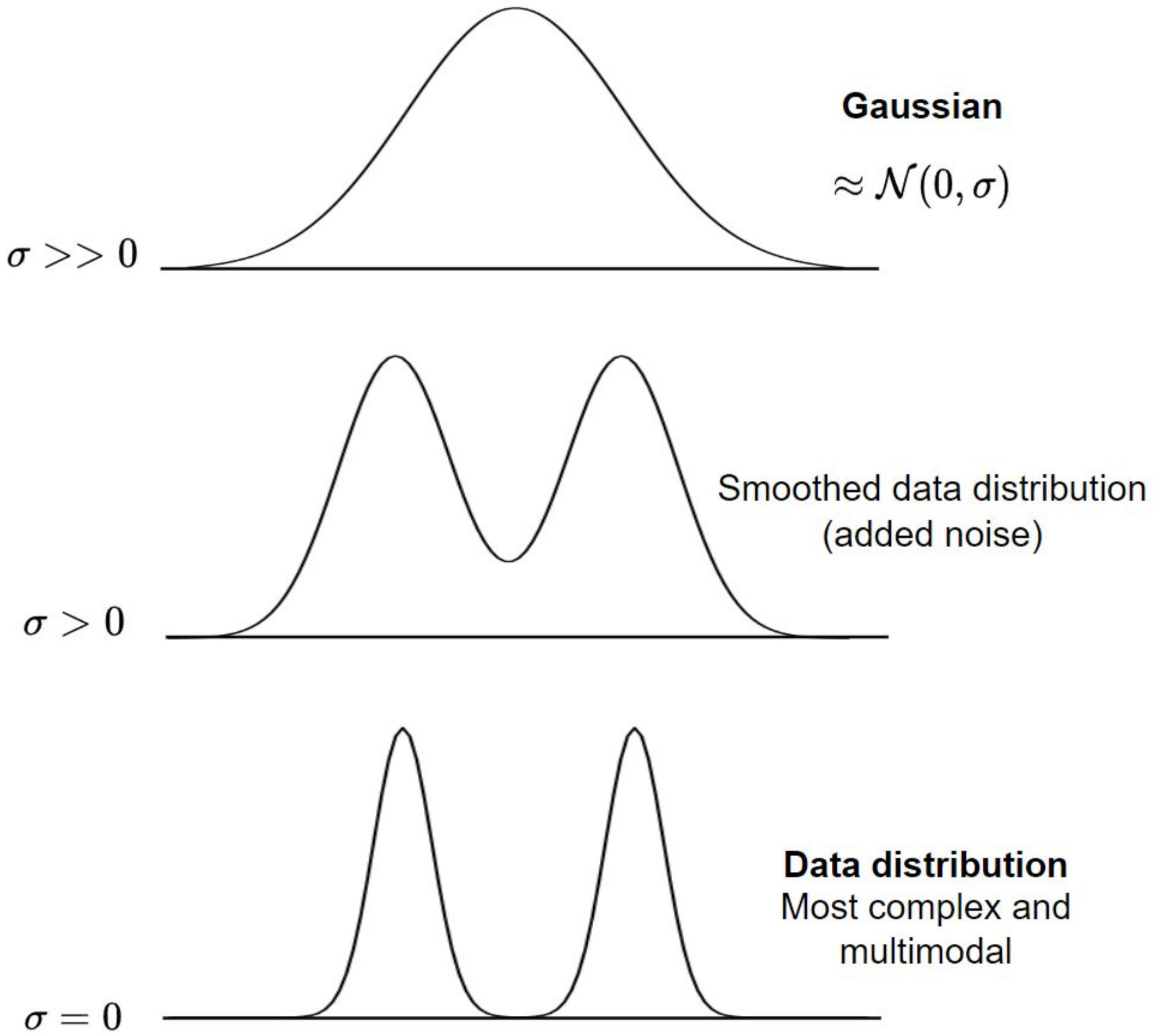
Noise levels

- Construct marginals



Noise levels

- Construct marginals

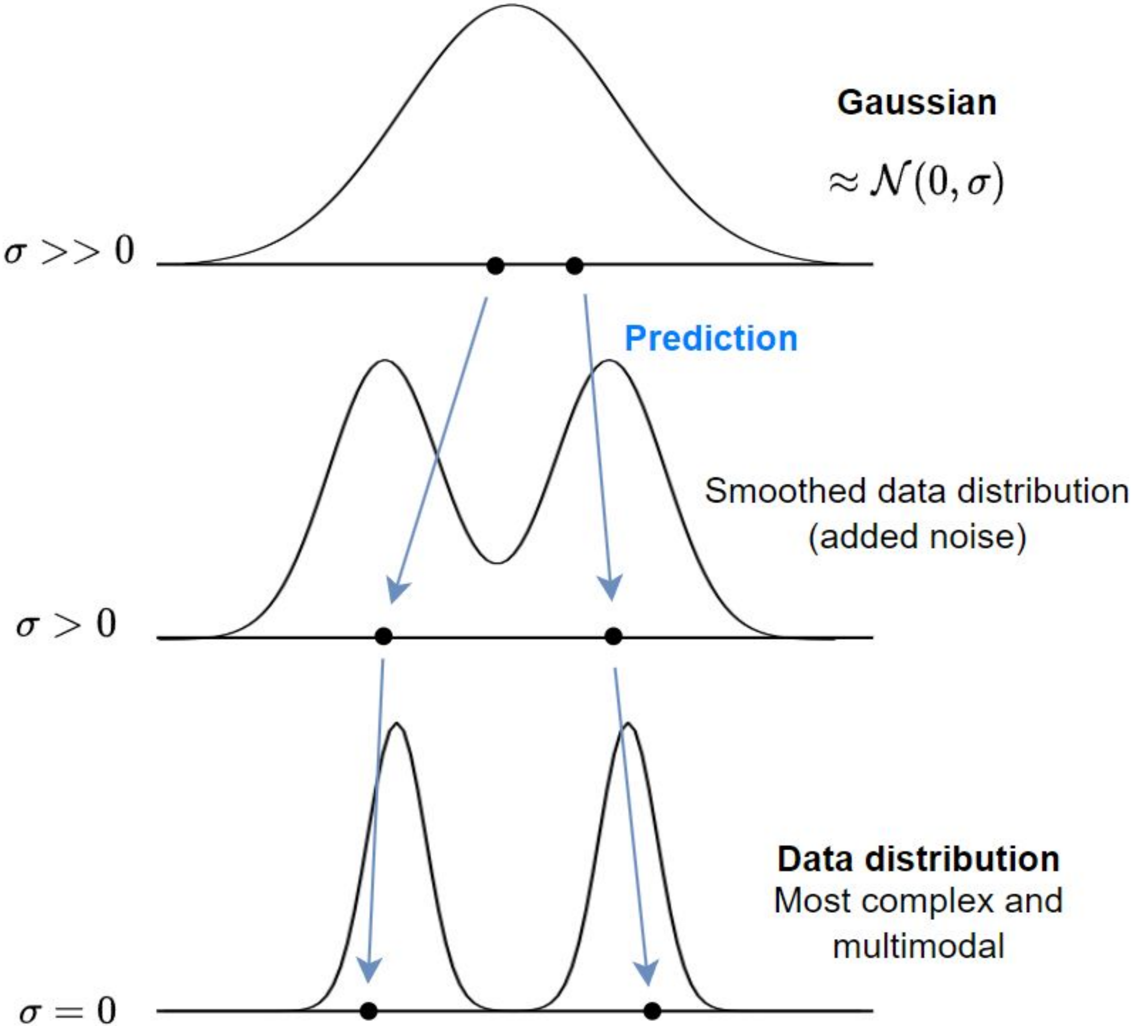


Noise levels

- Construct marginals

Types of steps

1) **Prediction steps**

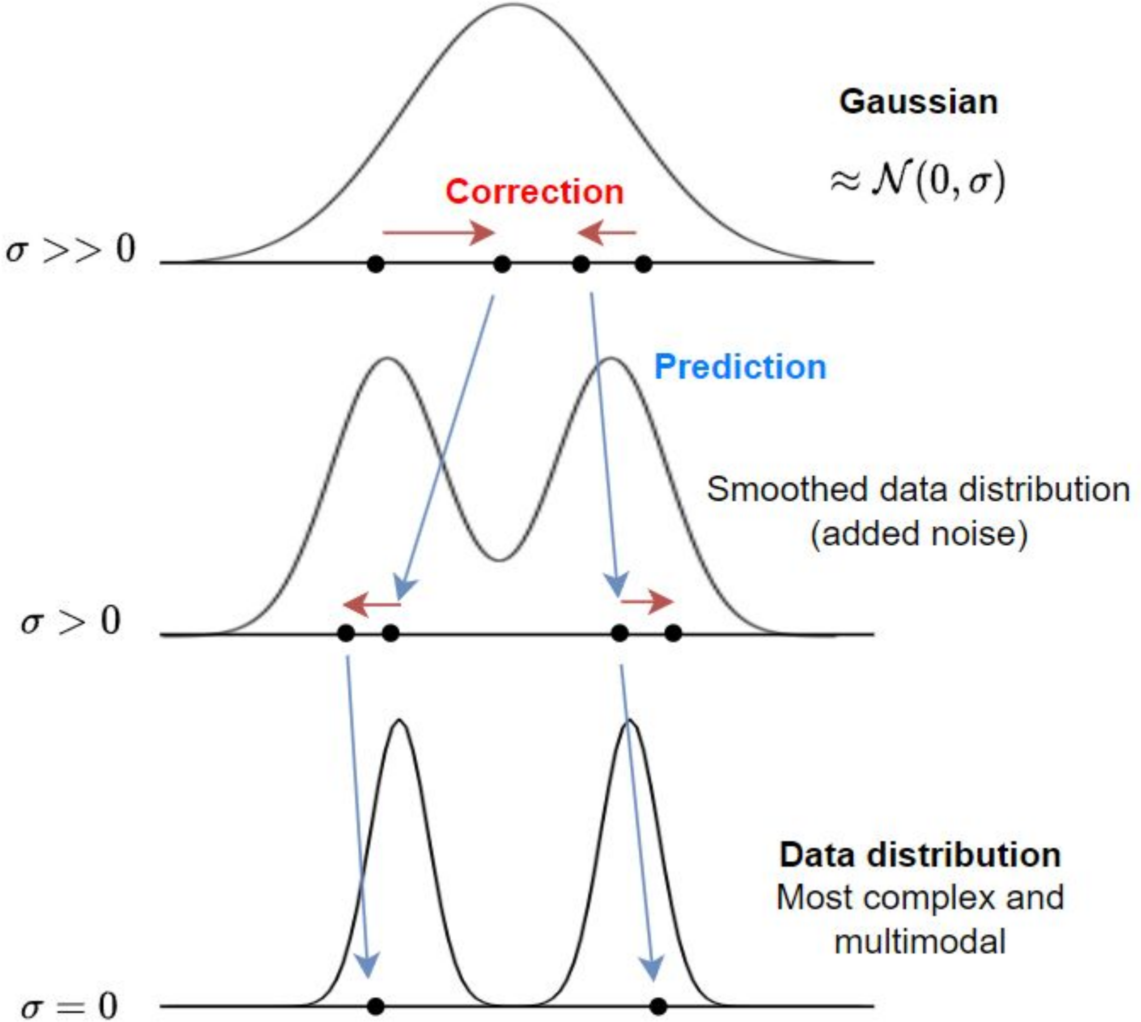


Noise levels

- Construct marginals

Types of steps

1) Prediction steps
2) Correction steps

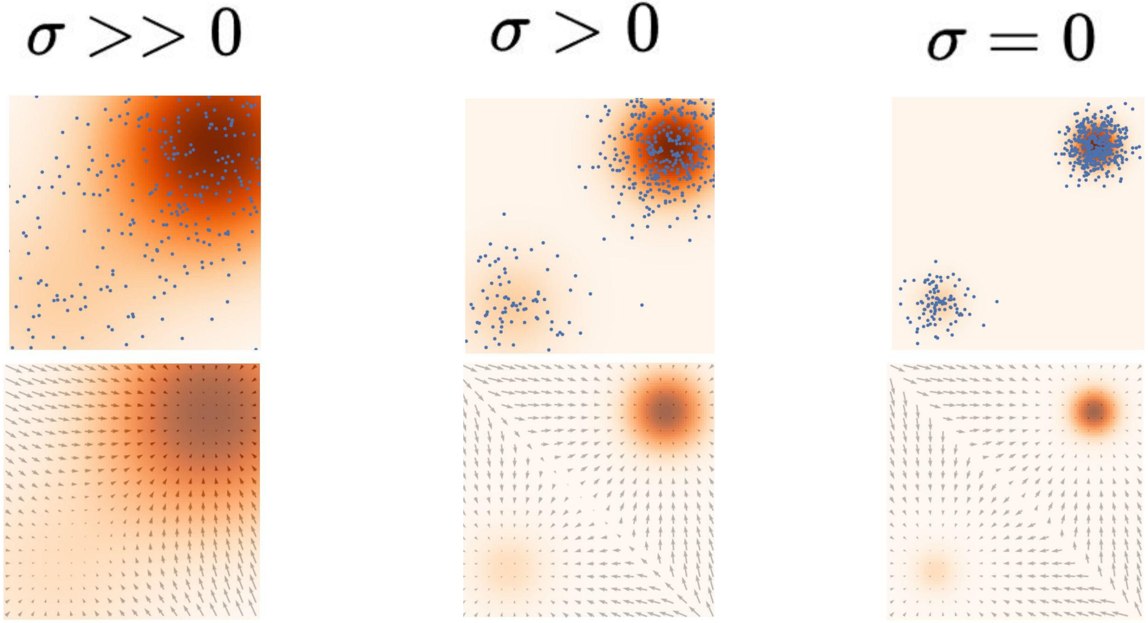


Sampling from $p(x)$

Samples

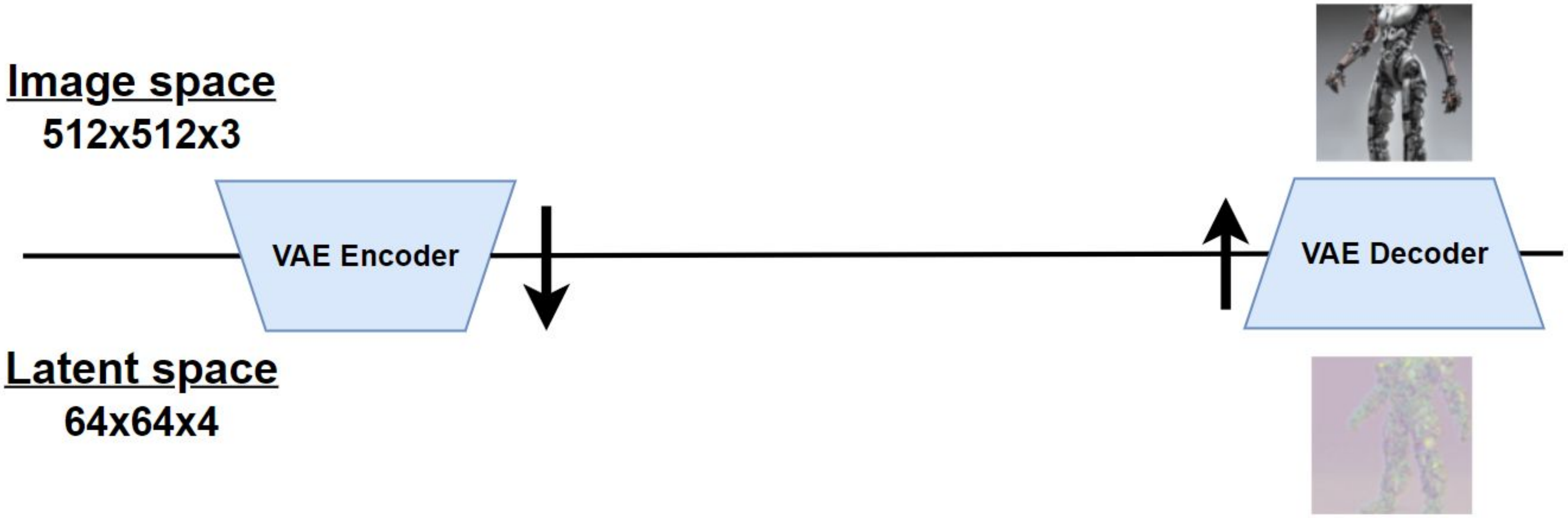
Scores

$\nabla \log p(x)$

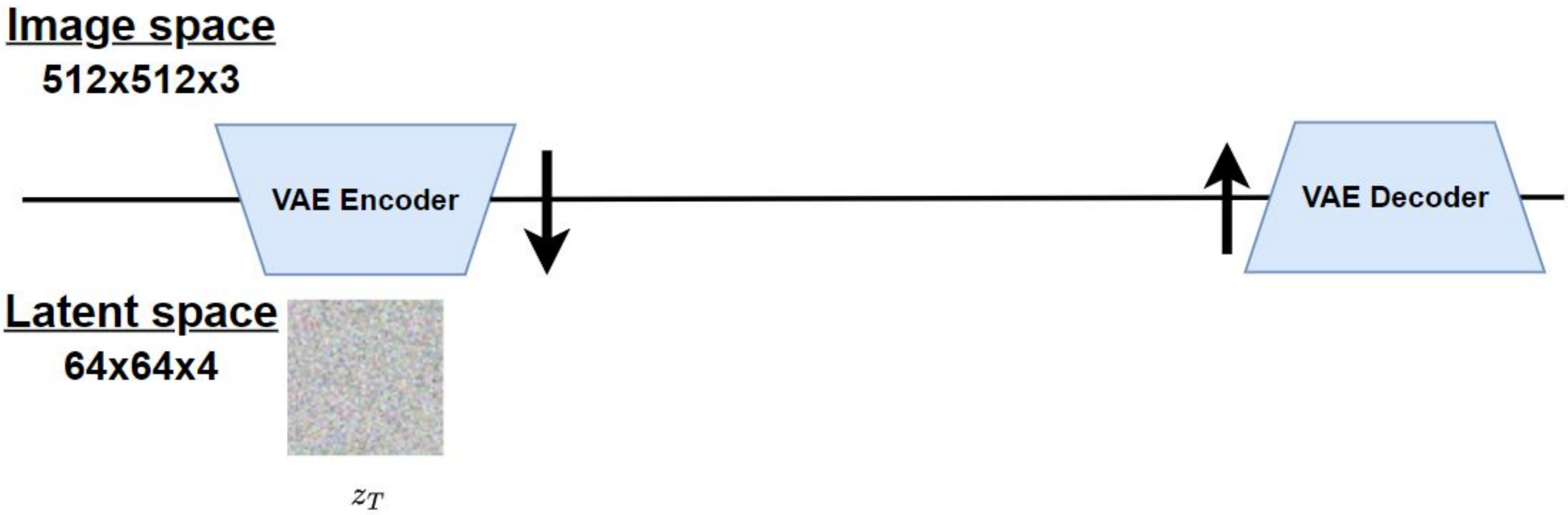


Diagrams by Song (2021)

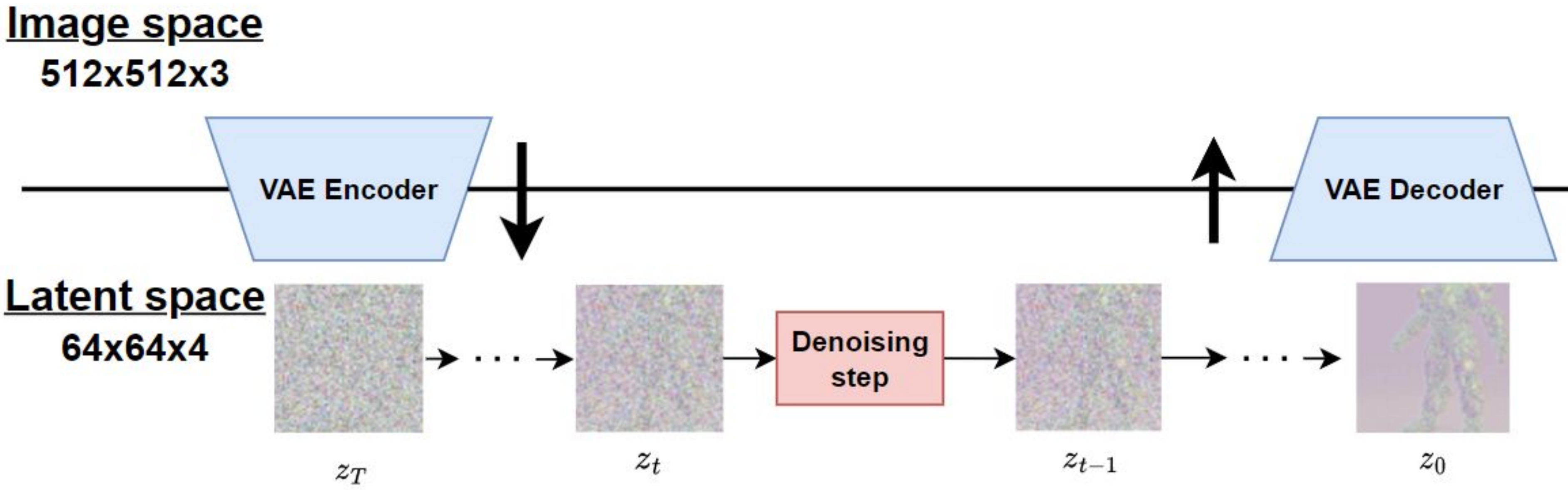
Stable Diffusion 2



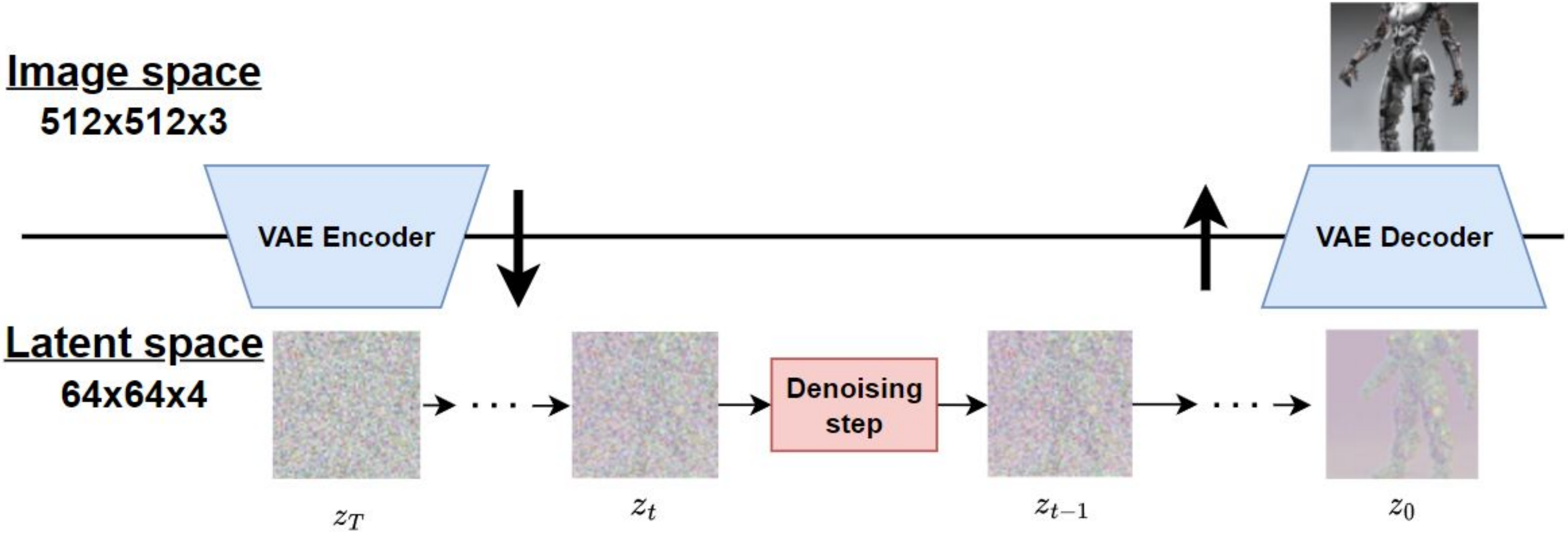
Stable Diffusion 2



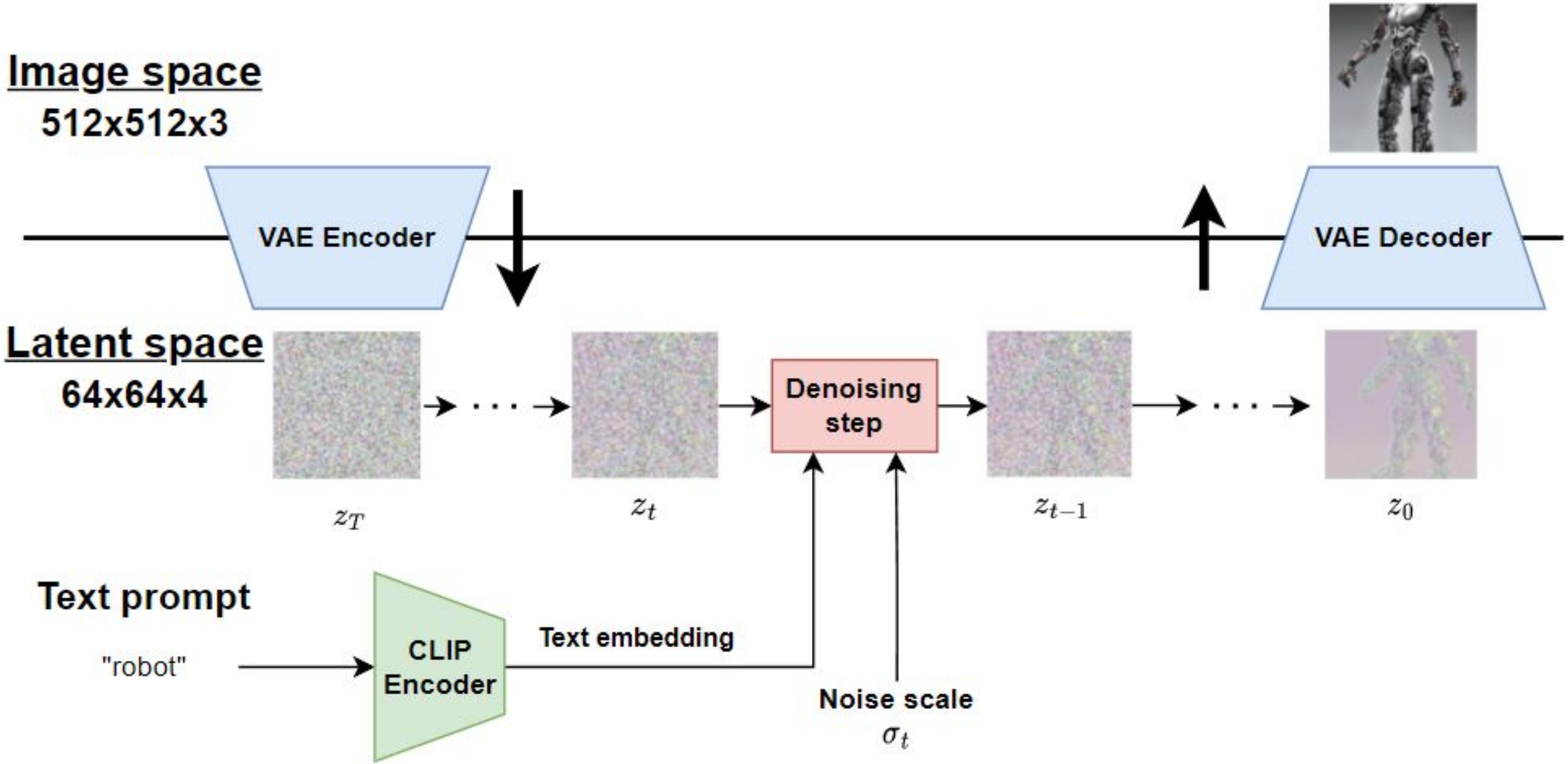
Stable Diffusion 2



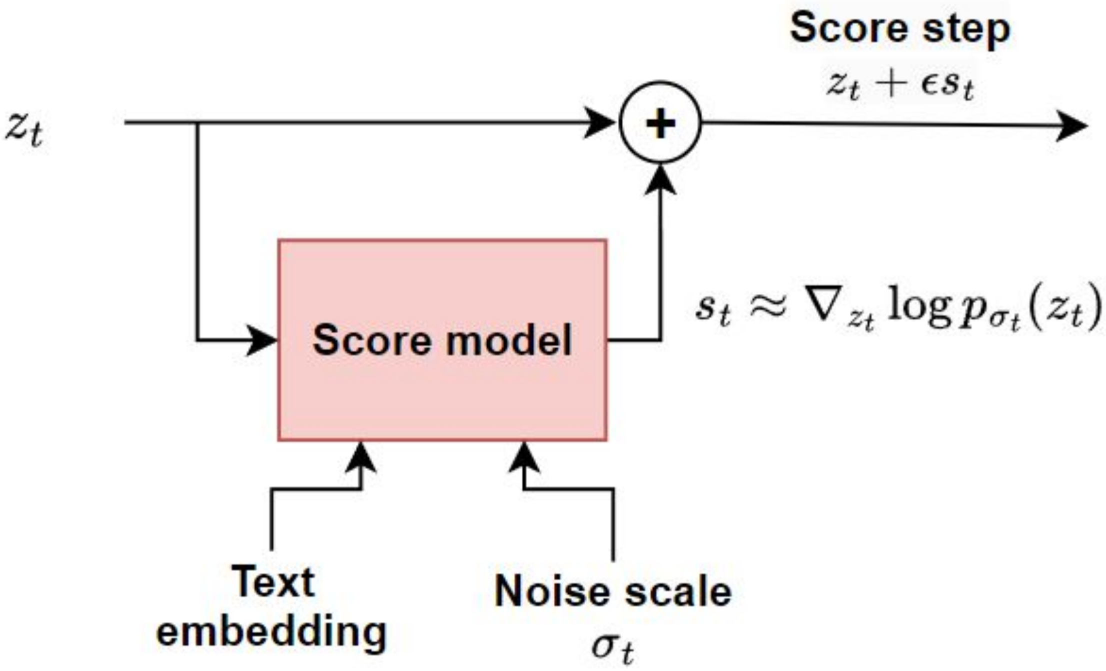
Stable Diffusion 2



Stable Diffusion 2



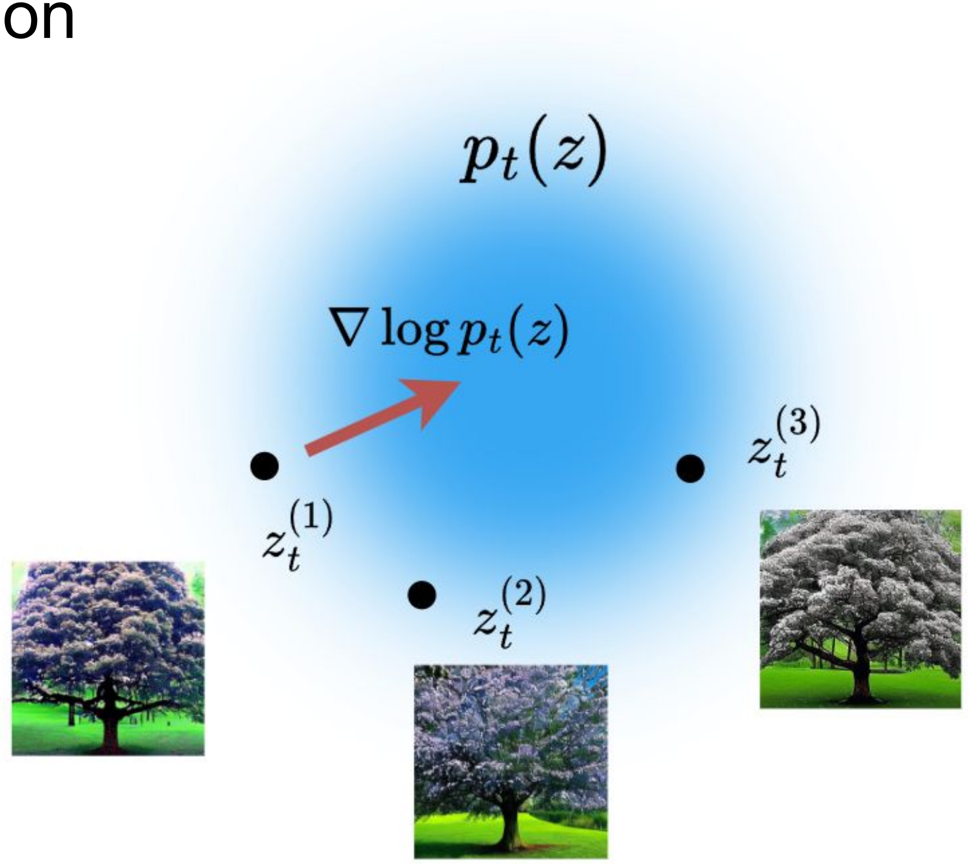
Denoising step



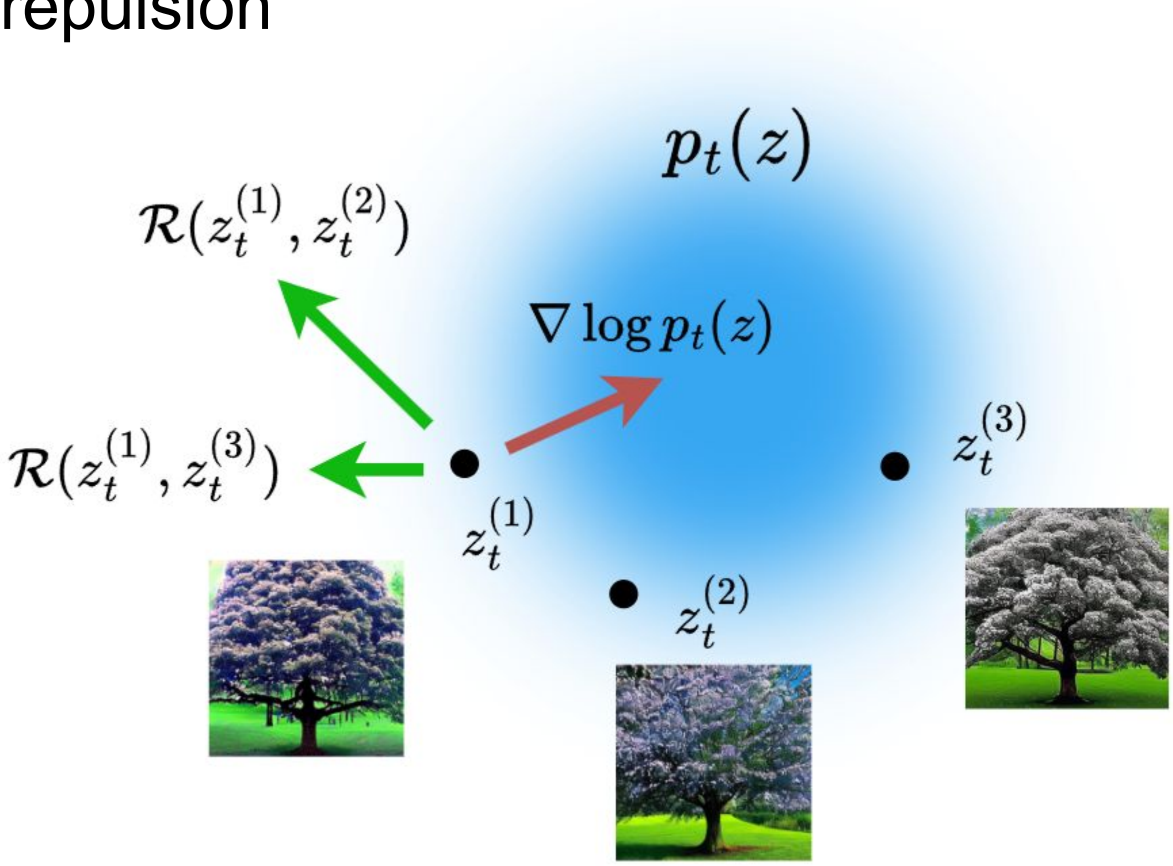
$$z_{t+1}^{(i)} = z_t^{(i)} + \epsilon_t \nabla_{z_t^{(i)}} \log p(z_t^{(i)})$$

Repulsion

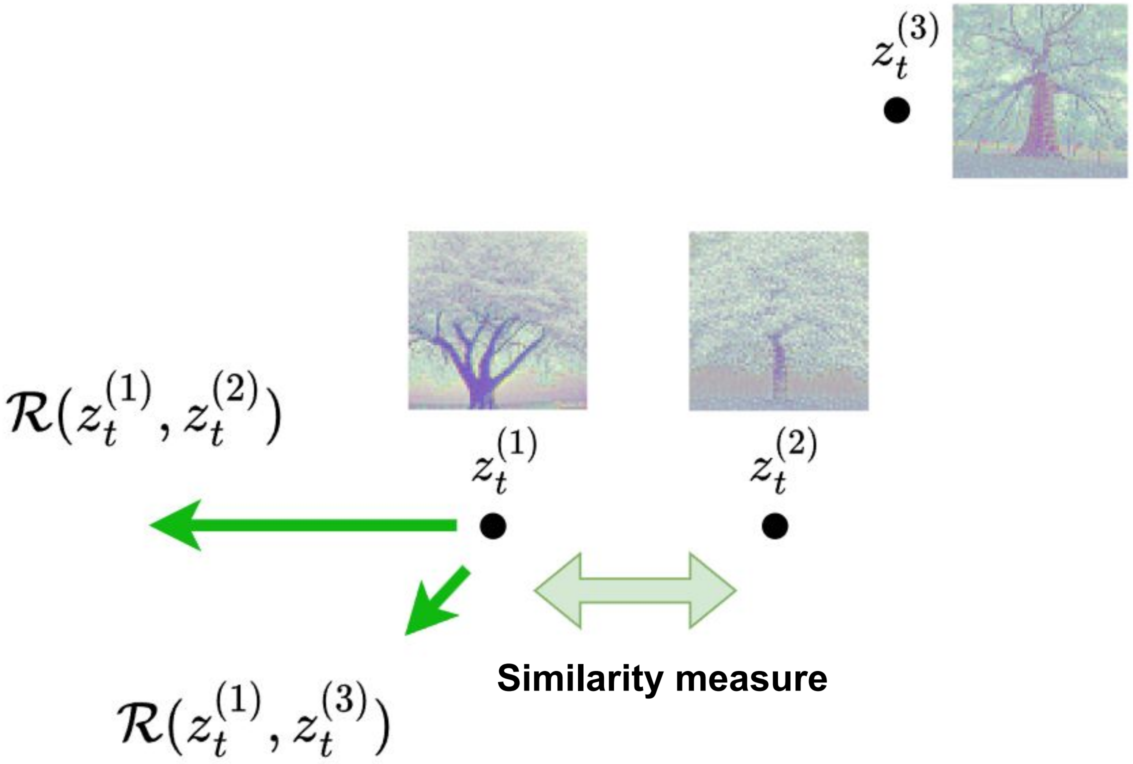
Particle repulsion



Particle repulsion



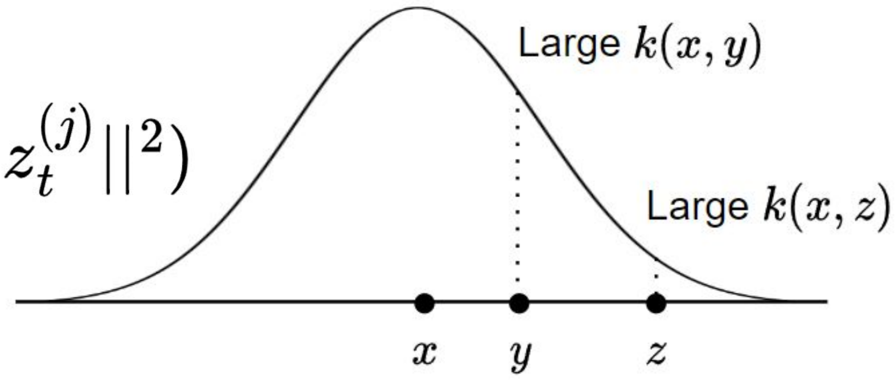
Similarity repulsion



Kernel gradient repulsion term

- RBF kernel

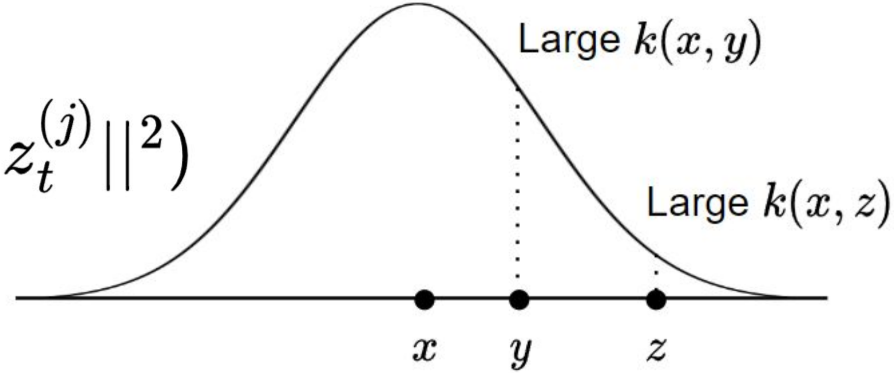
$$k(z_t^{(i)}, z_t^{(j)}) = exp(-\frac{1}{h}||z_t^{(i)} - z_t^{(j)}||^2)$$



Kernel gradient repulsion term

- RBF kernel

$$k(z_t^{(i)}, z_t^{(j)}) = exp(-\frac{1}{h}||z_t^{(i)} - z_t^{(j)}||^2)$$



- Take gradient

$$\mathcal{R}(z_t^{(i)}, z_t^{(j)}) = \nabla_{z_t^{(i)}} k(z_t^{(i)}, z_t^{(j)})$$

Repulsive step

- Add repulsive forces for N particles
- Repulsive force α

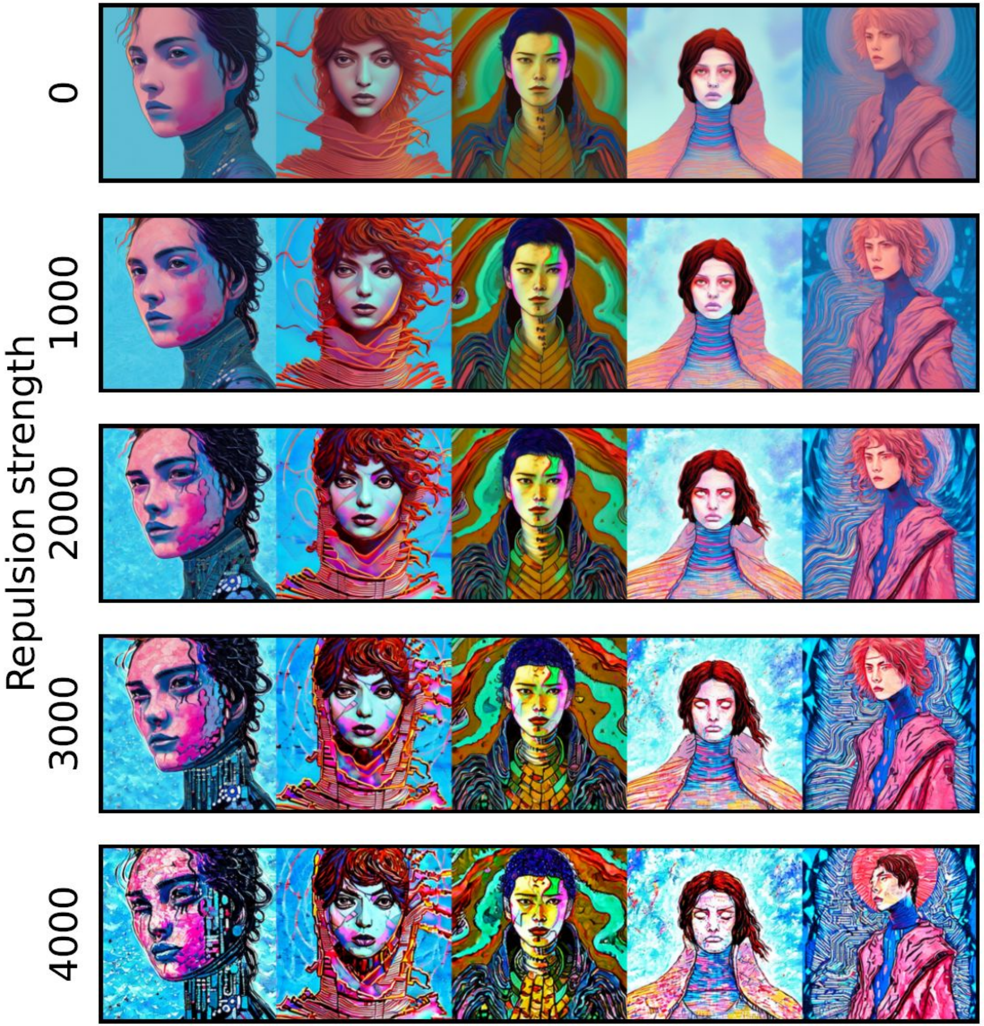
Denoising step: Score

$$z_{t+1}^{(i)} = z_t^{(i)} + \epsilon_t \nabla_{z_t^{(i)}} \log p(z_t^{(i)})$$

Denoising step: Score + Repulsion

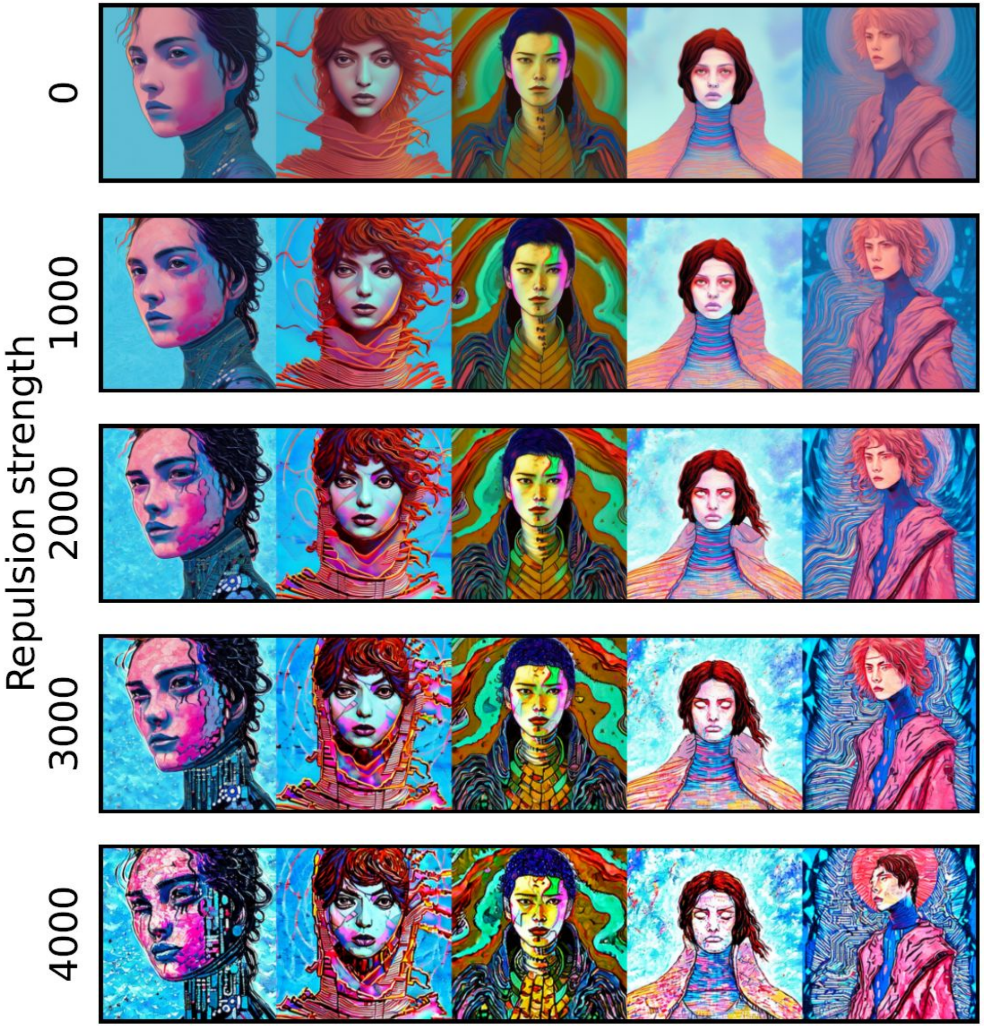
$$z_{t+1}^{(i)} = z_t^{(i)} + \epsilon_t [\nabla_{z_t^{(i)}} \log p(z_t^{(i)}) - \frac{\alpha}{N} \sum_{j=1}^{N} \nabla_{z_t^{(i)}} k(z_t^{(i)}, z_t^{(j)})]$$

Latent repulsion



Latent repulsion

- High dimensionality issue
- Unsuitable similarity measure



Feature similarity measure

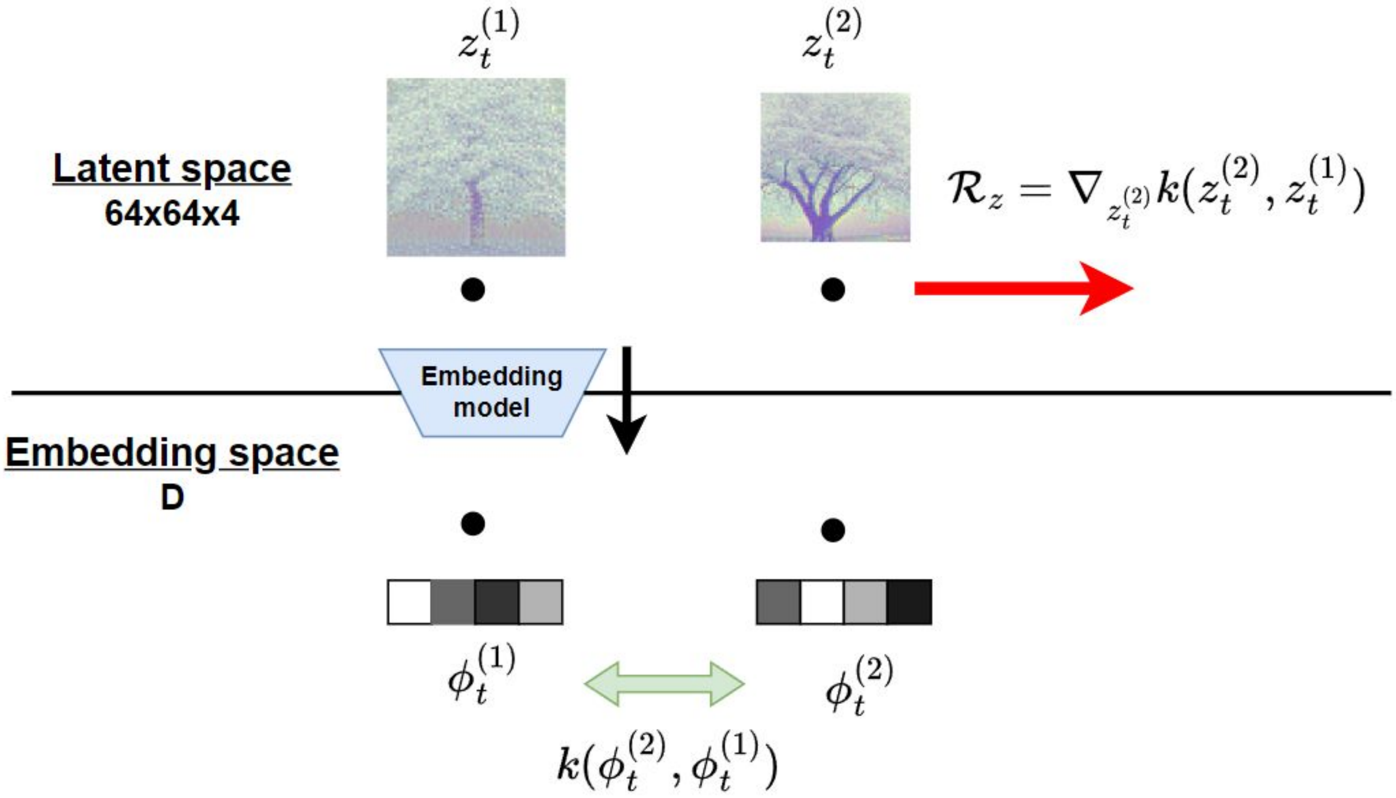
Latent space
64x64x4

$z_t^{(1)}$

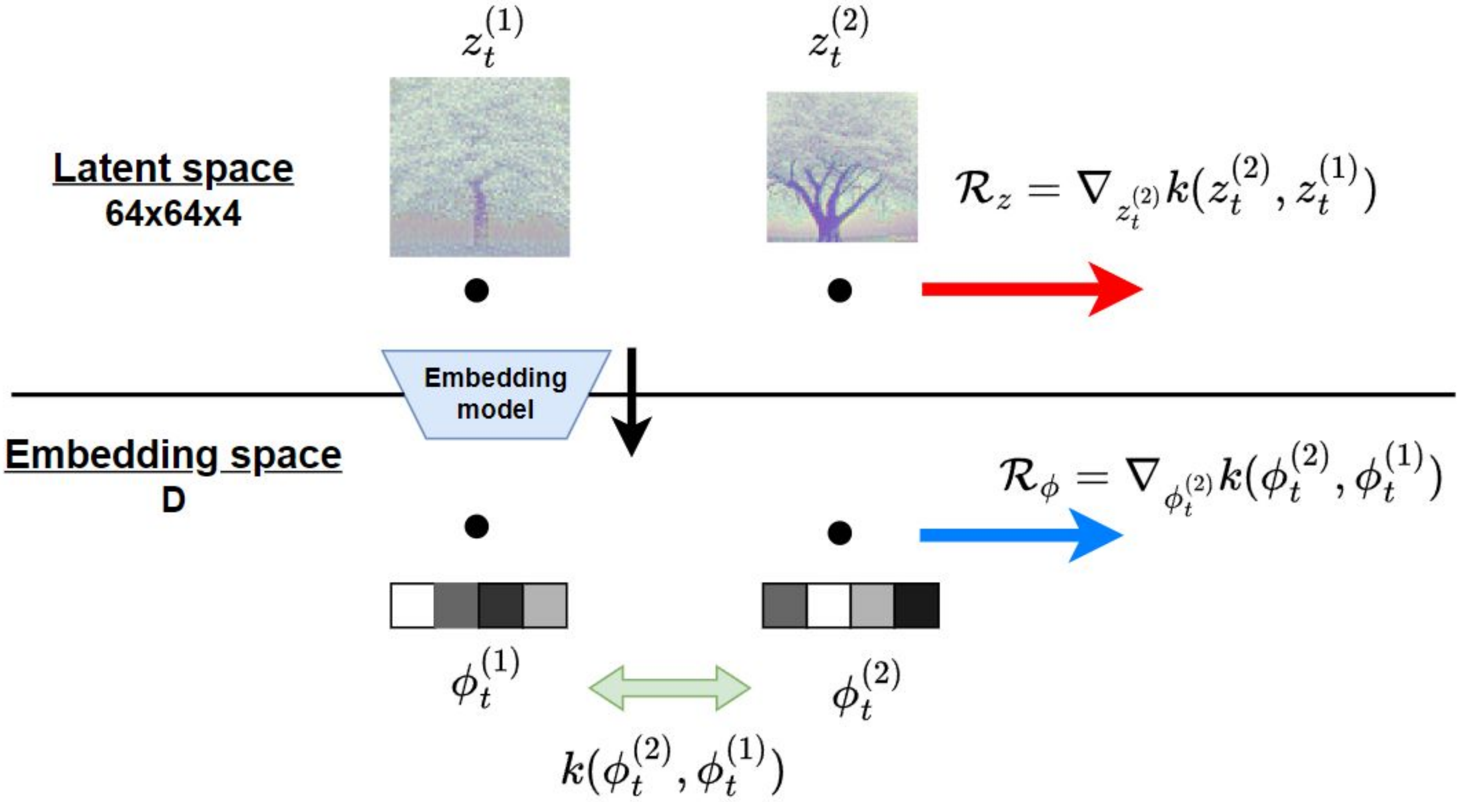
$z_t^{(2)}$

$$\mathcal{R}_z = \nabla_{z_t^{(2)}} k(z_t^{(2)}, z_t^{(1)})$$

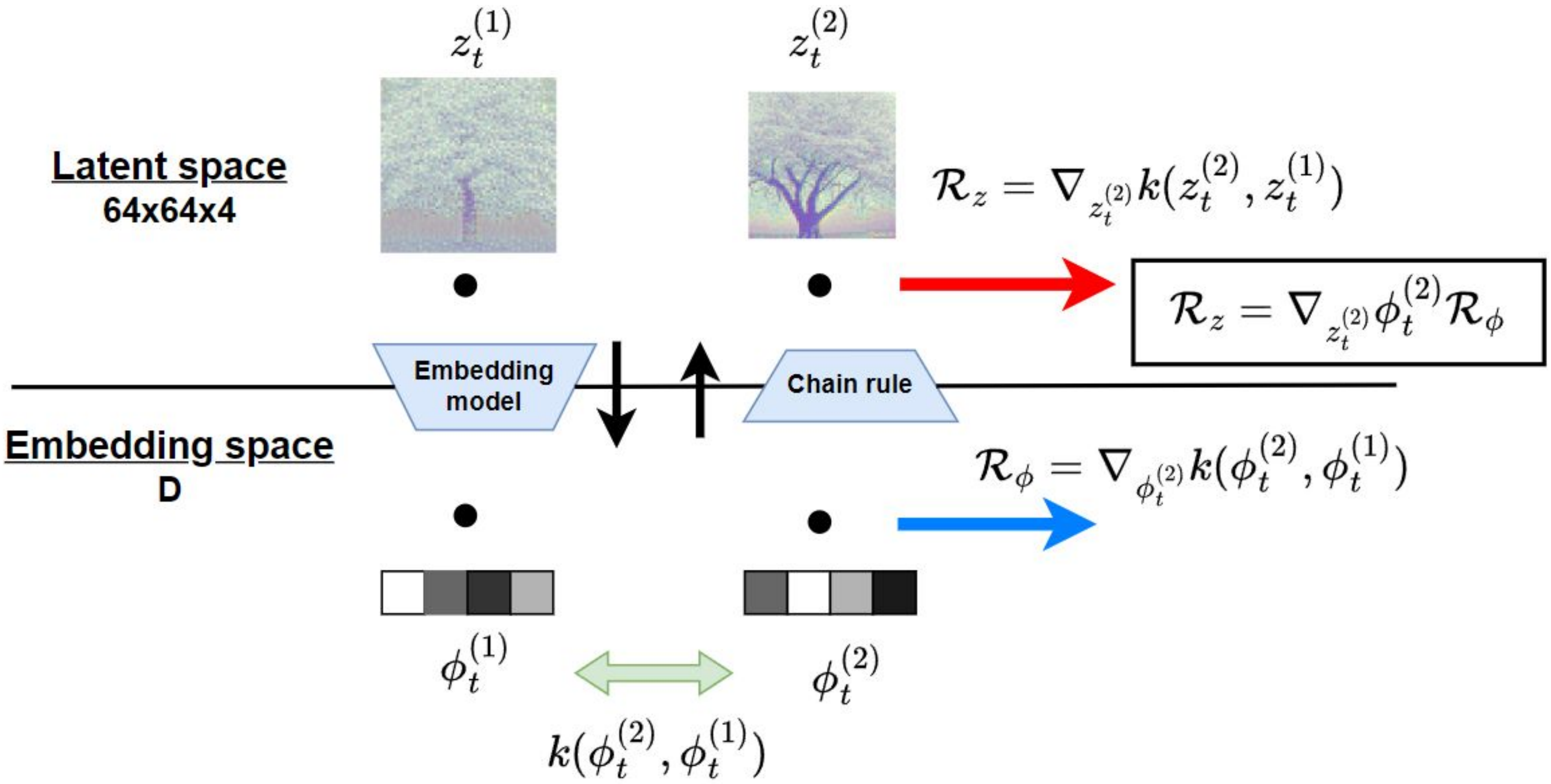
Feature similarity measure



Feature similarity measure



Feature similarity measure



Feature Repulsive steps

Latent repulsive step

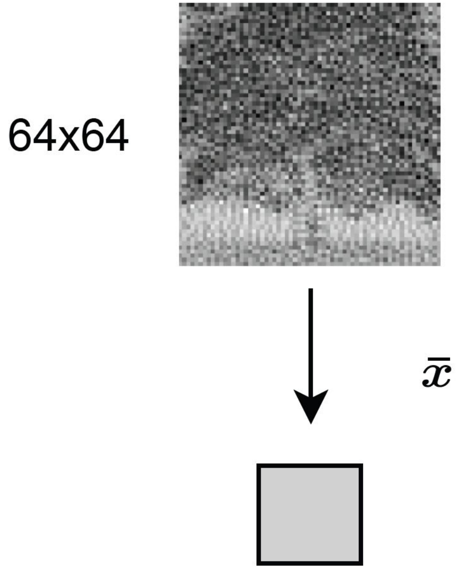
$$z_{t+1}^{(i)} = z_t^{(i)} + \epsilon_t [\nabla_{z_t^{(i)}} \log p(z_t^{(i)}) - \frac{\alpha}{N} \sum_{j=1}^{N} \nabla_{z_t^{(i)}} k(z_t^{(i)}, z_t^{(j)})]$$

Embedded repulsive step

$$z_{t+1}^{(i)} = z_t^{(i)} + \epsilon_t [\nabla_{z_t^{(i)}} \log p(z_t^{(i)}) - \frac{\alpha}{N} \sum_{j=1}^{N} \nabla_{z_t^{(i)}} \phi_t^{(i)} \nabla_{\phi_t^{(i)}} k(\phi_t^{(i)}, \phi_t^{(j)})]$$

Channel average repulsion

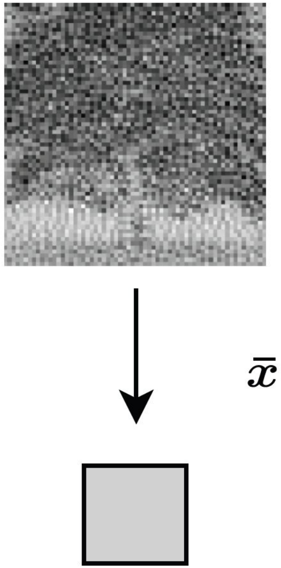
Average of each latent channel



Channel average repulsion

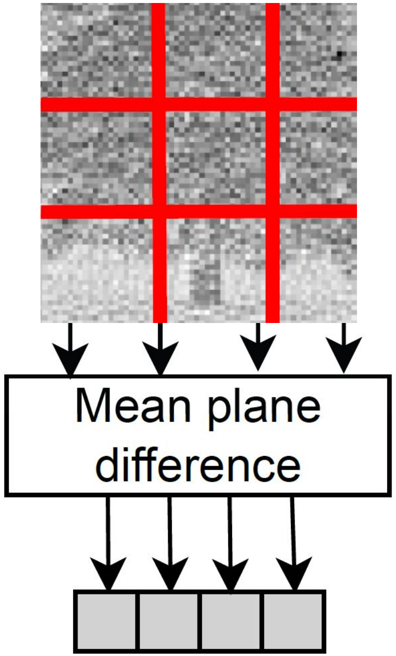
Average of each latent channel

64x64

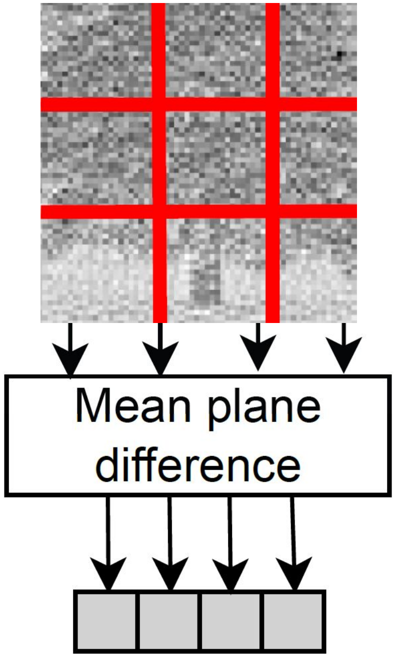


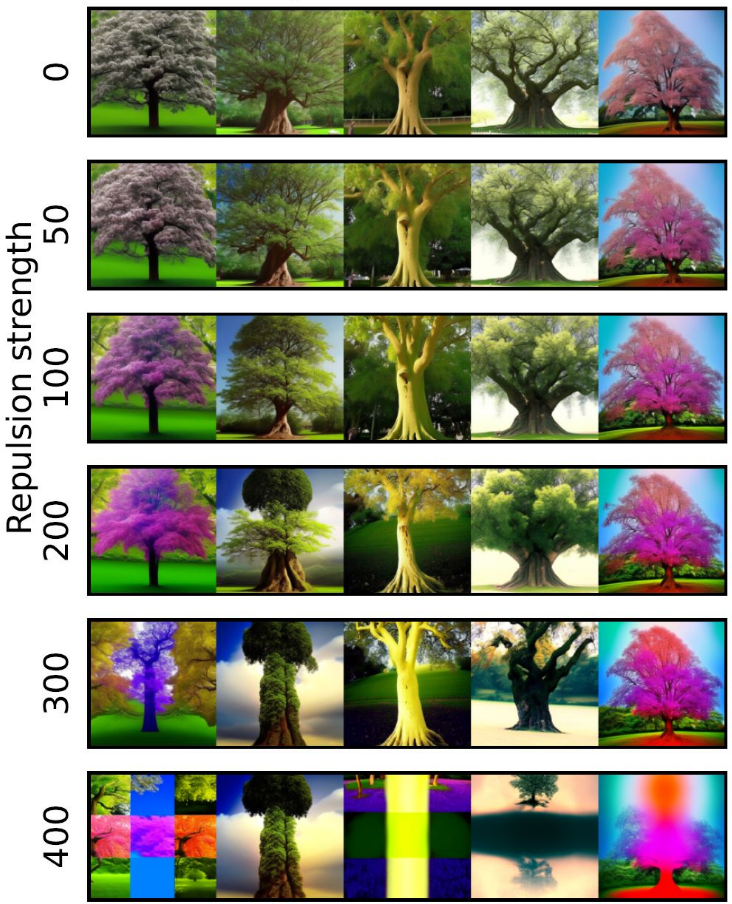
$\bar{x}$

Rule of thirds repulsion



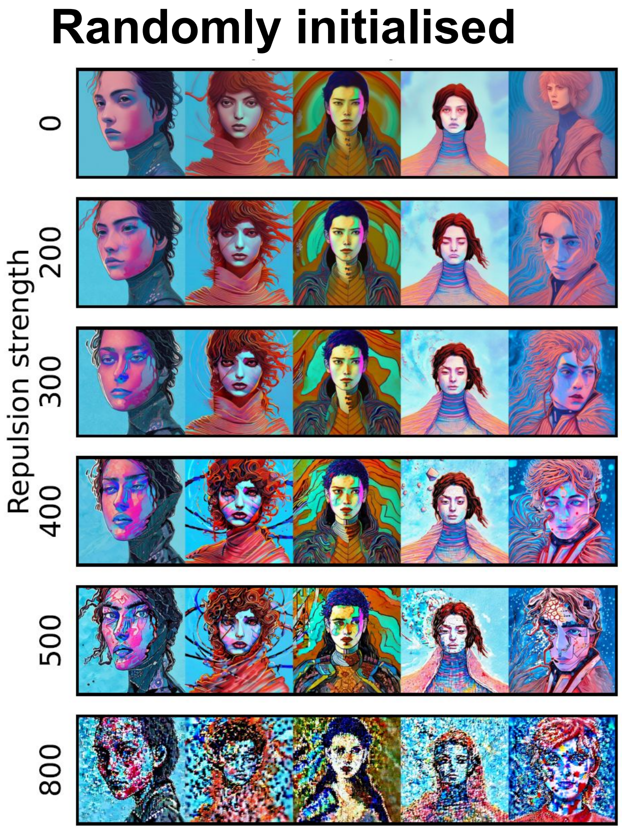
Rule of thirds repulsion





Convolutional Neural Network (CNN)

- Image property extractor
- Randomly initialise weights



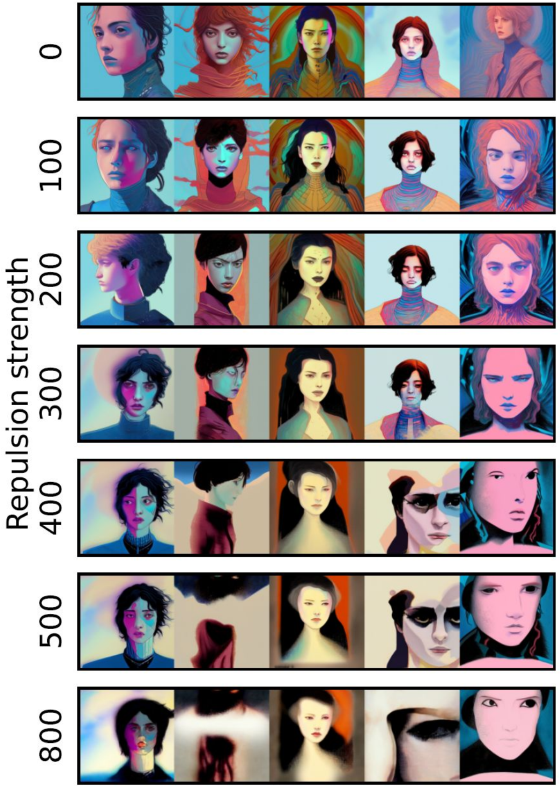
Style classifier trained on latents

- Train CNN to classify artist from latents

Style classifier trained on latents

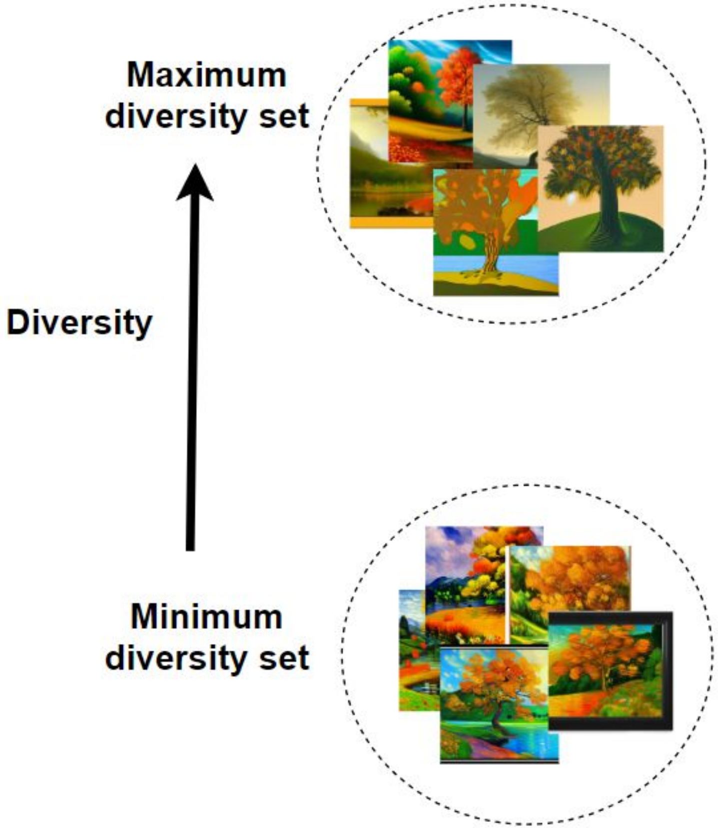
- Rule of thirds operation on style classifier embedding

Rule of thirds on feature space

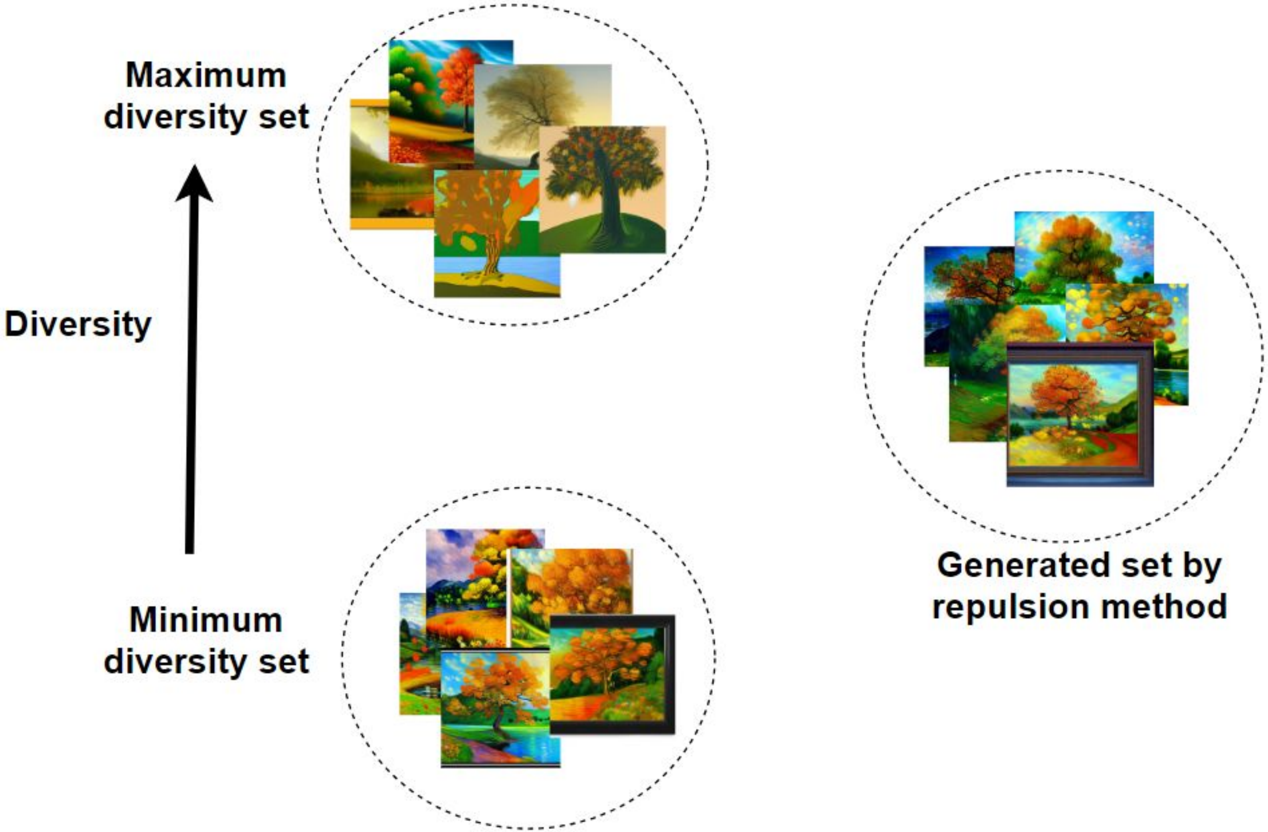


Diversity metrics

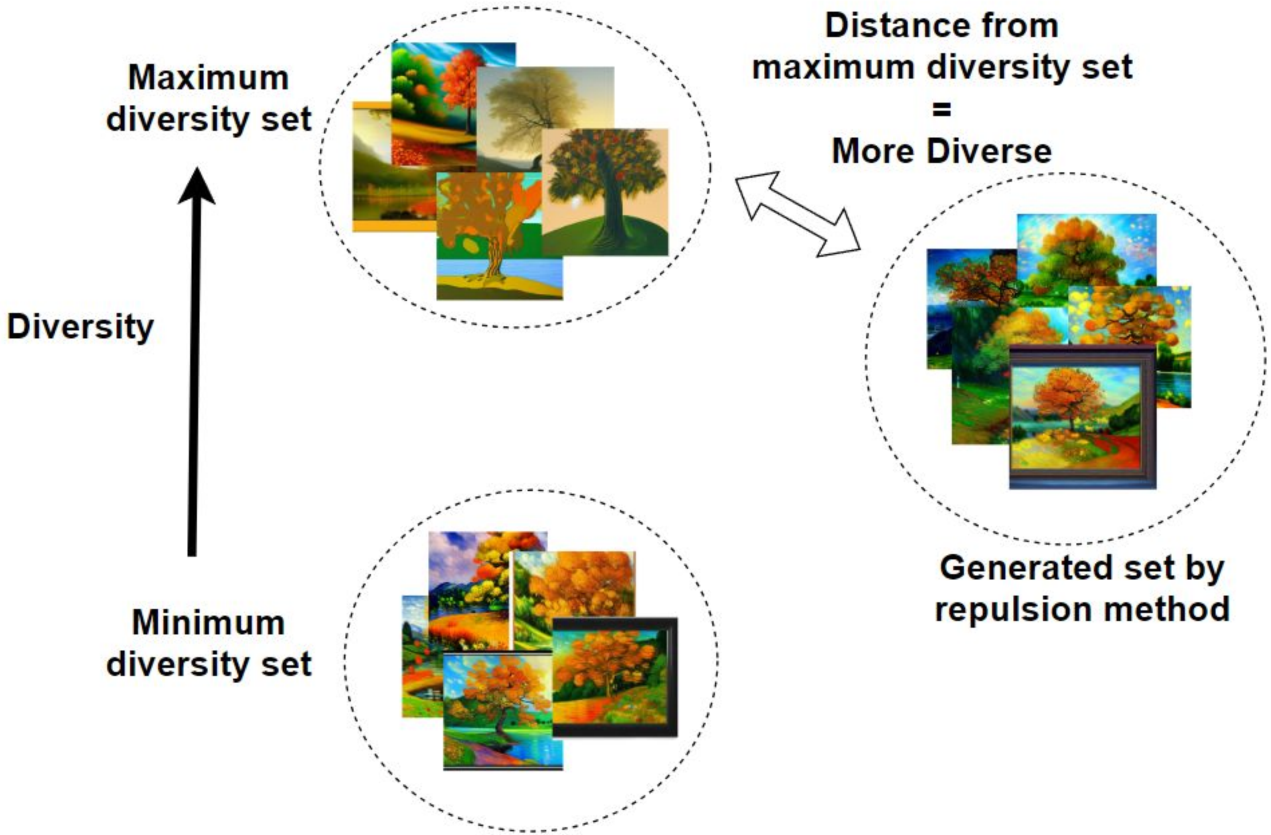
Maximum diversity distance



Maximum diversity distance



Maximum diversity distance

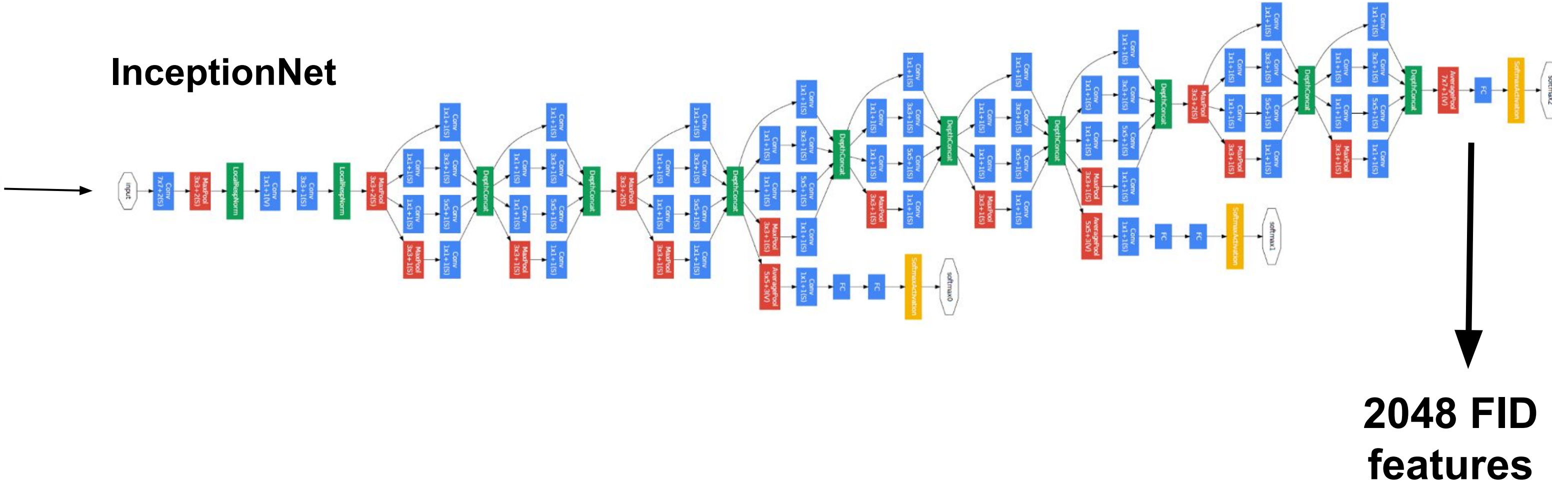


Frechet Inception Distance (FID) Feature vectors

- Compare image embeddings with FID
  - Default FID embedding

Frechet Inception Distance (FID) Feature vectors

- Compare image embeddings with FID
  - Default FID embedding

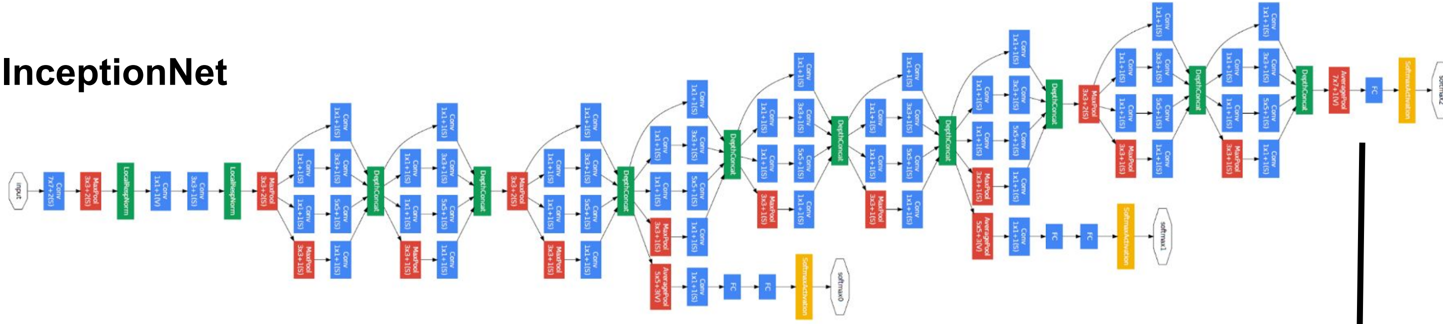


Model diagram:
JalFaizy Shaikh. `https://www.analyticsvidhya.com/blog/2018/10/understanding-inception-network-from-scratch/`. [Accessed 16th June 2023].

Frechet Inception Distance (FID) Feature vectors

- Compare image embeddings with FID
  - Default FID embedding
  - Introduce novel **location** and **style** embeddings

InceptionNet

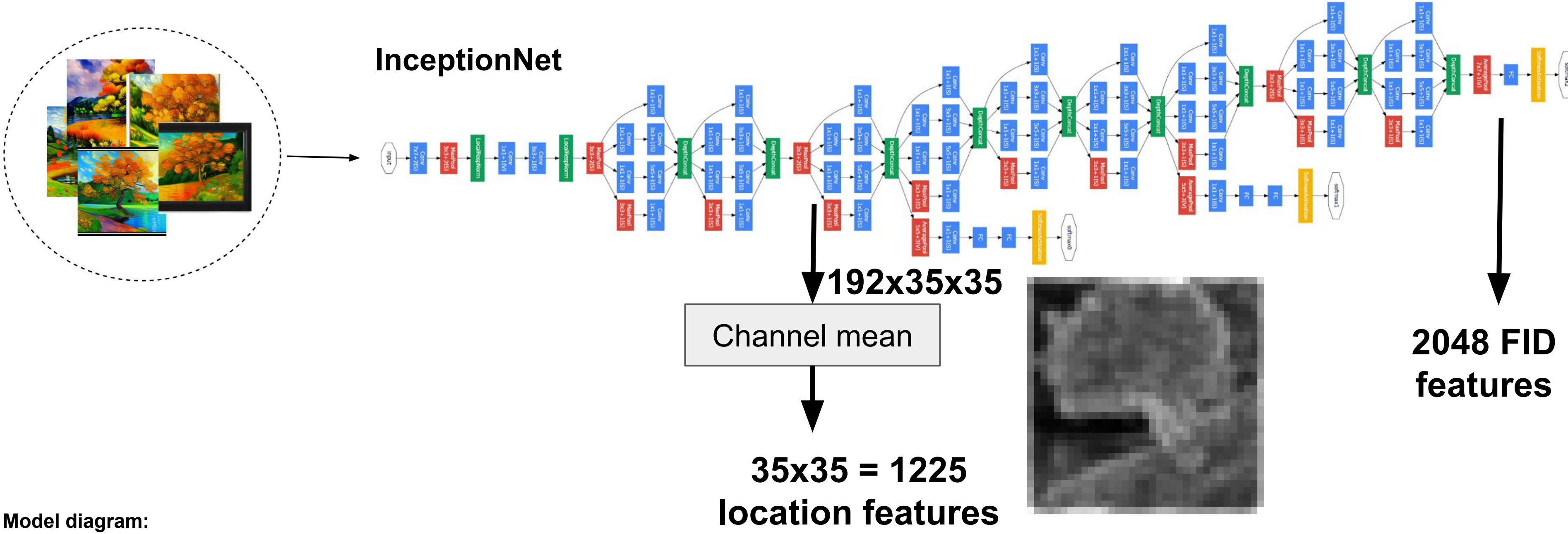


2048 FID features

Model diagram:
JalFaizy Shaikh. `https://www.analyticsvidhya.com/blog/2018/10/understanding-inception-network-from-scratch/`. [Accessed 16th June 2023].

Frechet Inception Distance (FID) Feature vectors

- Compare image embeddings with FID
  - Default FID embedding
  - Introduce novel **location** and **style** embeddings

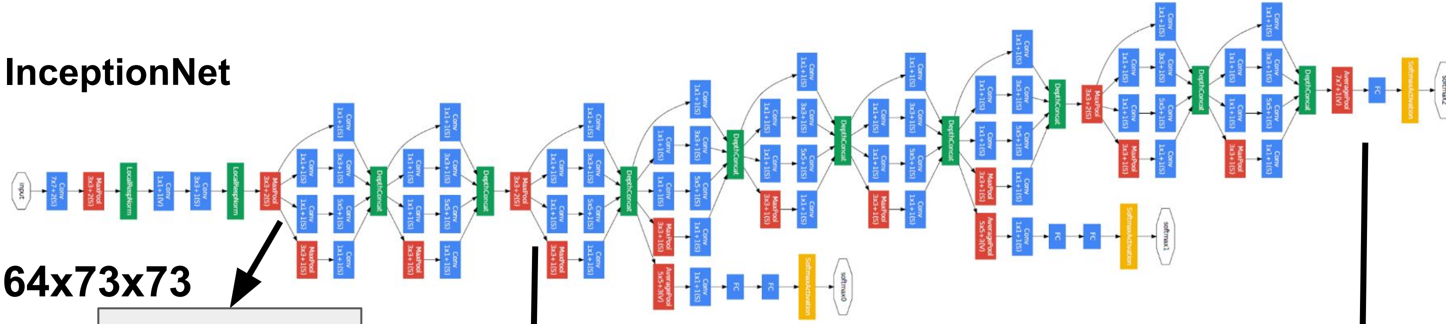


Model diagram:
JalFaizy Shaikh. `https://www.analyticsvidhya.com/blog/2018/10/understanding-inception-network-from-scratch/`. [Accessed 16th June 2023].

Frechet Inception Distance (FID) Feature vectors

- Compare image embeddings with FID
  - Default FID embedding
  - Introduce novel **location** and **style** embeddings

InceptionNet



64x73x73

Spatial mean

64

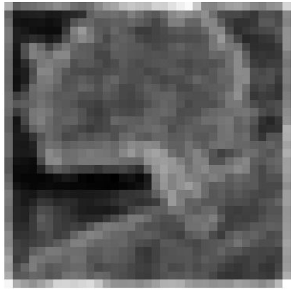
Gram matrix

2080 style features

192x35x35

Channel mean

35x35 = 1225
location features



2048 FID
features

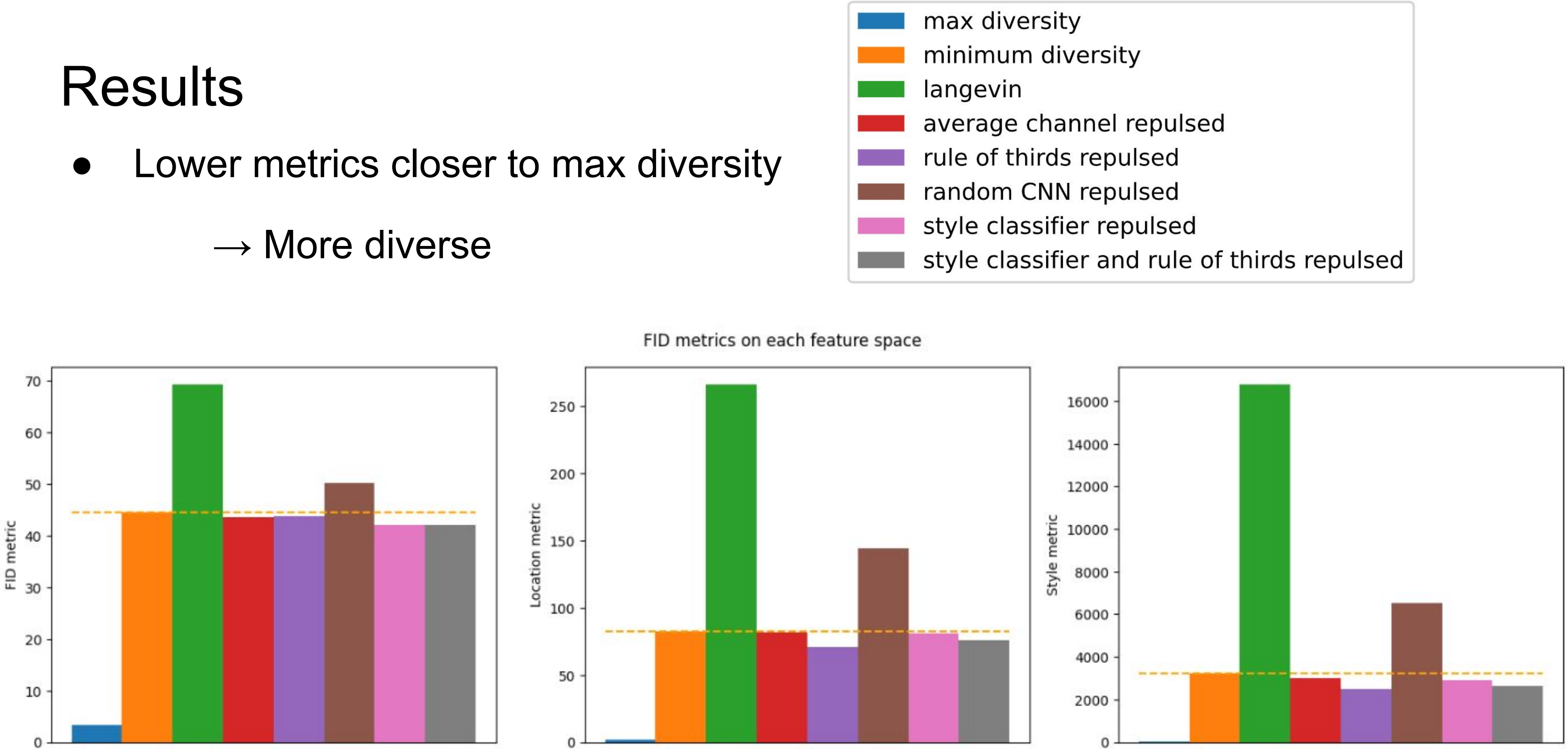
Model diagram:
JalFaizy Shaikh. `https://www.analyticsvidhya.com/blog/2018/10/understanding-inception-network-from-scratch/`. [Accessed 16th June 2023].

Evaluation results

Results

- Lower metrics closer to max diversity

  → More diverse



Conclusion

- Improved efficiency and control over exploration of image space
- Novel diversity metrics to separate location and style

Conclusion

- Improved efficiency and control over exploration of image space
- Novel diversity metrics to separate location and style

Applications

- Flexible sampling of pre-trained models
- Reduce redundancy e.g. recommendation systems

References

- **Stable Diffusion 2 Paper:** Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Bj¨orn Ommer. High-resolution image synthesis with latent diffusion models, 2021. URL https://arxiv.org/abs/2112.10752.
- **Stable Diffusion 2 Hugging Face API**, Jun 2022. URL https://huggingface.co/stabilityai/stable-diffusion-2. [Accessed 1st June 2023].

- Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, Advances in Neural Information Processing Systems, volume 33, pages 6840–6851. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf.
- Yang Song. Generative modeling by estimating gradients of the data distribution, May 2021. URL https://yang-song.net/blog/2021/score/.
- JalFaizy Shaikh. Deep learning in the trenches: Understanding inception network from scratch, May 2020. URL https://www.analyticsvidhya.com/blog/2018/10/understanding-inception-network-from-scratch/. [Accessed 16th June 2023].