



HYPER-TRANSFORMING LATENT DIFFUSION MODELS

Ignacio Peis

Technical University of Denmark

ipeaz@dtu.dk



Further details

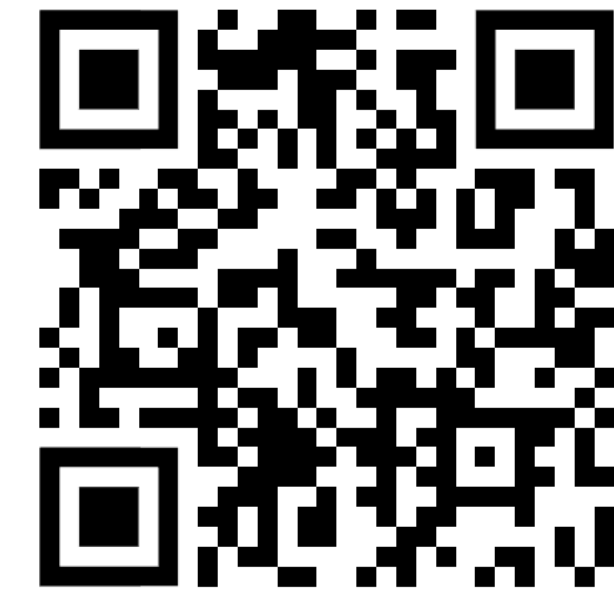
HYPER-TRANSFORMING LATENT DIFFUSION MODELS

Ignacio Peis*
Technical University of Denmark

Batuhan Koyuncu
Saarland University

Isabel Valera
Saarland University

Jes Frellsen
Technical University of Denmark



[\[Paper\]](#)

^[1] Peis et al., 2025

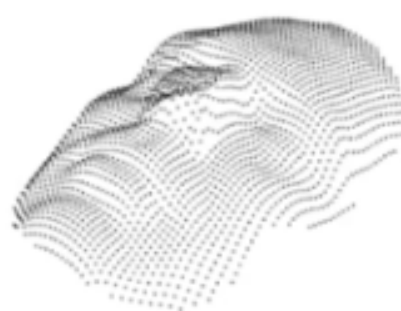
Introduction

- We typically discretized data that are continuous in nature.
- Real data can be expressed as a function over continuous coordinate systems.

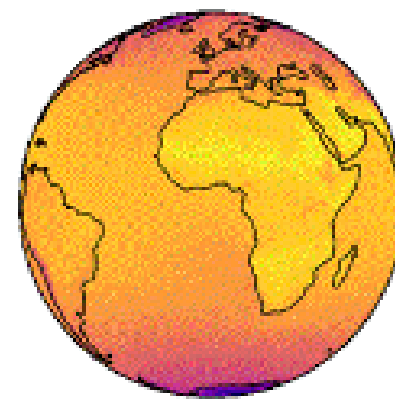
2D Images



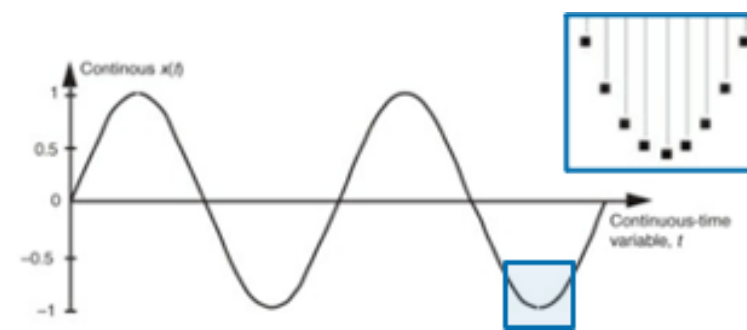
3D Images



Polar data



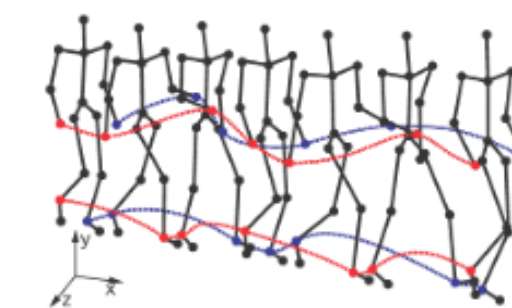
Time series



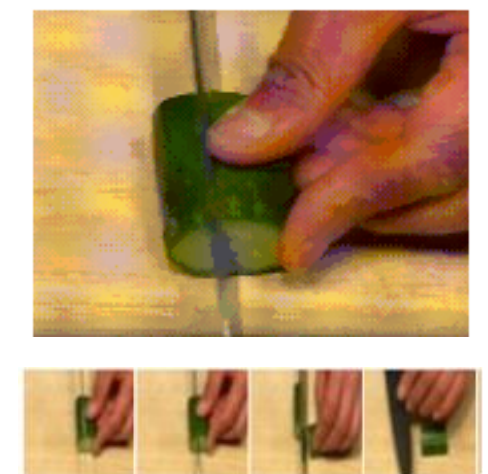
Audio



Motion sequences



Video



Spatial

Temporal

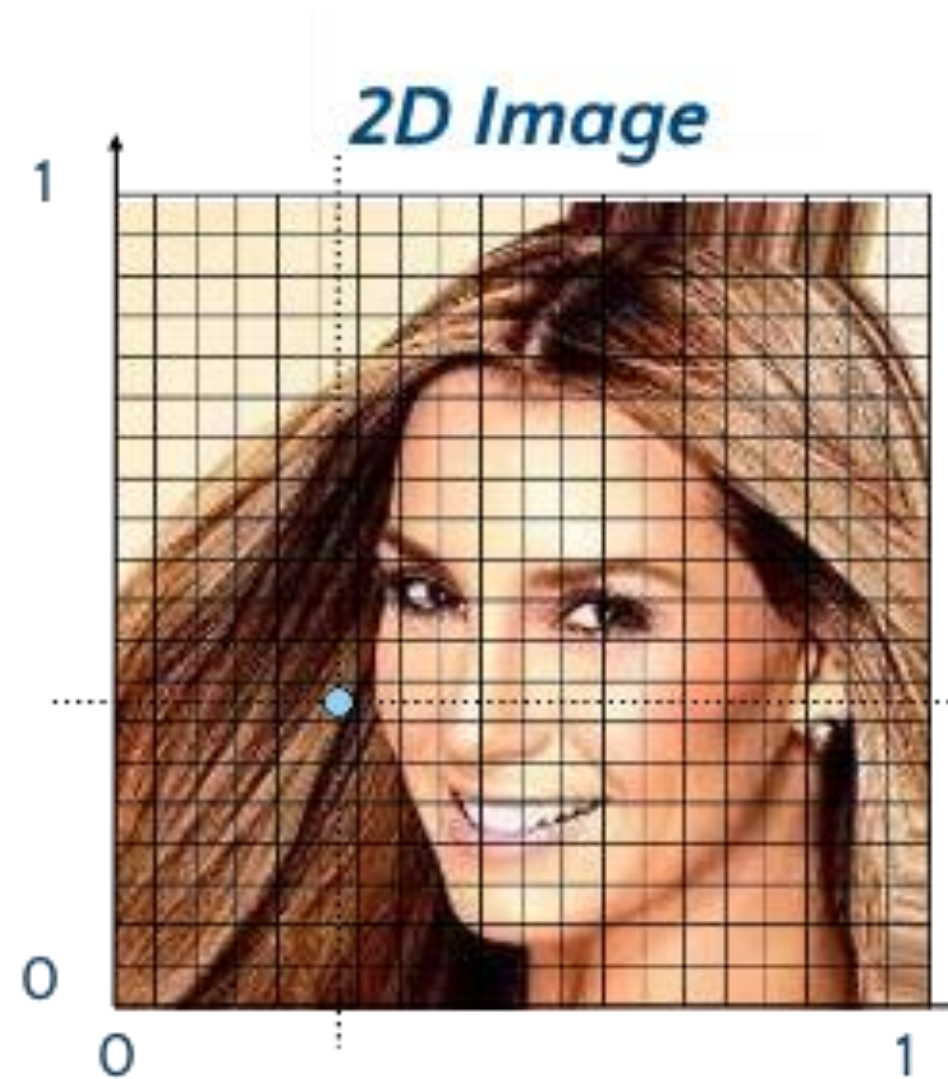
Spatio-temporal

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}^3, f(x_1, x_2) = (r, g, b) \quad f : \mathbb{R}^3 \rightarrow \{0, 1\}, f(x_1, x_2, x_3) = p \quad f : \mathbb{R}^2 \rightarrow \mathbb{R}, f(\varphi, \lambda) = T$$

$$f : \mathbb{R}^3 \rightarrow \mathbb{R}^3, f(x_1, x_2, t) = (r, g, b)$$

Introduction

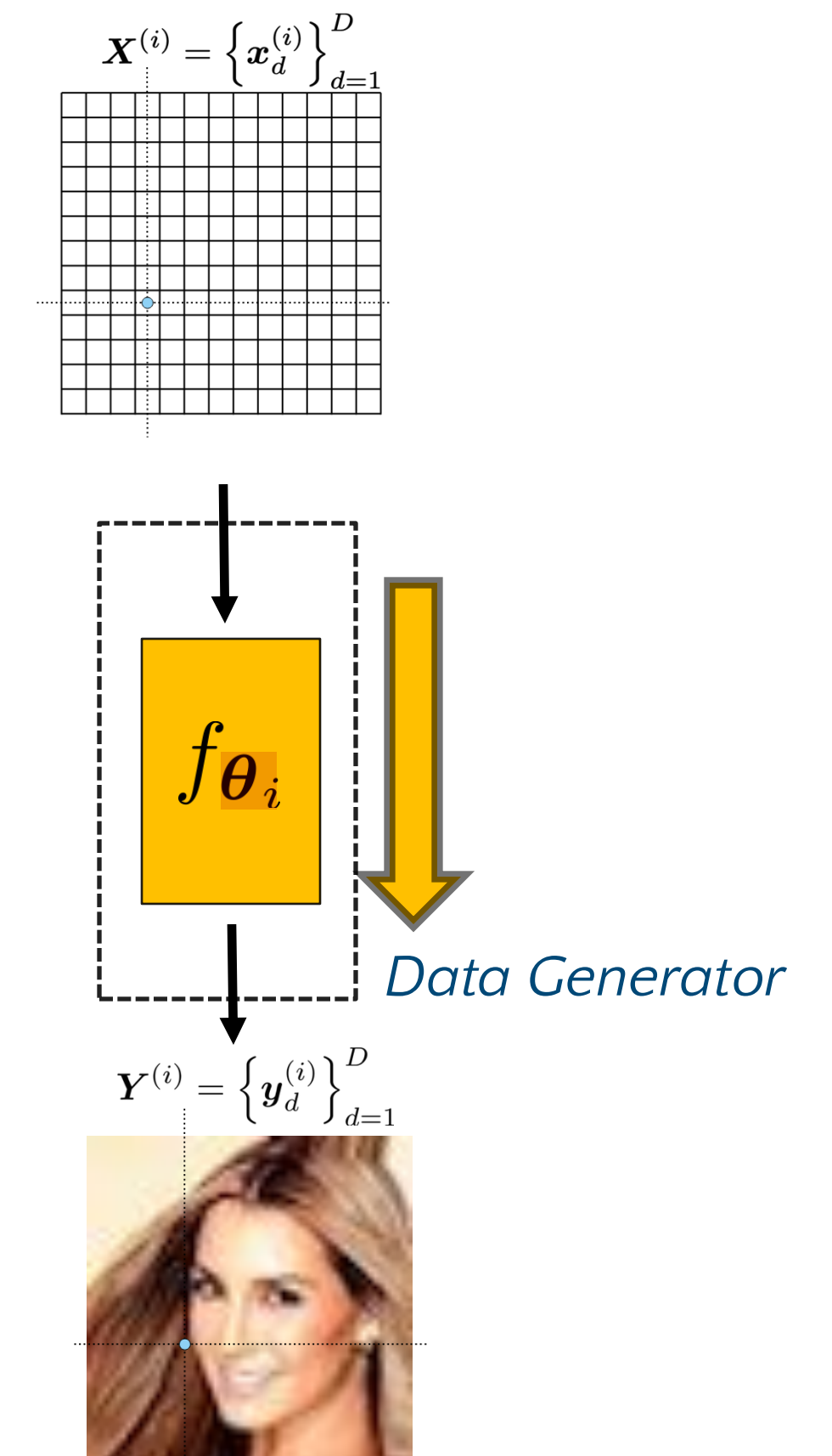
- Focusing on images:



- Generator function $f : \mathbf{X} \rightarrow \mathbf{Y}$ creates this specific image with the mapping $f(\mathbf{x}_d) = \mathbf{y}_d, d \in [1, \dots, D]$
- Each pixel is now a pair $\{\mathbf{x}_d, \mathbf{y}_d\}$ where $x_d \in \mathbb{R}^2, y_d \in \mathbb{R}^3$
- Full image is a pair of sets $\mathbf{X} = \{\mathbf{x}_d\}_{d=1}^D, \mathbf{Y}_d = \{\mathbf{y}_d\}_{d=1}^D$

Introduction

Implicit Neural Representations (INRs) [2-4]



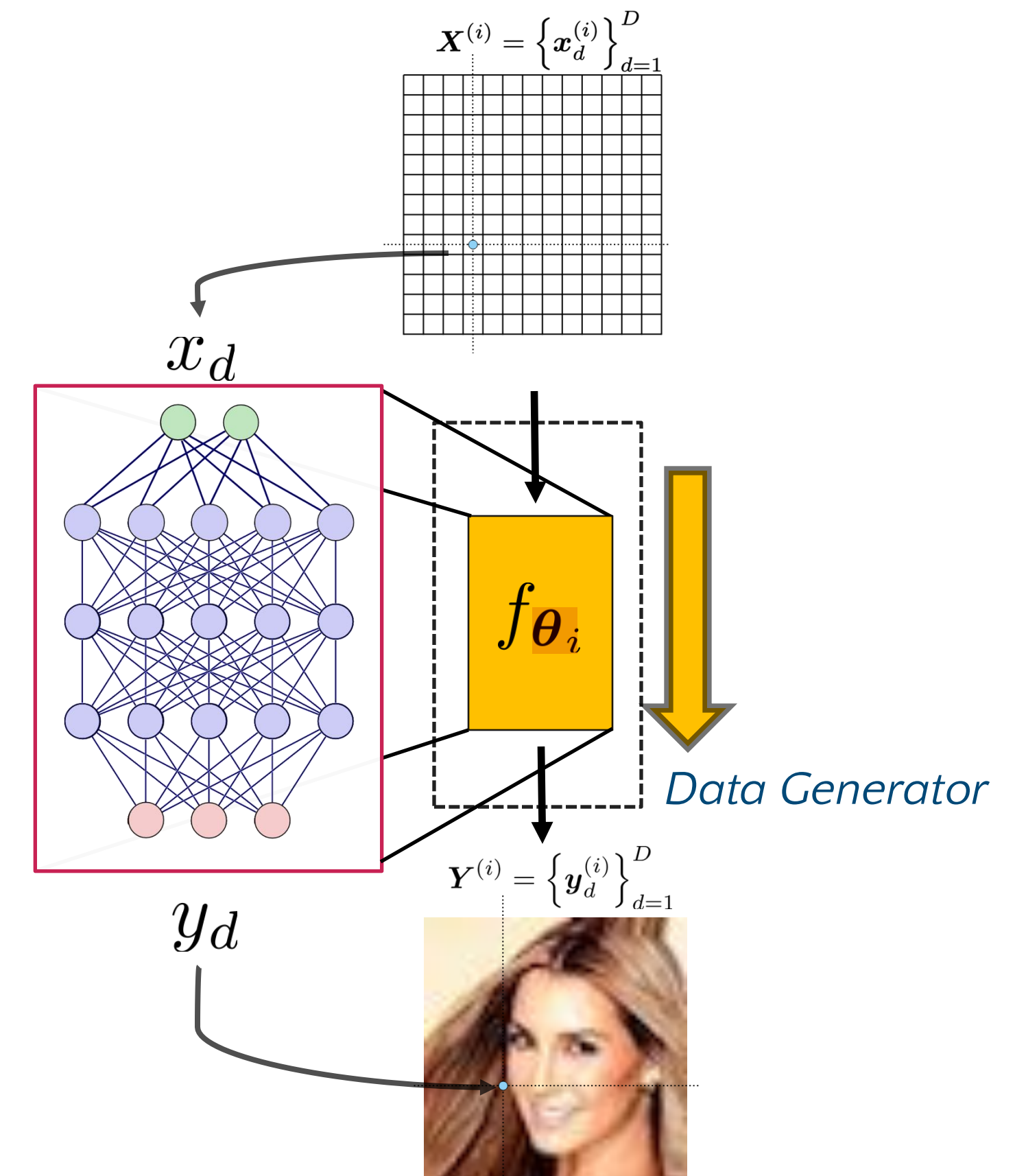
[2] Sitzmann et al., 2020

[3] Mescheder et al., 2019

[4] Sitzmann et al., 2019

Introduction

Implicit Neural Representations (INRs) [2-4]



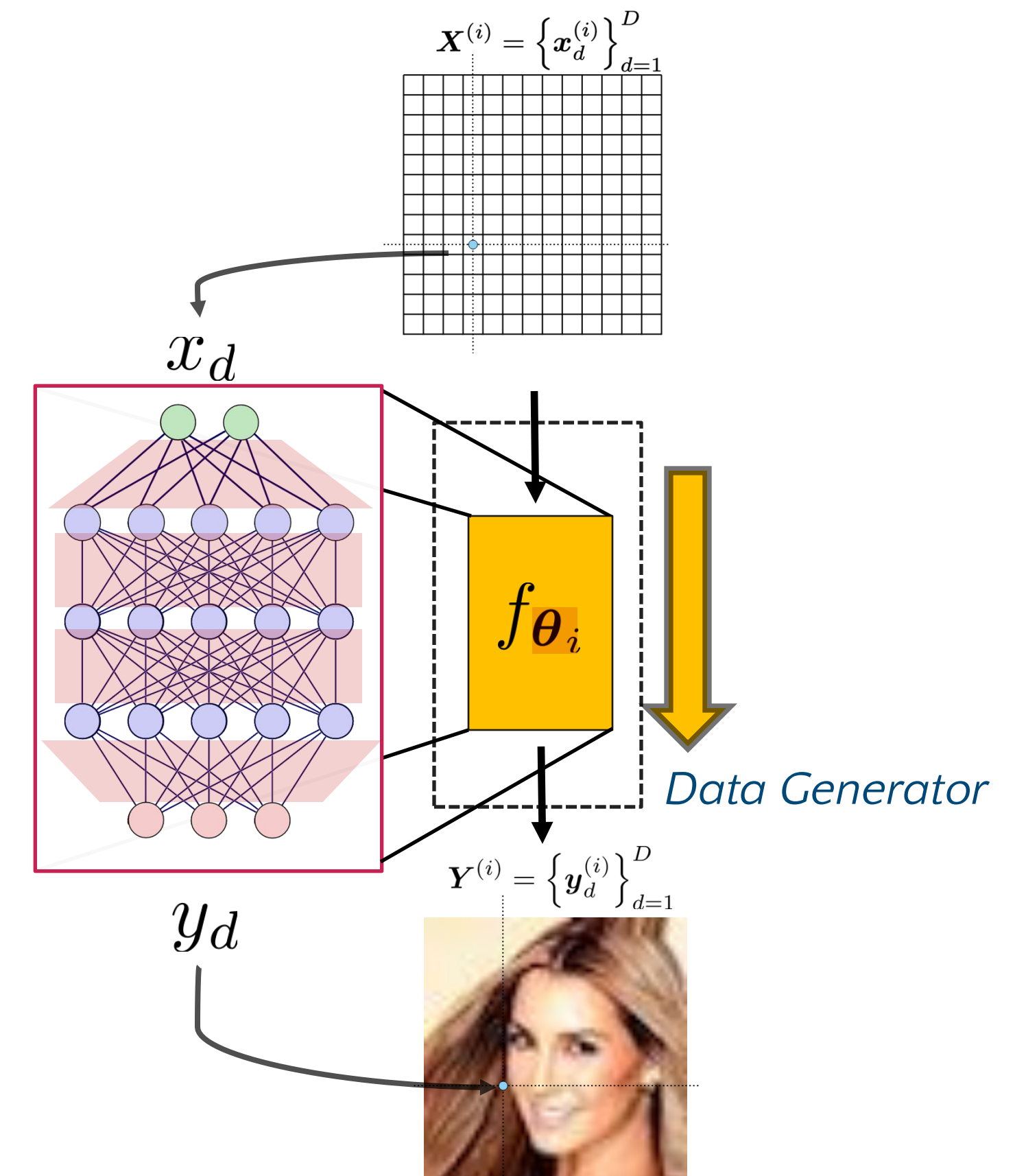
[2] Sitzmann et al., 2020

[3] Mescheder et al., 2019

[4] Sitzmann et al., 2019

Introduction

Implicit Neural Representations (INRs) [2-4]



[2] Sitzmann et al., 2020

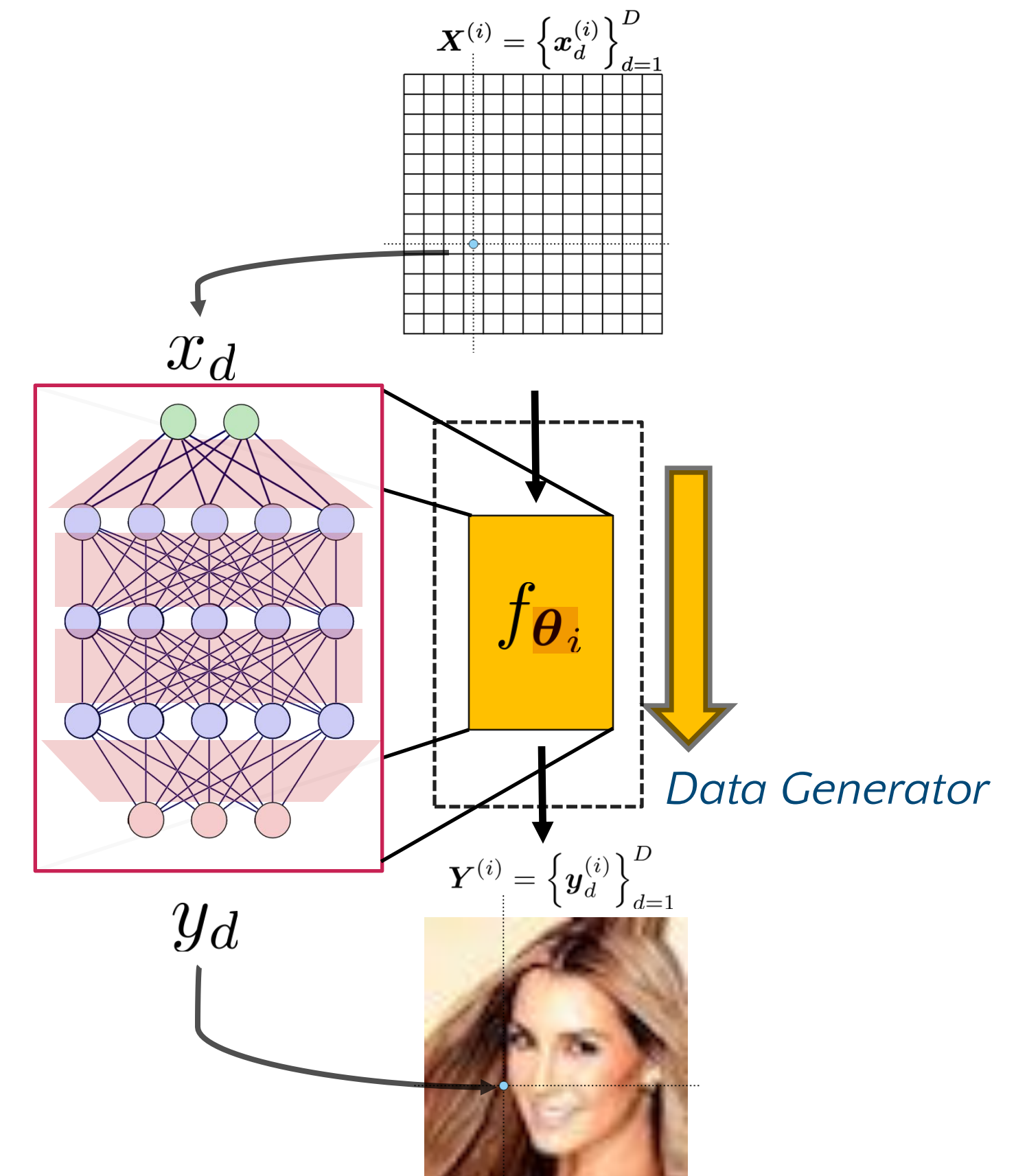
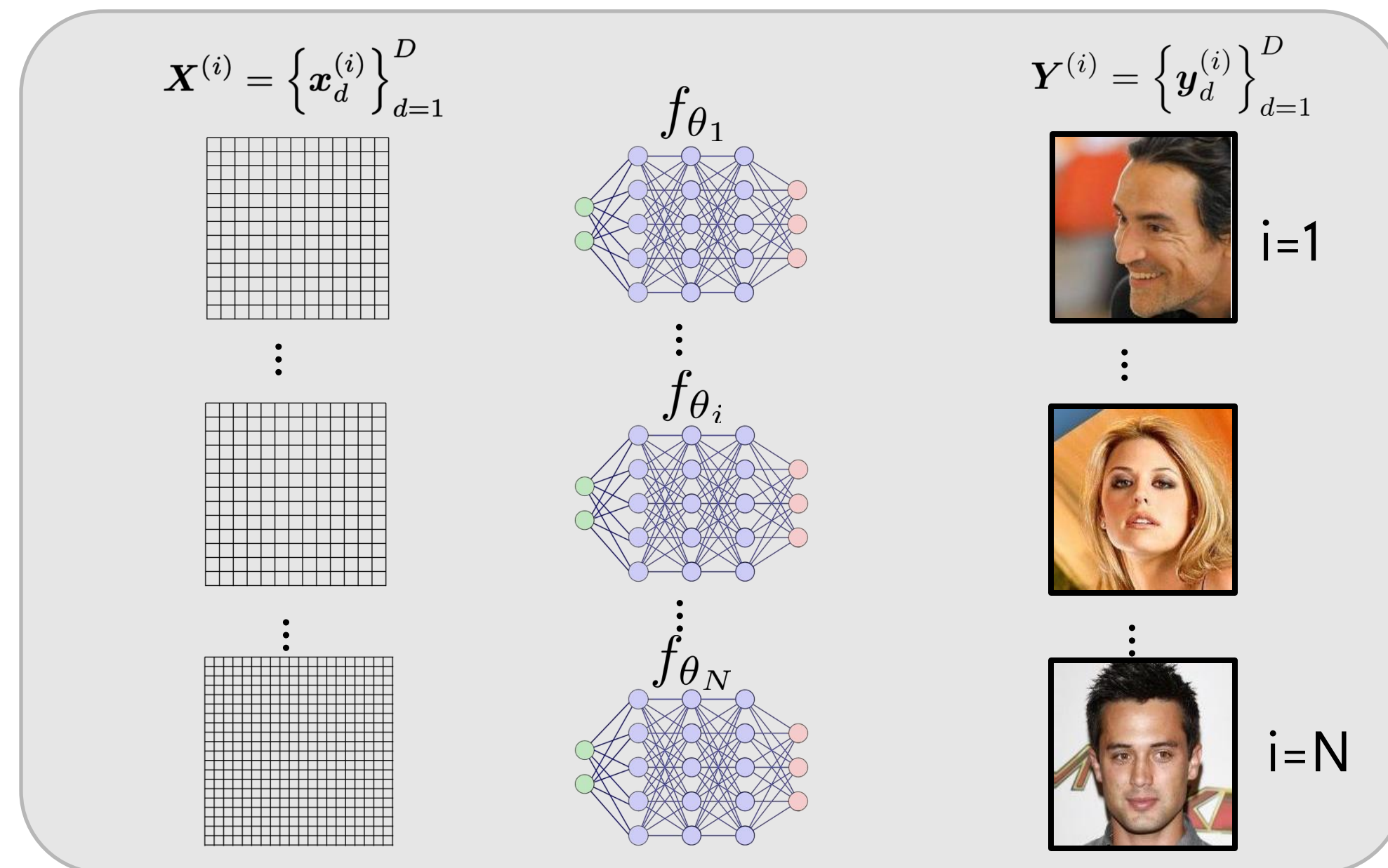
[3] Mescheder et al., 2019

[4] Sitzmann et al., 2019

Introduction

Implicit Neural Representations (INRs) [2-4]

Data generator f_{θ_i} is unique to each image



[2] Sitzmann et al., 2020

[3] Mescheder et al., 2019

[4] Sitzmann et al., 2019

Introduction

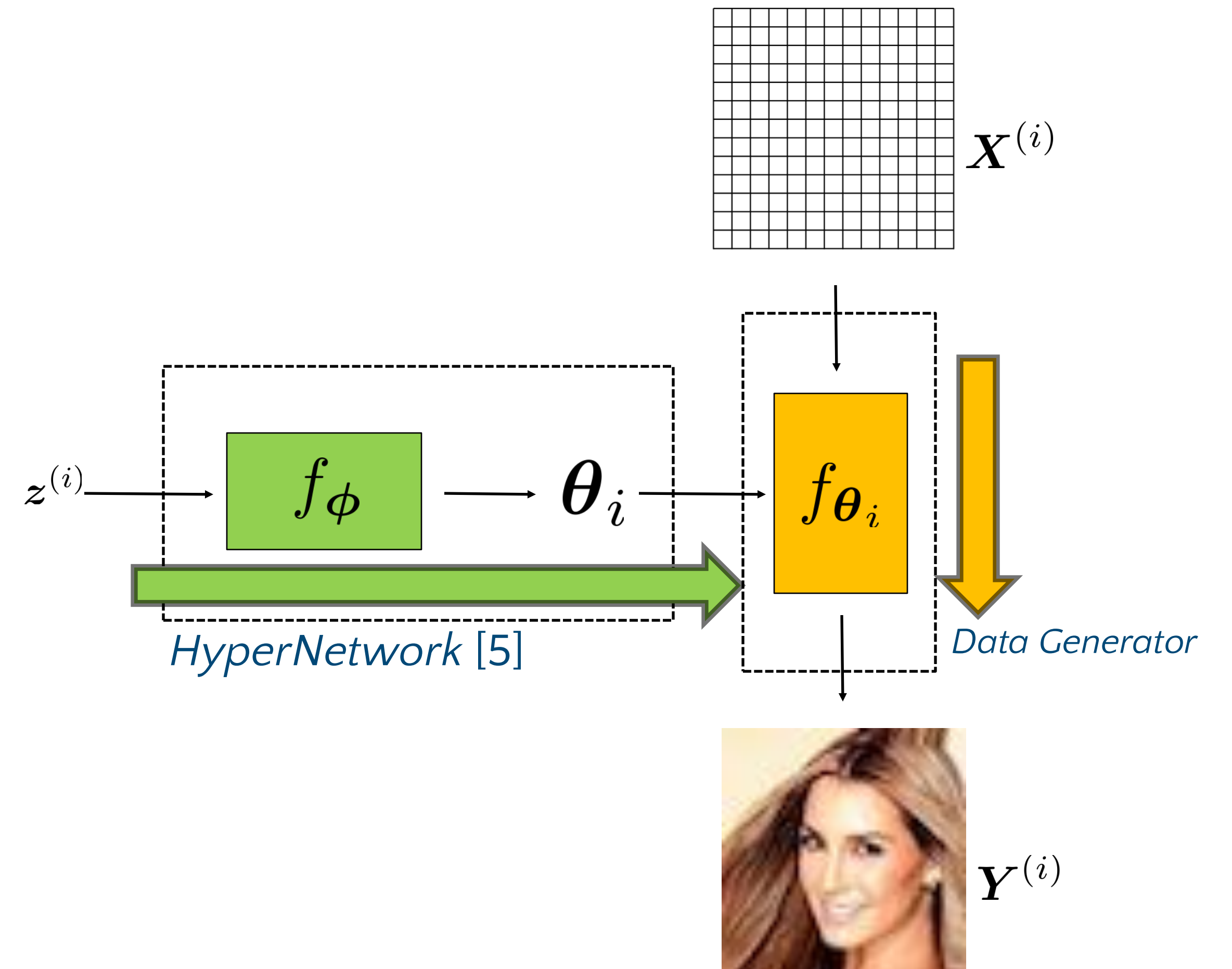
How to scale to large datasets?

How to map data to an INR?

Introduction

Hypernetworks [5]

Have $z^{(i)}$, a summary representation of image.



[5] Ha et al., 2017

Introduction

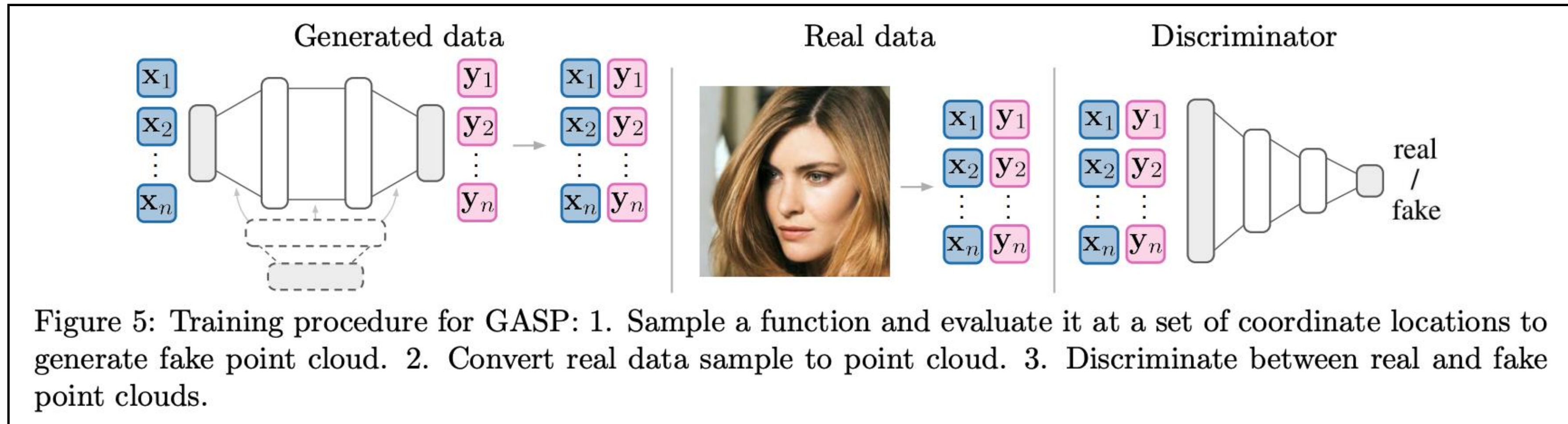
How to infer the latent representation $\mathbf{z}$?

$$p_{\theta}(\mathbf{z})$$

Related Work

GASP^[6]

- Adversarial training:



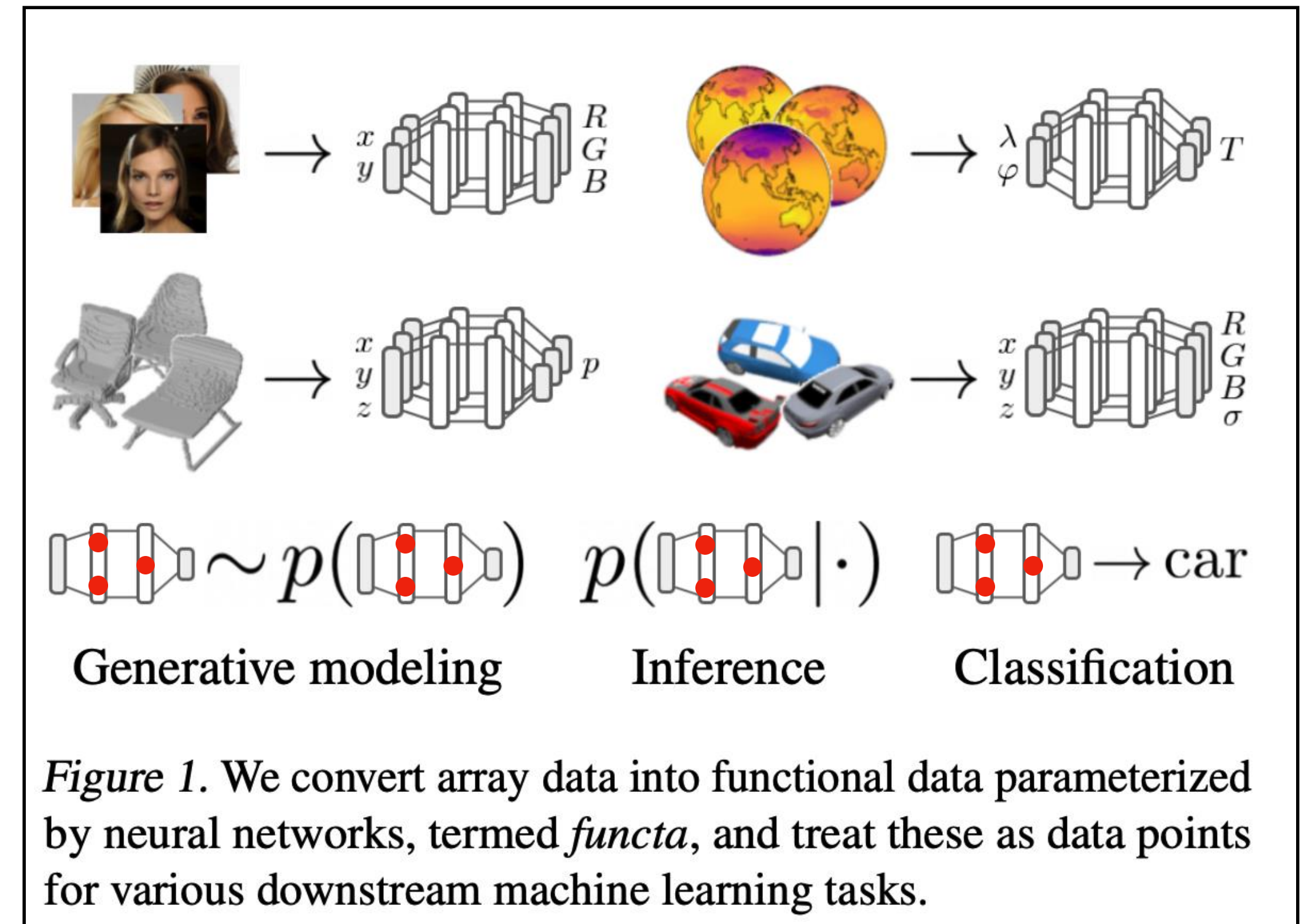
✗ Can't tackle inference related tasks.

^[6] Dupont et al., 2020

Related Work

Functa^[7]

- Decoupled training:
 1. Fit an INR per datapoint using SIREN^[2] and **modulation vectors**, named **functas**.
 2. Train any generative model on the functa dataset of vectors.
- ✗ Computationally expensive inference.



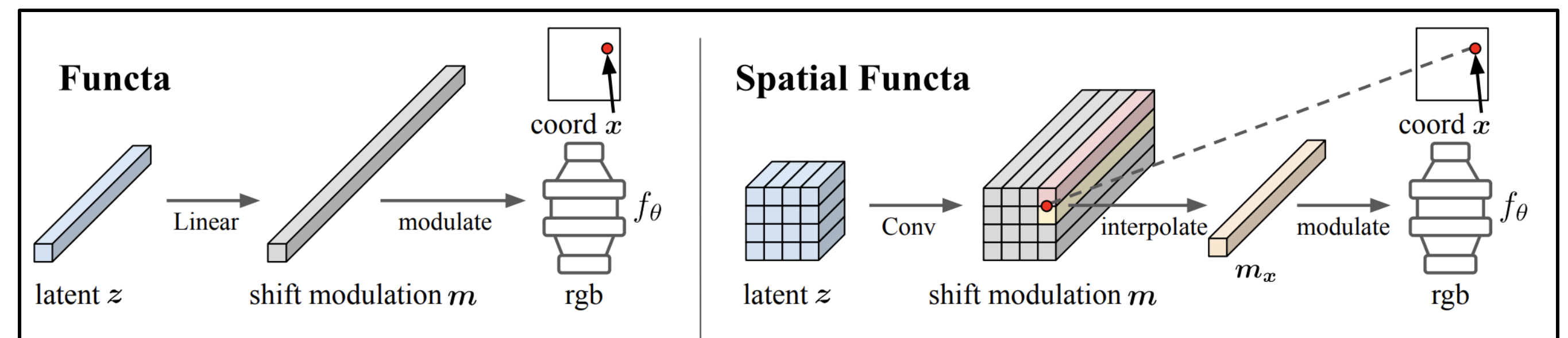
^[7] Dupont et al., 2022

^[2] Sitzmann et al., 2020

Related Work

Spatial Functa^[8]

- Decoupled training:
 1. Fit an INR per datapoint using SIREN^[8] and **modulation tensor**.
 2. Train any generative model on the functa dataset of tensors.
- ✗ Computationally expensive inference.



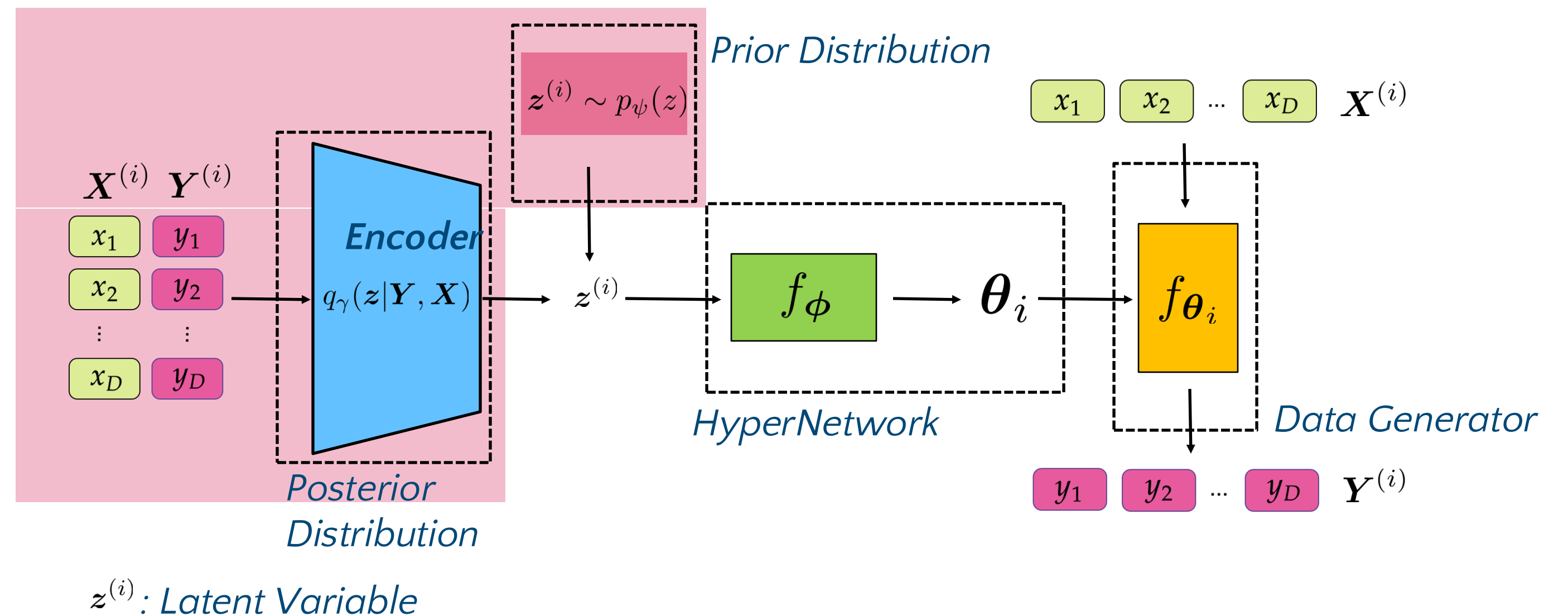
^[8] Bauer et al., 2023

^[2] Sitzmann et al., 2020

Related Work

VAMoH [9]

- Variational Inference
 - Requires flexible, learnable prior.



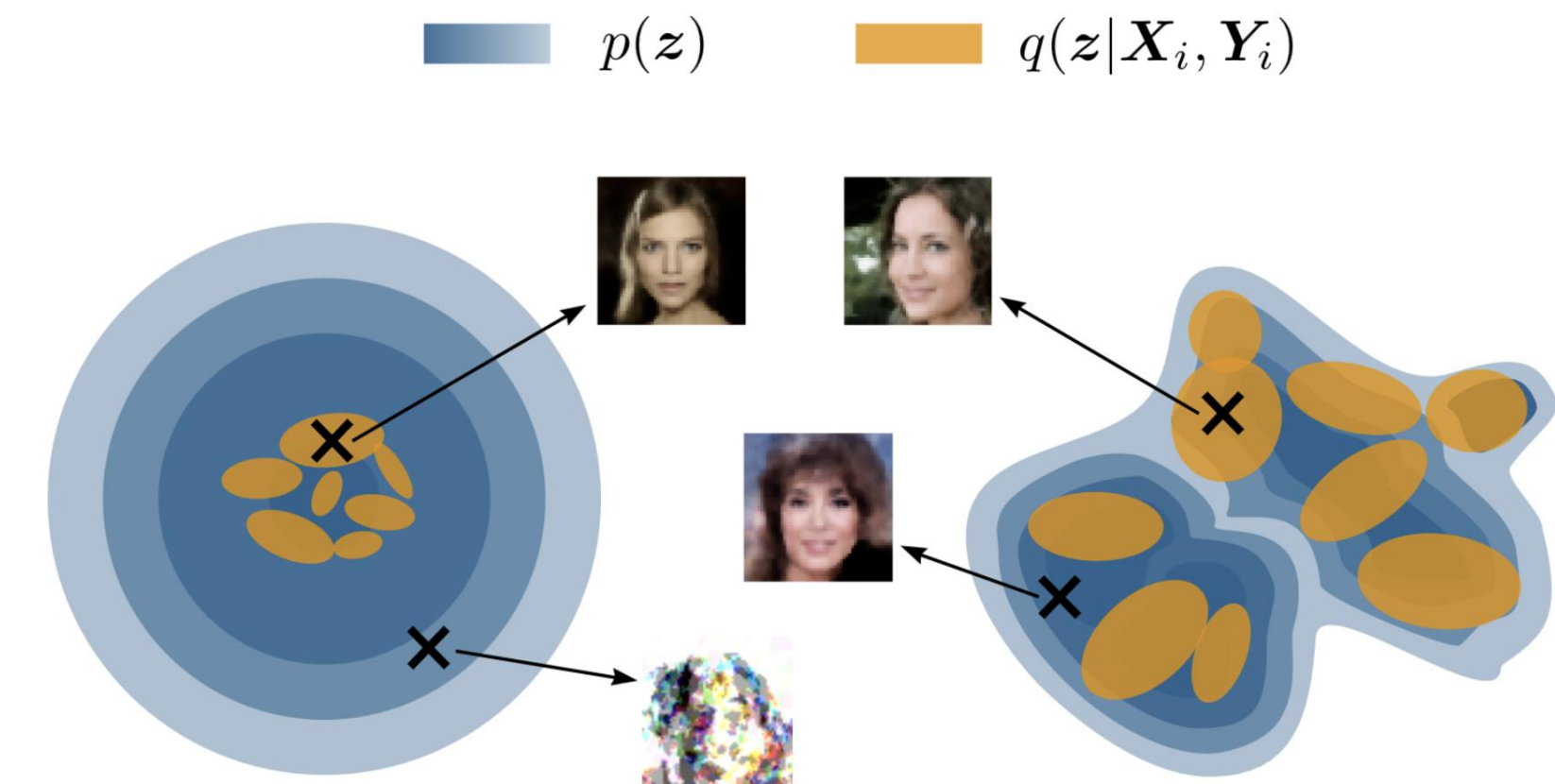
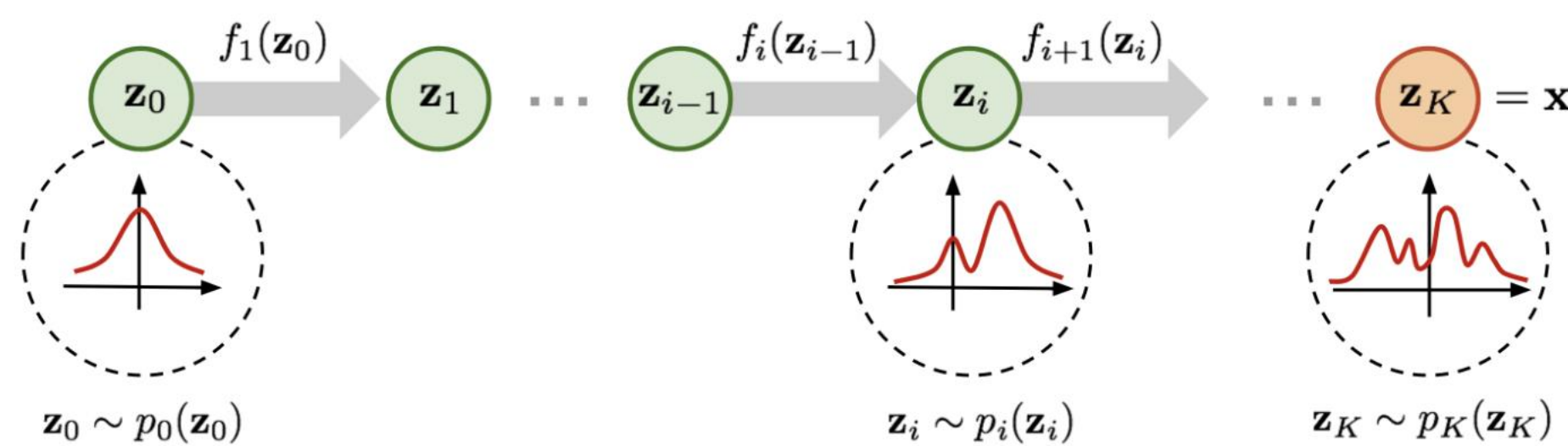
$$\mathcal{L}(\mathbf{Y}, \mathbf{X}; \psi, \phi, \gamma) = \sum_{d=1}^D \mathbb{E}_{q_{\gamma_z}(z|\mathbf{Y}, \mathbf{X})} \left[\sum_{k=1}^K \log p_{\theta_k}(\mathbf{y}_d | \mathbf{x}_d) \cdot \pi_{dk} \right] - D_{KL}(q_{\gamma_z}(z | \mathbf{X}, \mathbf{Y}) \| p_{\psi_z}(z))$$

[9] Koyuncu et al., 2023

Related Work

VAMoH [9]

- Variational Inference
 - Requires flexible, learnable prior.



The *hole* problem

$$\mathcal{L}(Y, X; \psi, \phi, \gamma) = \sum_{d=1}^D \mathbb{E}_{q_{\gamma_z}(z|Y, X)} \left[\sum_{k=1}^K \log p_{\theta_k}(y_d | x_d) \cdot \pi_{dk} \right] - D_{KL}(q_{\gamma_z}(z | X, Y) \| p_{\psi_z}(z))$$

[9] Koyuncu et al., 2023

Motivation

Flexibility of the latent space in [6, 7, 9]

✗ Poor generation quality.

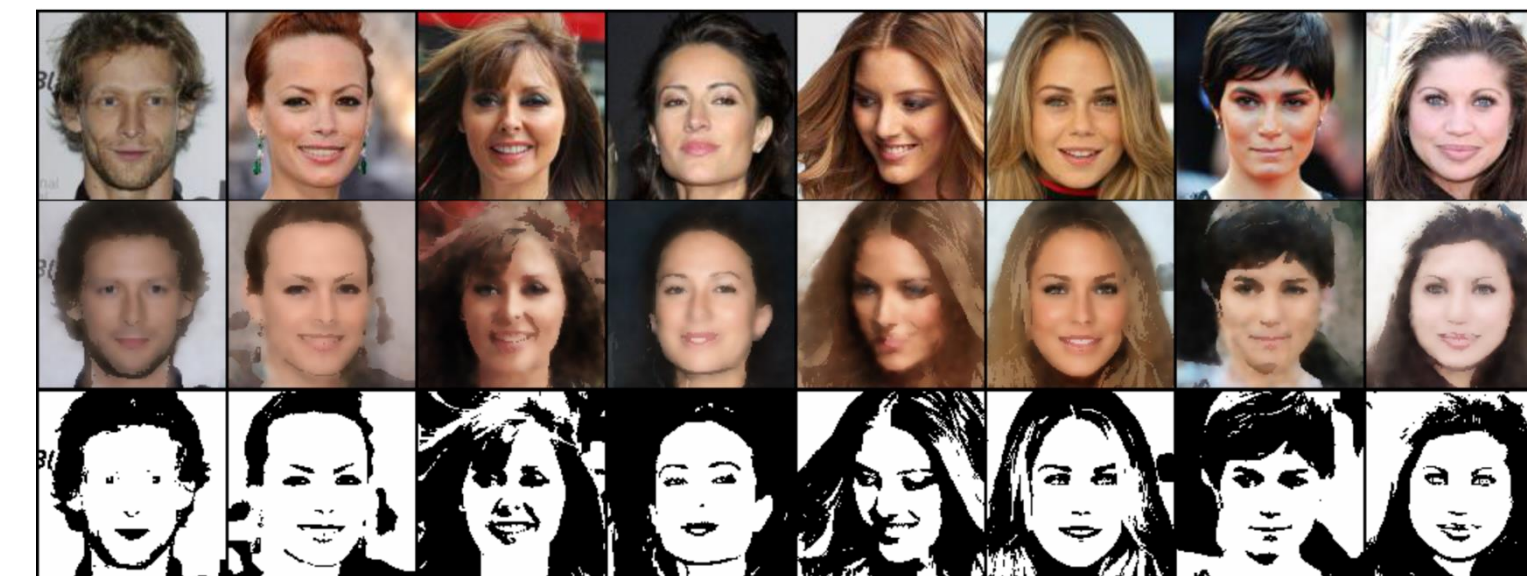
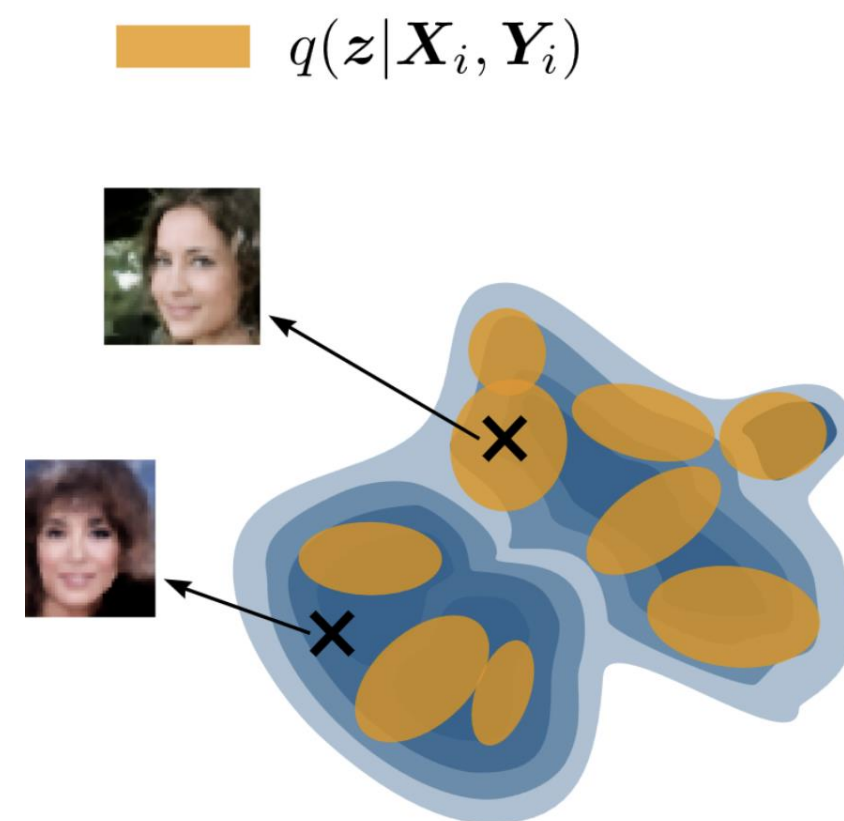
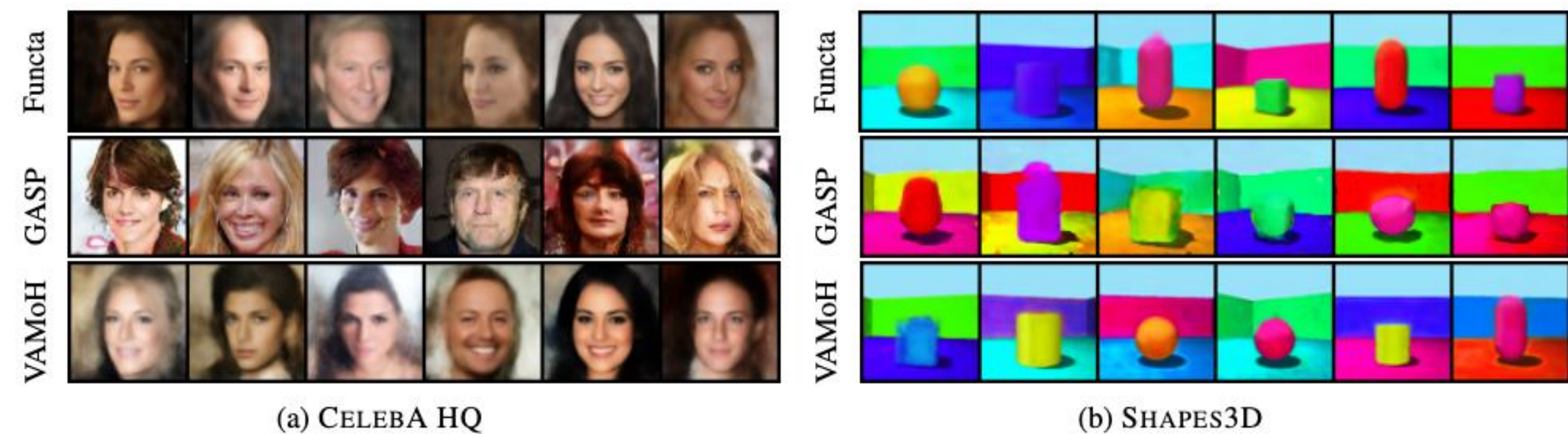


Image Reconstruction with VAMoH



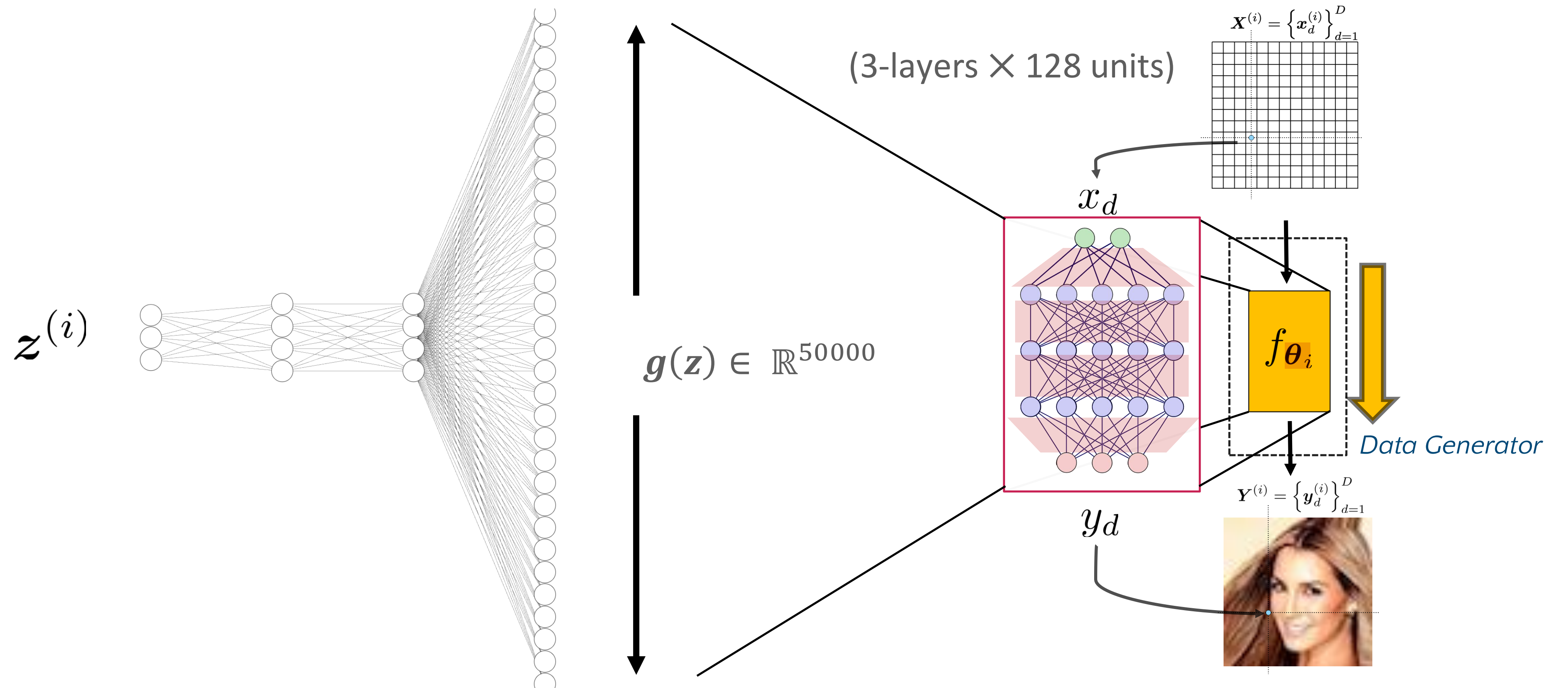
[6] Dupont et al., 2020

[7] Dupont et al., 2022

[9] Koyuncu et al., 2023

Motivation

Hypernetwork bottleneck in [6, 9]



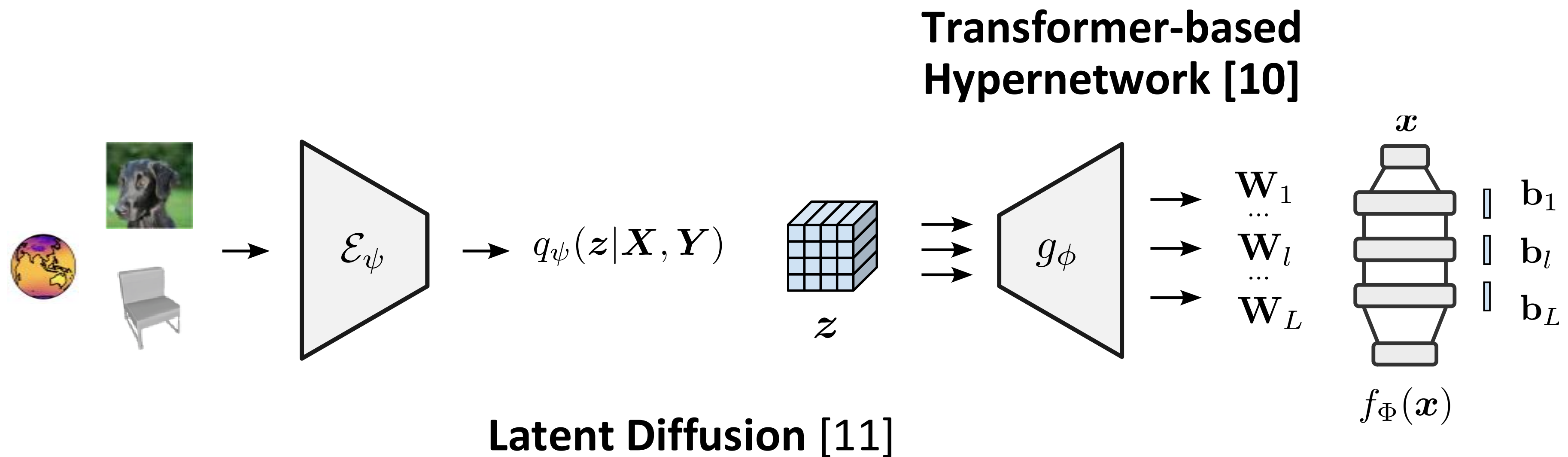
[6] Dupont et al., 2020

[9] Koyuncu et al., 2023

Proposed method

Latent Diffusion Models of INRs

LDMI [1]



[1] Peis et al., 2025

[10] Chen et al., 2024

[11] Rombach et al., 2021

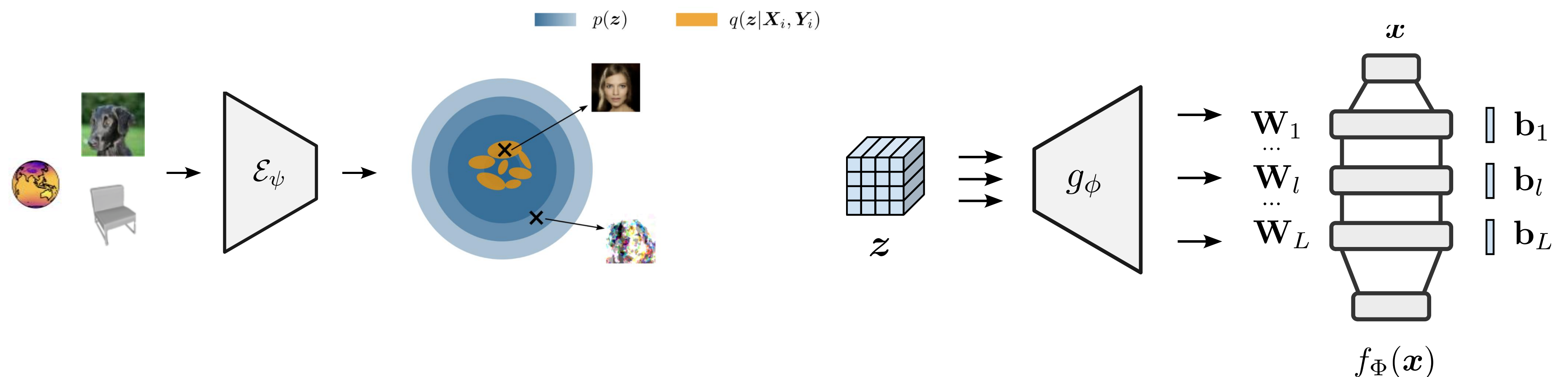
Latent Diffusion Models of INRs

The HD decoder

We will **firstly** train an “*under-regularized*” autoencoder to accurately represent data in a (tensor-shaped) latent space.

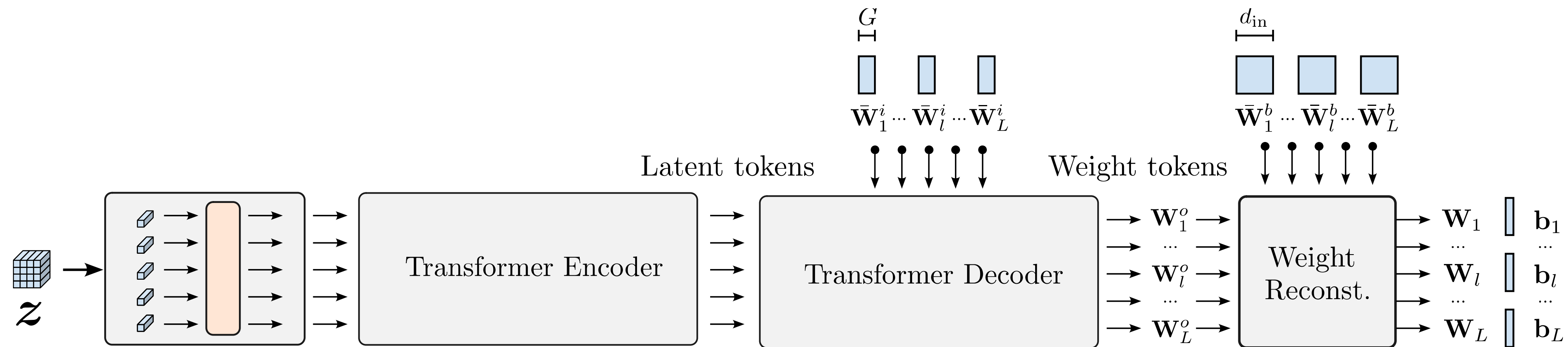
- The latents are mapped into INRs using our **transformer-based hypernetwork decoder**.

$$\mathcal{L}_{\text{VAE}}(\phi, \psi) = \mathbb{E}_{q_{\psi}(z|\mathbf{X})} [\log p_{\Phi}(\mathbf{X})] - \beta \cdot D_{\text{KL}}(q_{\psi}(z | \mathbf{X}) \| p(z)) + \mathcal{L}_{\text{perceptual}} + \mathcal{L}_{\text{GAN}}$$



Latent Diffusion Models of INRs

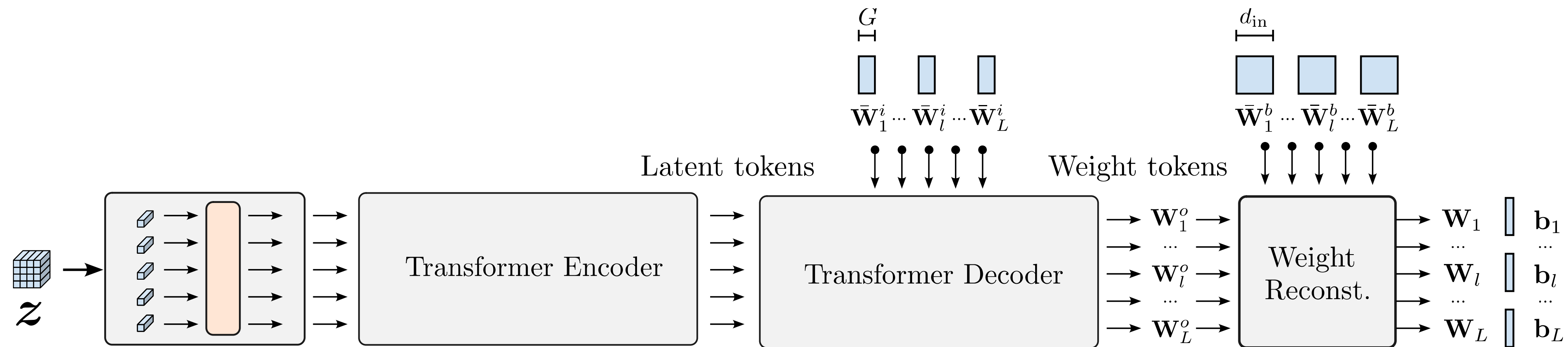
The HD decoder



- The **latent tokens** come from tokenizing $\mathbf{z}$ (following ViT [32]) and processing them via a Transformer Encoder.
- The **weight tokens** are columns of the weight matrices.

Latent Diffusion Models of INRs

The HD decoder



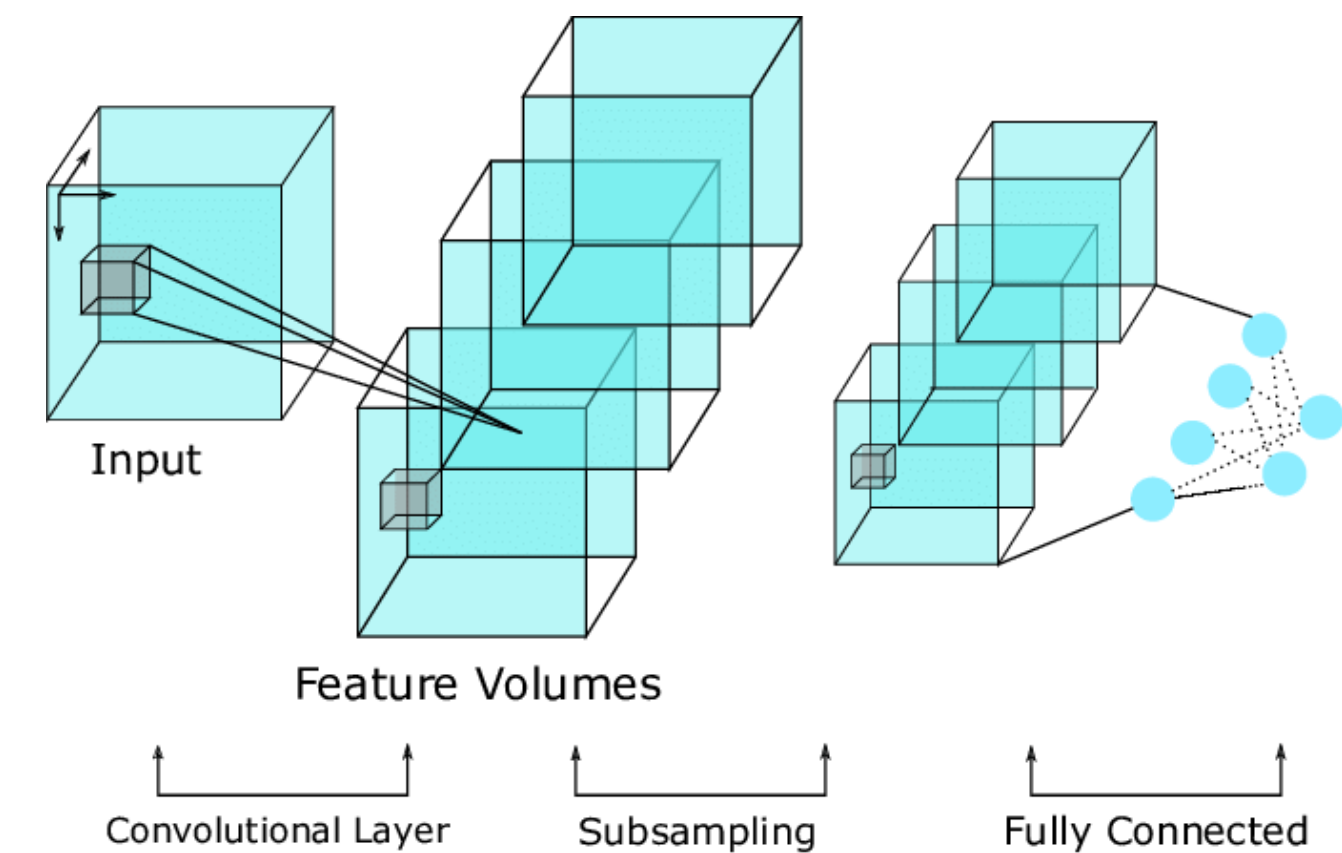
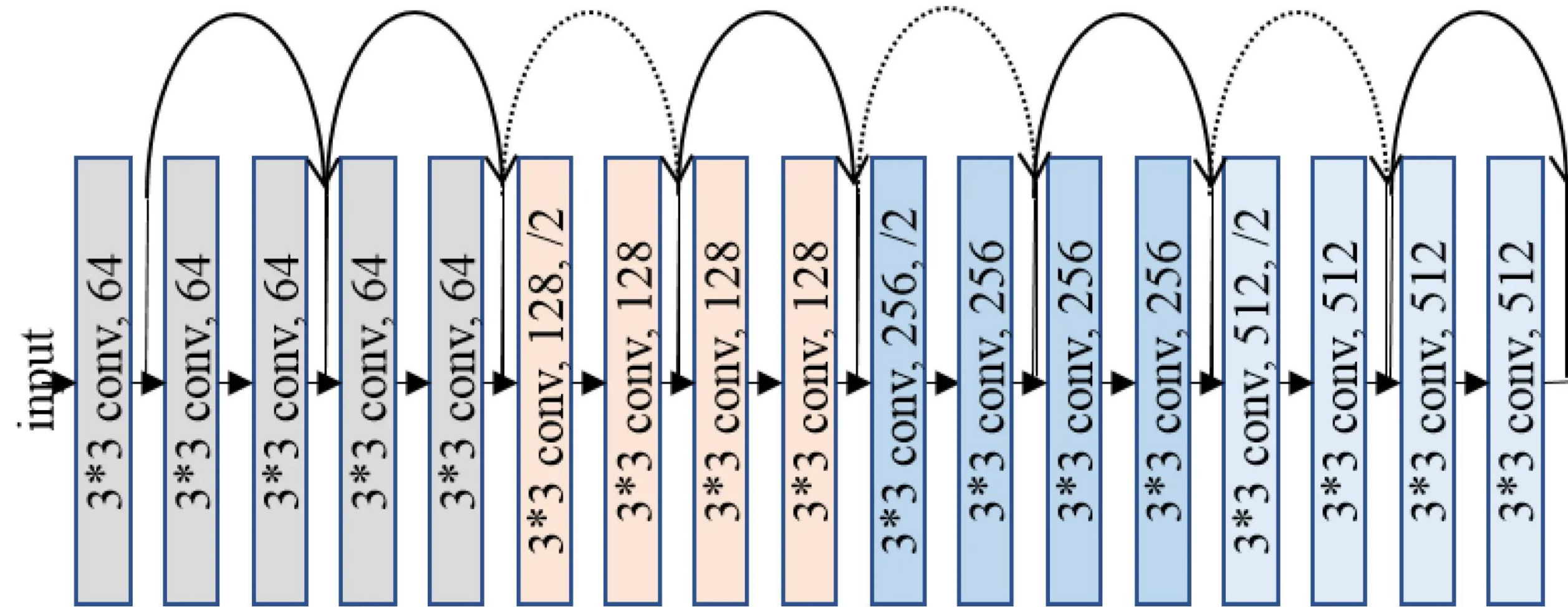
- A set of G initial weight tokens $\bar{\mathbf{w}}^i$ per layer (learnable, globally shared) cross-attend the *latent tokens*, $\bar{\mathbf{w}}^i$, to produce specific weights $\mathbf{w}^o$.
- The final weight tokens, $\mathbf{w}$, are obtained by expanding $\mathbf{w}^o$ using a learnable, globally shared template $\bar{\mathbf{w}}^b$:

$$w_c = \mathcal{R} \left(\mathbf{w}_{\lfloor c/k \rfloor}^o, \bar{\mathbf{w}}_c^b \right) = \left(1 + \mathbf{w}_{\lfloor c/k \rfloor}^o \right) \odot \bar{\mathbf{w}}_c^b$$

Latent Diffusion Models of INRs

Encoder

- The data is stored in a structured representation.
- We can make use of powerful encoders tailored to structured data (like ResNets).

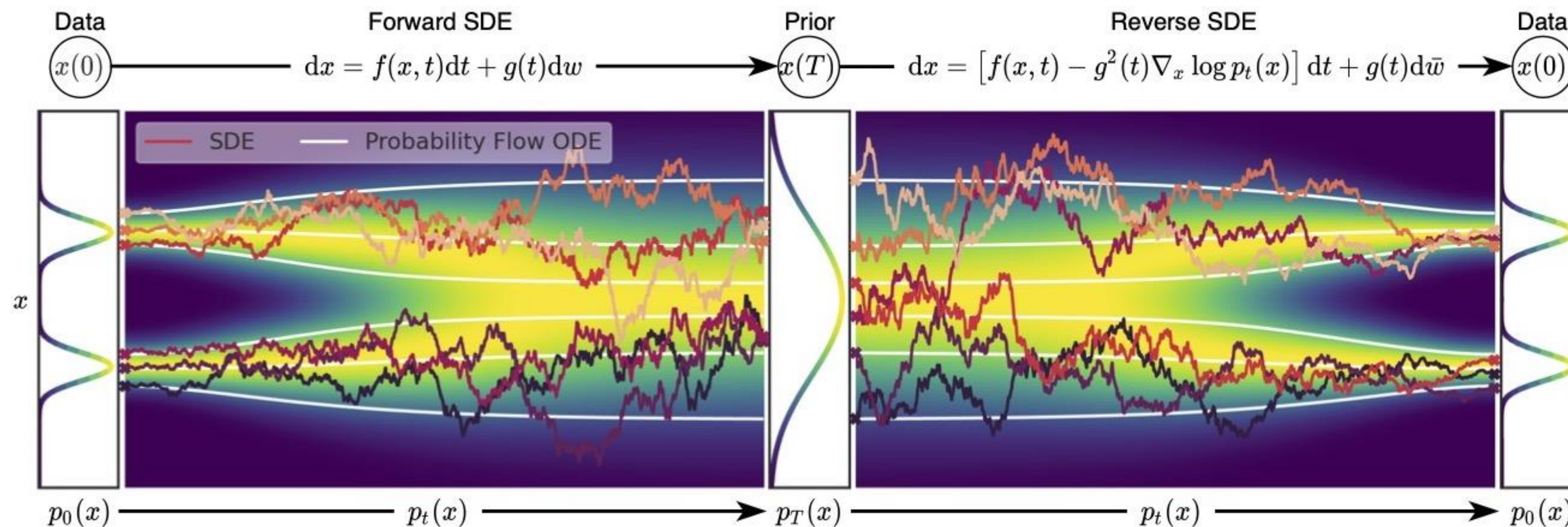


Latent Diffusion Models of INRs

Diffusion Models [12]

Denoising Score Matching

$$\theta^* = \arg \min_{\theta} \mathbb{E}_t \left\{ \lambda(t) \mathbb{E}_{\mathbf{x}(0)} \mathbb{E}_{\mathbf{x}(t)|\mathbf{x}(0)} \left[\left\| \mathbf{s}_{\theta}(\mathbf{x}(t), t) - \nabla_{\mathbf{x}(t)} \log p_{0t}(\mathbf{x}(t) | \mathbf{x}(0)) \right\|_2^2 \right] \right\}$$

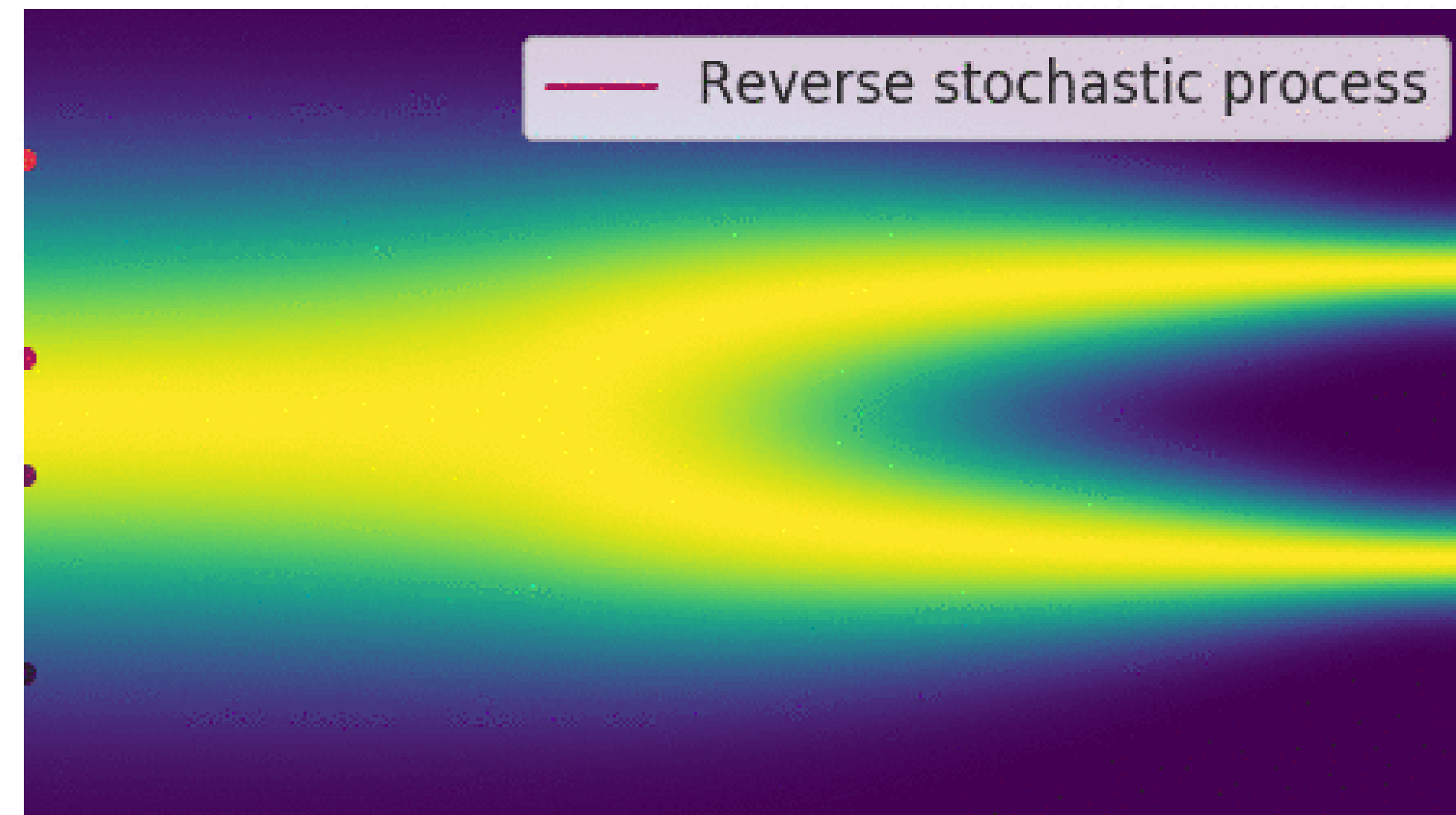


[12] Song et al., 2020

Latent Diffusion Models of INRs

Diffusion Models [12]

$$s_{\theta}(x_t, t)$$



[12] Song et al., 2020

Latent Diffusion Models of INRs

DDPM [13]

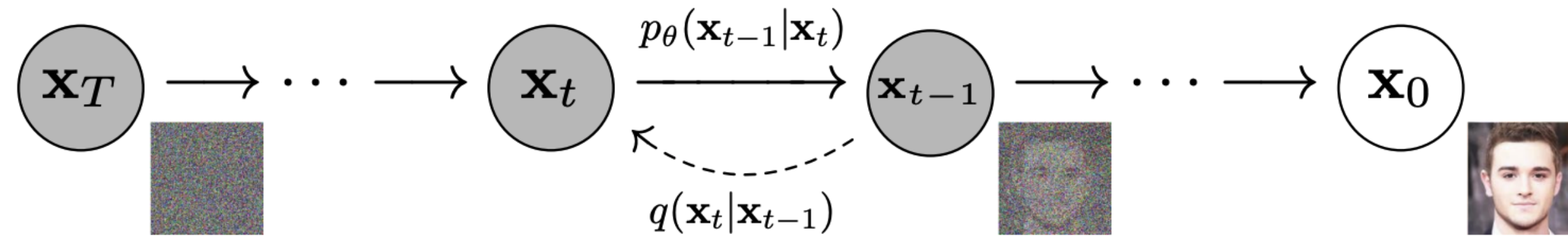


Figure 2: The directed graphical model considered in this work.

$$p_{\theta}(\mathbf{x}_{0:T}) := p(\mathbf{x}_T) \prod_{t=1}^T p_{\theta}(\mathbf{x}_{t-1} | \mathbf{x}_t), \quad p_{\theta}(\mathbf{x}_{t-1} | \mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_{\theta}(\mathbf{x}_t, t), \boldsymbol{\Sigma}_{\theta}(\mathbf{x}_t, t))$$

$$q(\mathbf{x}_{1:T} | \mathbf{x}_0) := \prod_{t=1}^T q(\mathbf{x}_t | \mathbf{x}_{t-1}), \quad q(\mathbf{x}_t | \mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t} \mathbf{x}_{t-1}, \beta_t \mathbf{I})$$

$$q(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t} \mathbf{x}_0, (1 - \bar{\alpha}_t) \mathbf{I}) \quad \alpha_t := 1 - \beta_t \quad \bar{\alpha}_t := \prod_{s=1}^t \alpha_s$$

$$\mathbb{E}_q \left[\underbrace{D_{\text{KL}}(q(\mathbf{x}_T | \mathbf{x}_0) || p(\mathbf{x}_T))}_{L_T} + \sum_{t>1} \underbrace{D_{\text{KL}}(q(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0) || p_{\theta}(\mathbf{x}_{t-1} | \mathbf{x}_t))}_{L_{t-1}} - \underbrace{\log p_{\theta}(\mathbf{x}_0 | \mathbf{x}_1)}_{L_0} \right]$$

[13] Ho et al., 2020

Latent Diffusion Models of INRs

DDPM [13]

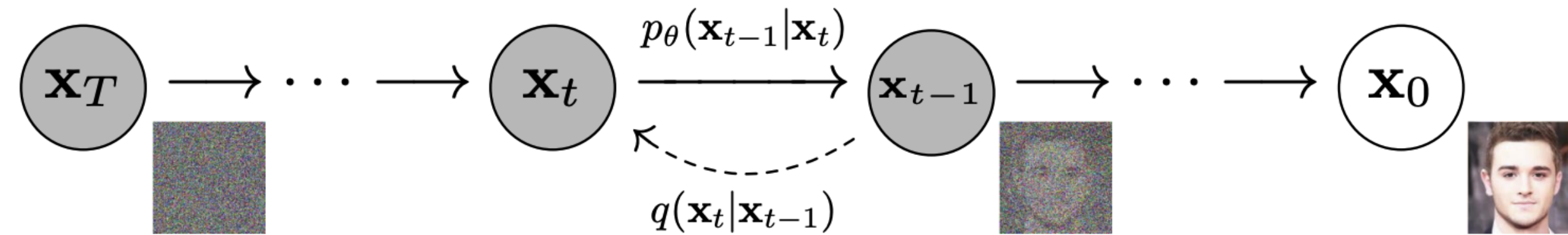


Figure 2: The directed graphical model considered in this work.

$$\mathbb{E}_q \left[\underbrace{D_{\text{KL}}(q(\mathbf{x}_T | \mathbf{x}_0) \| p(\mathbf{x}_T))}_{L_T} + \sum_{t>1} \underbrace{D_{\text{KL}}(q(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0) \| p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t))}_{L_{t-1}} - \underbrace{\log p_\theta(\mathbf{x}_0 | \mathbf{x}_1)}_{L_0} \right]$$



$$\mathbb{E}_{\mathbf{x}_0, \epsilon} \left[\frac{\beta_t^2}{2\sigma_t^2 \alpha_t (1 - \bar{\alpha}_t)} \left\| \epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, t) \right\|^2 \right]$$

[13] Ho et al., 2020

Latent Diffusion Models of INRs

DDPM [13]

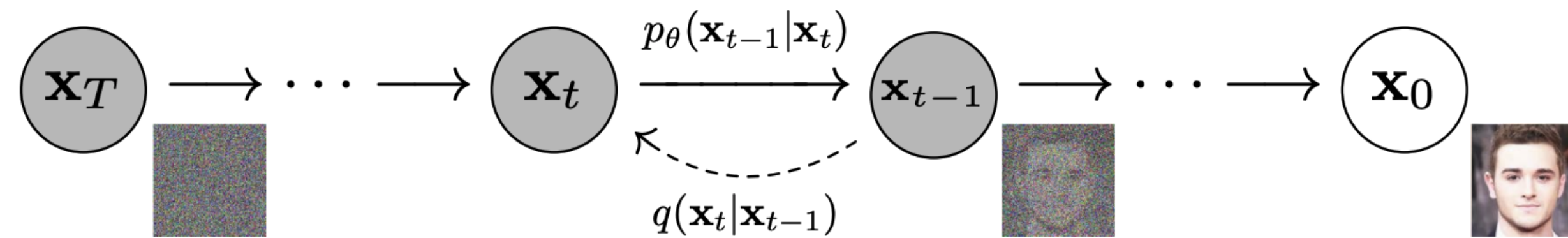


Figure 2: The directed graphical model considered in this work.

$$\mathbb{E}_{\mathbf{x}_0, \epsilon} \left[\frac{\beta_t^2}{2\sigma_t^2 \alpha_t (1 - \bar{\alpha}_t)} \left\| \epsilon - \epsilon_\theta \left(\sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, t \right) \right\|^2 \right]$$



$$L_{\text{simple}}(\theta) := \mathbb{E}_{t, \mathbf{x}_0, \epsilon} \left[\left\| \epsilon - \epsilon_\theta \left(\sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, t \right) \right\|^2 \right]$$

[13] Ho et al., 2020

Latent Diffusion Models of INRs

DDIM [14]

- Define a Non-Markovian Inference Model:

$$p_{\theta}(z_{t-1} | z_t) = \begin{cases} \mathcal{N}(\hat{z}, \sigma_1^2 \mathbf{I}) & \text{if } t = 1 \\ q(z_{t-1} | z_t, \hat{z}) & \text{otherwise} \end{cases}$$

- The objective is the same!

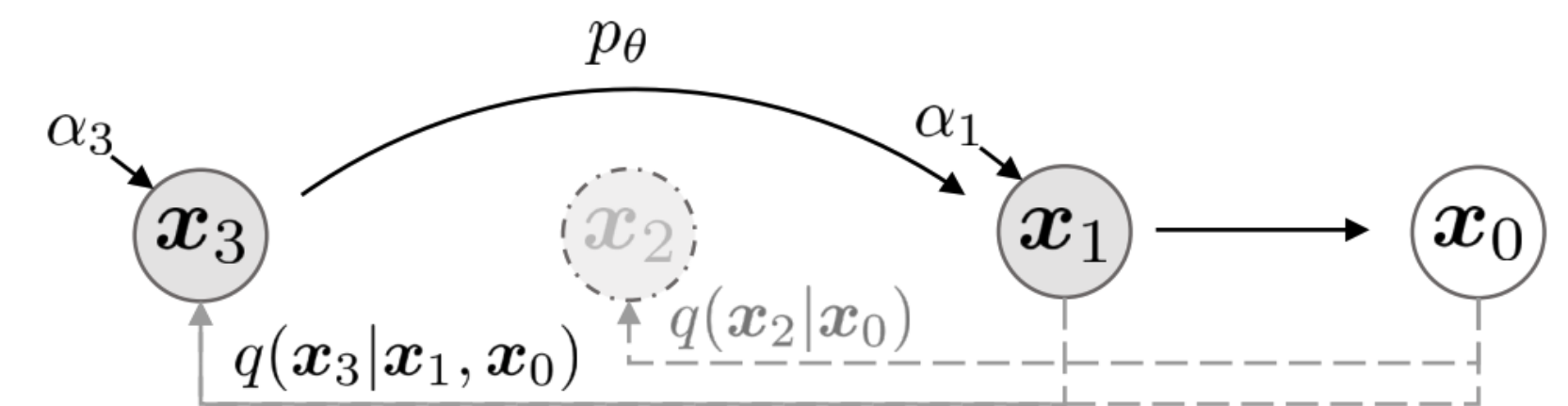
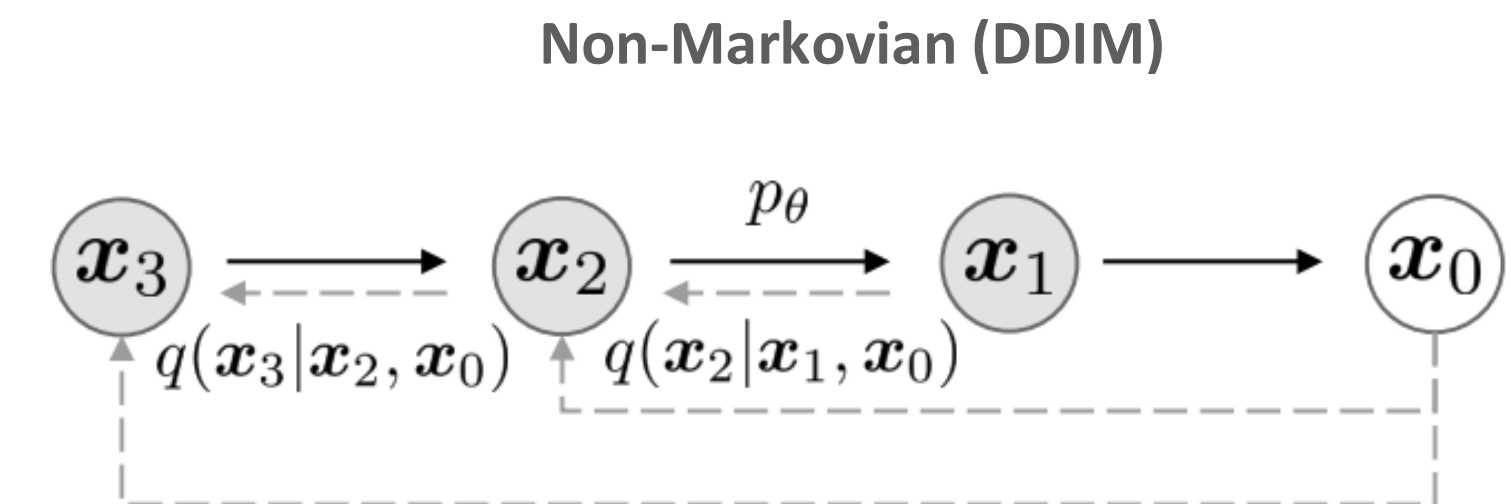
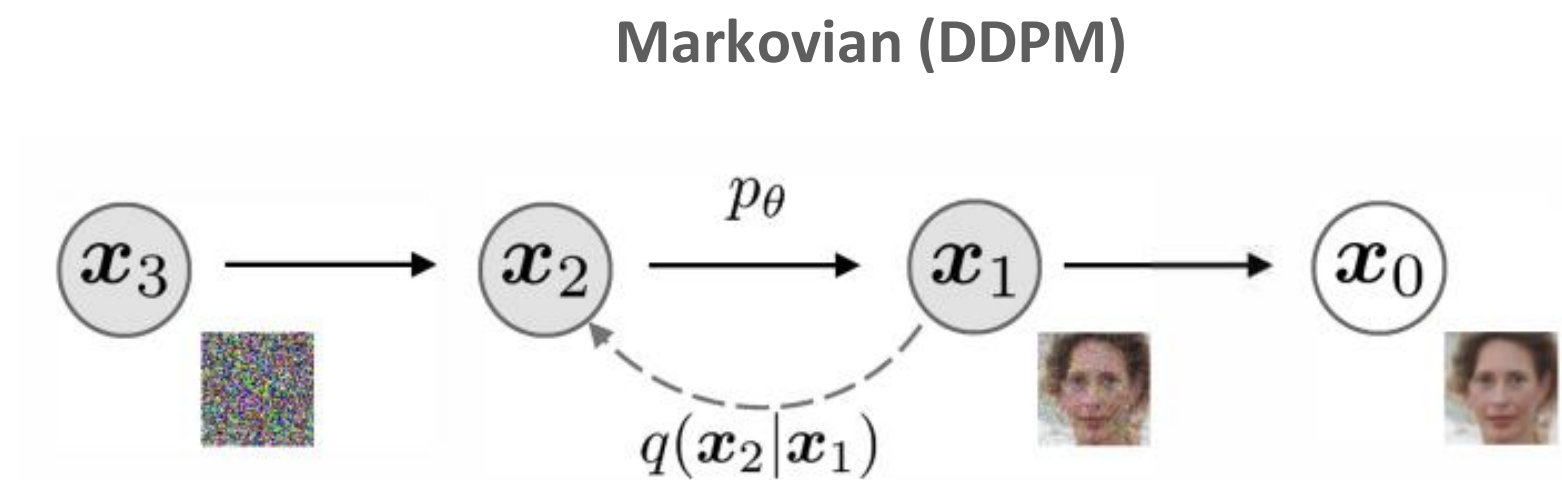
$$L_{\text{simple}}(\theta) := \mathbb{E}_{t, \mathbf{x}_0, \epsilon} \left[\left\| \epsilon - \epsilon_{\theta}(\sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, t) \right\|^2 \right]$$

- Using the same model, you can sample in fewer steps!

$$p_{\theta}^{(\tau_i)}(\mathbf{x}_{\tau_{i-1}} | \mathbf{x}_{\tau_i}) = q_{\sigma, \tau}(\mathbf{x}_{\tau_{i-1}} | \mathbf{x}_{\tau_i}, f_{\theta}^{(\tau_i)}(\mathbf{x}_{\tau_{i-1}})) \quad \text{if } i \in [S], i > 1$$

$$p_{\theta}^{(t)}(\mathbf{x}_0 | \mathbf{x}_t) = \mathcal{N}(f_{\theta}^{(t)}(\mathbf{x}_t), \sigma_t^2 \mathbf{I}) \quad \text{otherwise,}$$

[14] Song et al., 2021



Latent Diffusion Models of INRs

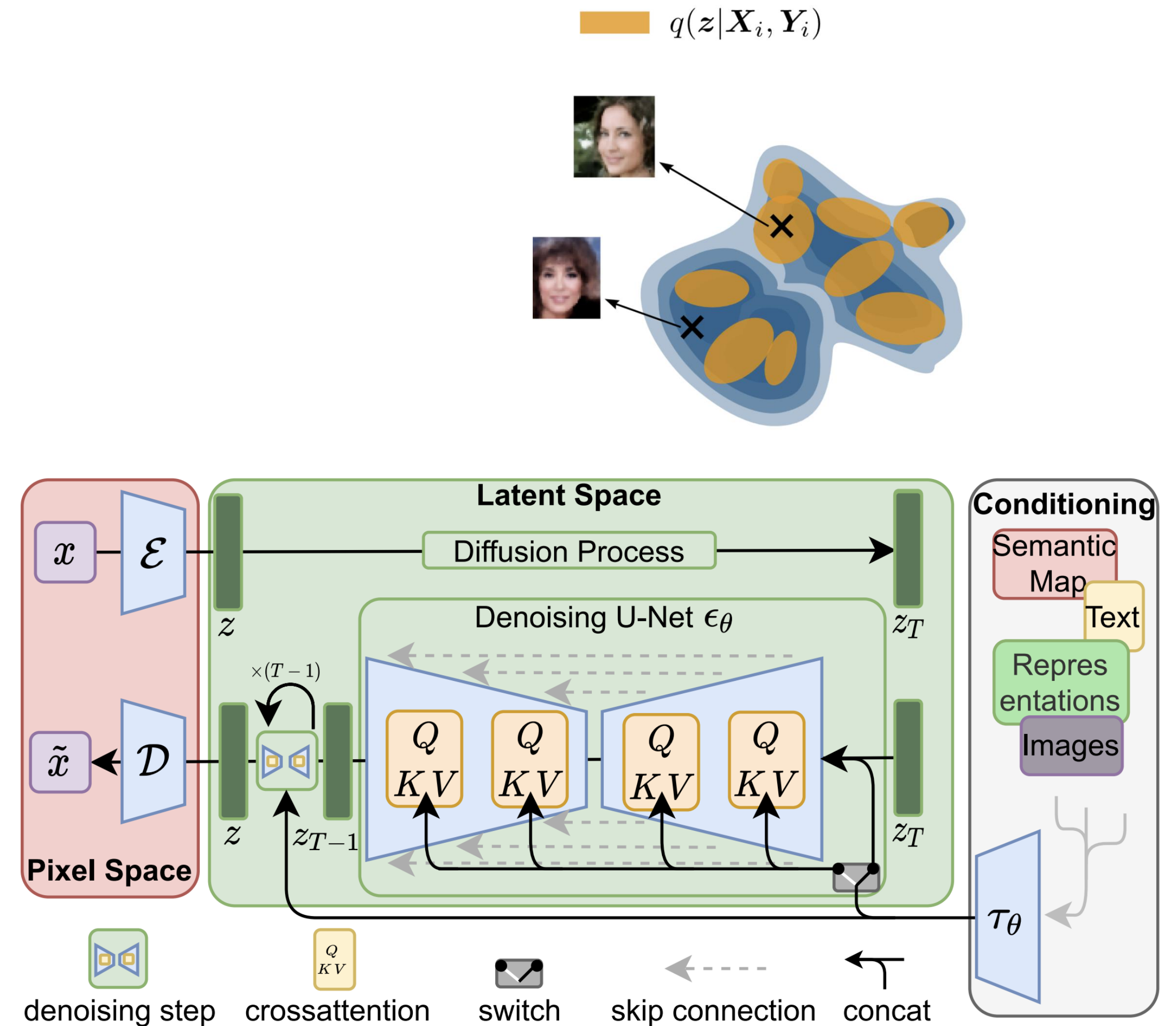
Latent Diffusion Models [11]

- First stage:

$$\begin{aligned} \mathcal{L}_{\text{VAE}}(\phi, \psi) = & \mathbb{E}_{q_{\psi}(z|\mathbf{X})} [\log p_{\Phi}(\mathbf{X})] \\ & - \beta \cdot D_{\text{KL}}(q_{\psi}(z|\mathbf{X}) \| p(z)), \\ & + \mathcal{L}_{\text{perceptual}} + \mathcal{L}_{\text{GAN}} \end{aligned}$$

- Second stage:

$$\mathcal{L}_{\text{DDPM}} = \mathbb{E}_{\mathbf{X}, z, \epsilon, t} \left[\lambda(t) \|\epsilon - \epsilon_{\theta}(z_t, t)\|^2 \right],$$



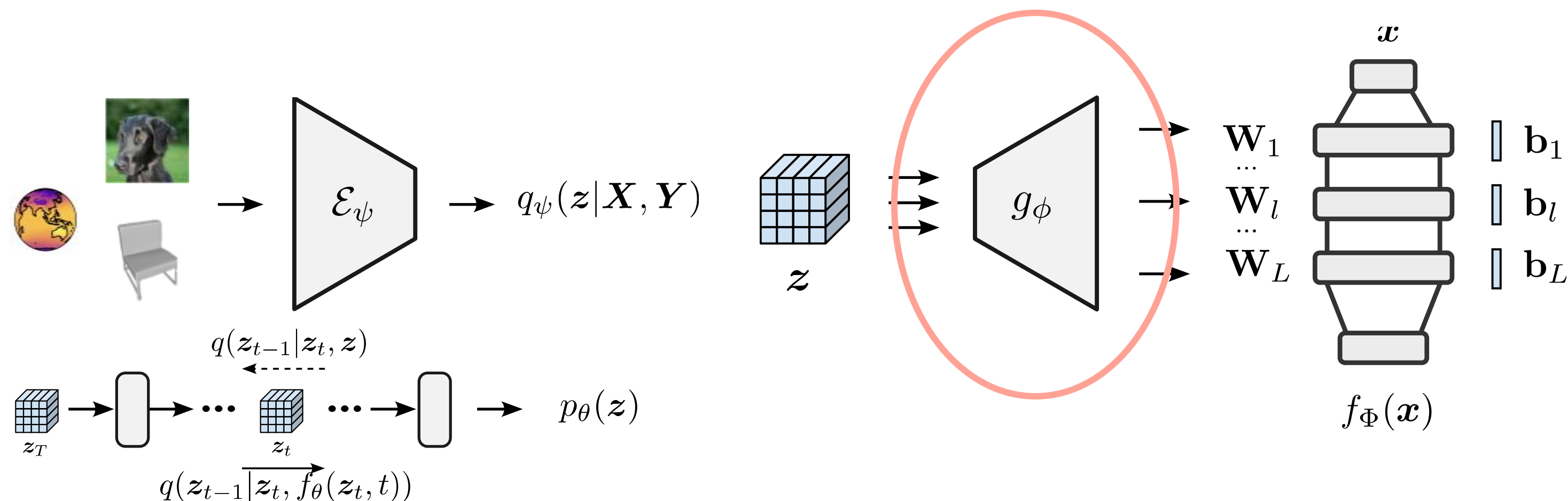
[11] Rombach et al., 2021

Latent Diffusion Models of INRs

Hyper-Transforming

- We can download pre-trained LDMs and just re-train only our decoder!

$$\mathcal{L}_{\text{HT}}(\phi) = \mathbb{E}_{q_{\psi}(z|\mathbf{X}, \mathbf{Y})} [\log p_{\Phi}(\mathbf{Y} | \mathbf{X})] + \mathcal{L}_{\text{perceptual}} + \mathcal{L}_{\text{GAN}}$$



Pretrained LDMs

Datset	Task	Model	FID	IS	Prec	Recall	
CelebA-HQ	Unconditional Image Synthesis	LDM-VQ-4 (200 DDIM steps, eta=0)	5.11 (5.11)	3.29	0.72	0.49	https://omr-diffusion/ci
FFHQ	Unconditional Image Synthesis	LDM-VQ-4 (200 DDIM steps, eta=1)	4.98 (4.98)	4.50 (4.50)	0.73	0.50	https://omr-diffusion/ff
LSUN-Churches	Unconditional Image Synthesis	LDM-KL-8 (400 DDIM steps, eta=0)	4.02 (4.02)	2.72	0.64	0.52	https://omr-diffusion/ls

Experiments

Datasets

CelebA-HQ (64x64)



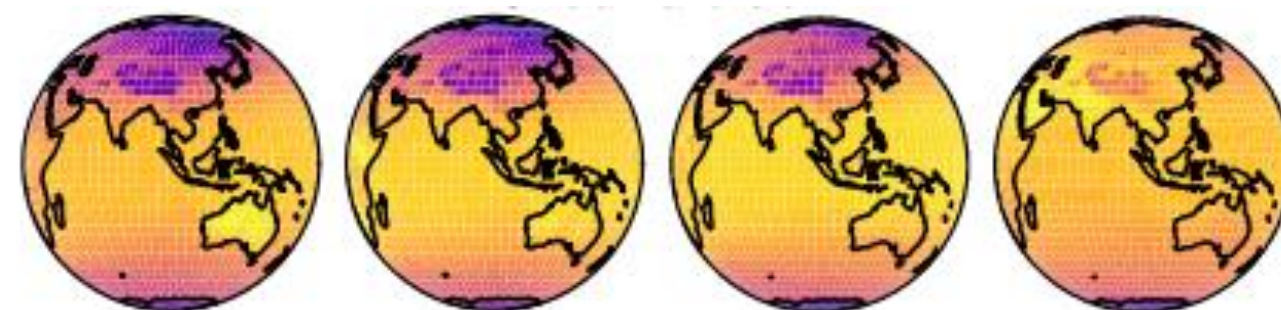
CelebA-HQ (256x256)



ImageNet (256x256)



ERA5 (Polar)



ShapeNET (Voxels)



Experiments

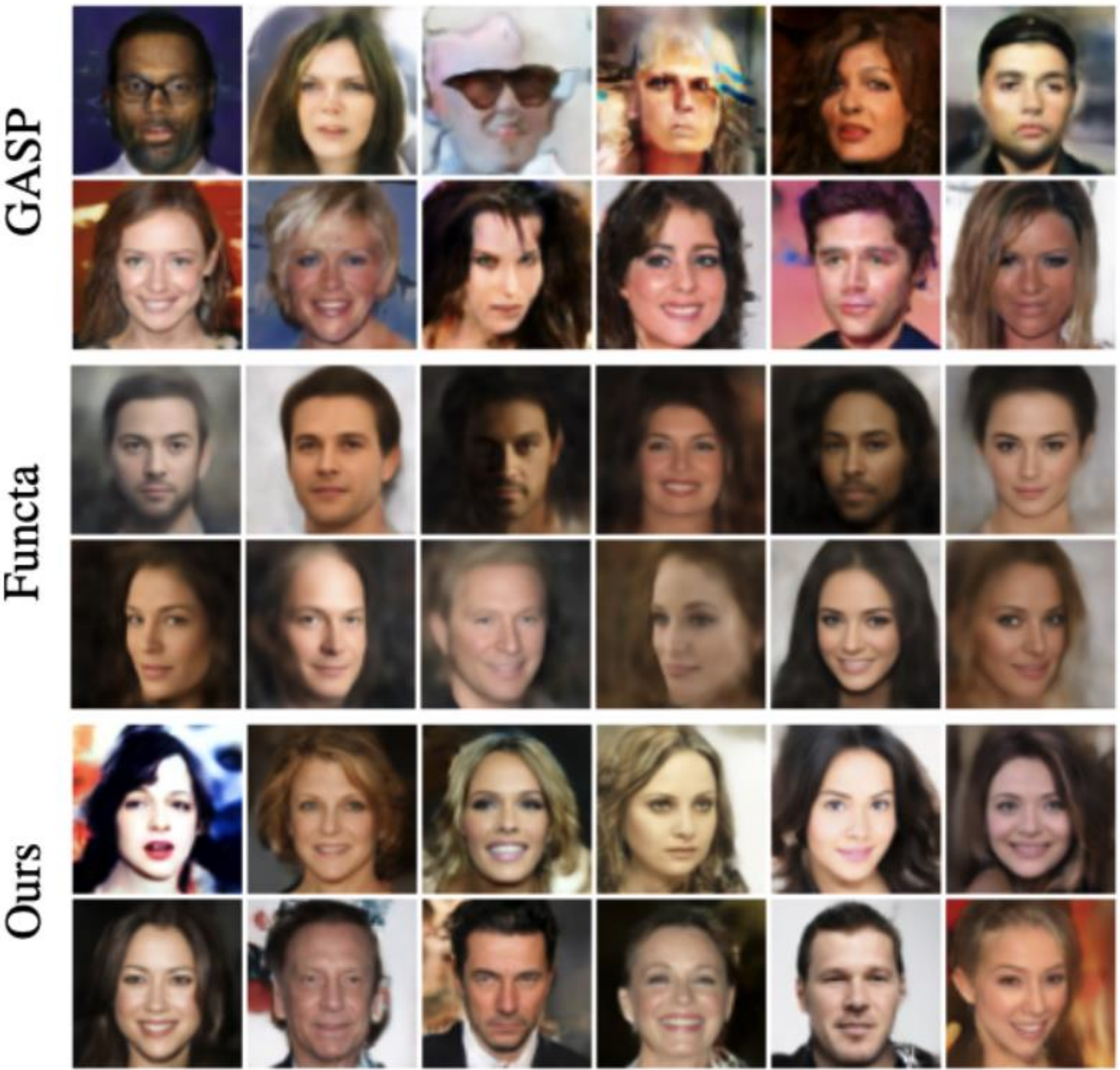
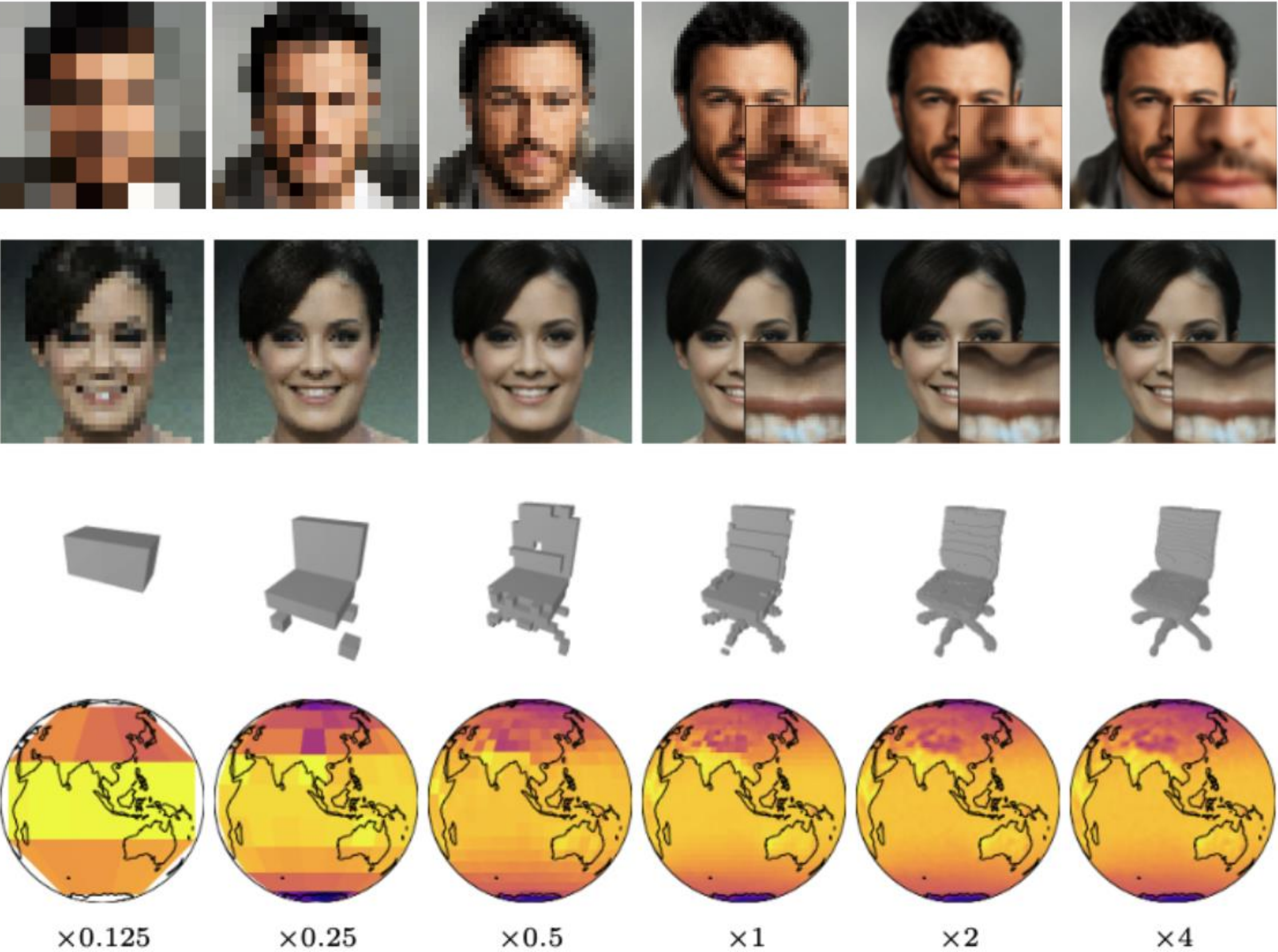
Baselines

Model	Approach	Training Procedure	Generation	Reconstruction, Imputation, Super Resolution	Scalable	Flexible
GASP (2021) [5]	GAN	Minimax	Forward Pass	✗	✗	✗
Functa (2022) [6]	Flow-based	Bilevel optimization	+ Extra Generative Model	Optimization procedure(s) per sample	✗	✗
VaMoH (ours)	VAE-based	Single optimization	Forward Pass	Forward pass	✗	✗
LDMI (ours)	LDM-based	Hyper-Transforming	Forward Pass	Forward pass	✓	✓

LDMI enhances efficiency, scalability quality of the learned representations.

Experiments

Generation



(a) CelebA-HQ

Experiments

Reconstruction

Model	PSNR (dB) $\uparrow$	FID $\downarrow$	HN Params $\downarrow$
CelebA-HQ (64 $\times$ 64)			
GASP [Dupont et al., 2022a]	-	7.42	25.7M
Functa [Dupont et al., 2022b]	≤ 30.7	40.40	-
VAMoH [Koyuncu et al., 2023]	23.17	66.27	25.7M
LDMI	24.80	18.06	8.06M
ImageNet (256 $\times$ 256)			
Spatial Functa [Bauer et al., 2023]	≤ 38.4	≤ 8.5	-
LDMI	20.69	6.94	102.78M

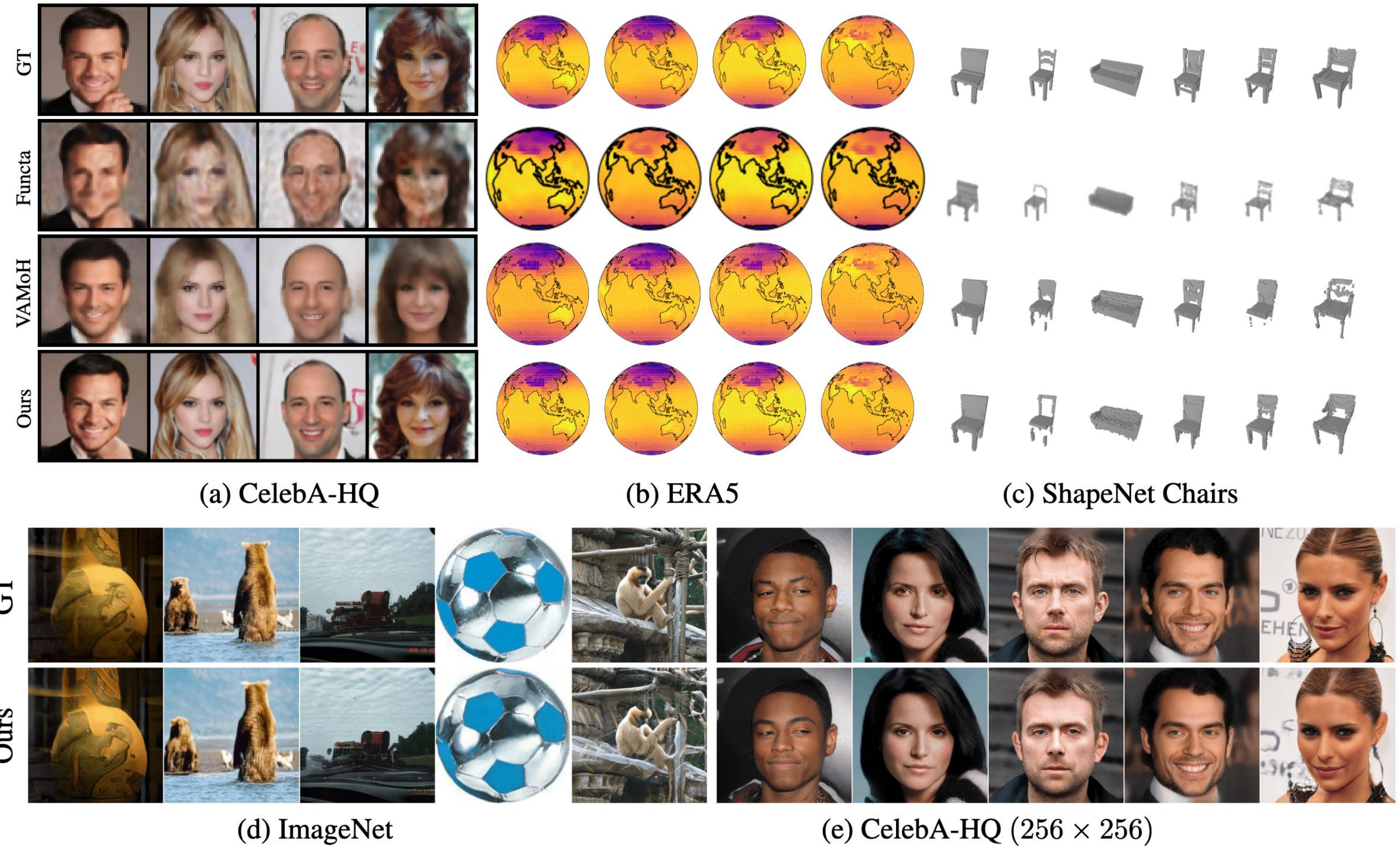
Table 1: Metrics on CelebA-HQ and ImageNet.

Model	Chairs (PSNR) $\uparrow$	ERA5 (PSNR) $\uparrow$
Functa [Dupont et al., 2022b]	29.2	34.9
VAMoH [Koyuncu et al., 2023]	38.4	39.0
LDMI	38.8	44.6

Table 2: Reconstruction quality (PSNR in dB) on ShapeNet Chairs and ERA5 climate data, demonstrating LDMI’s strong generalization capabilities across modalities. Note that GASP is omitted as it is not applicable to INR reconstruction tasks.

Method	HN Params	INR Weights	Ratio (INR/HN)
GASP/VAMoH	25.7M	50K	0.0019
LDMI	8.06M	330K	0.0409

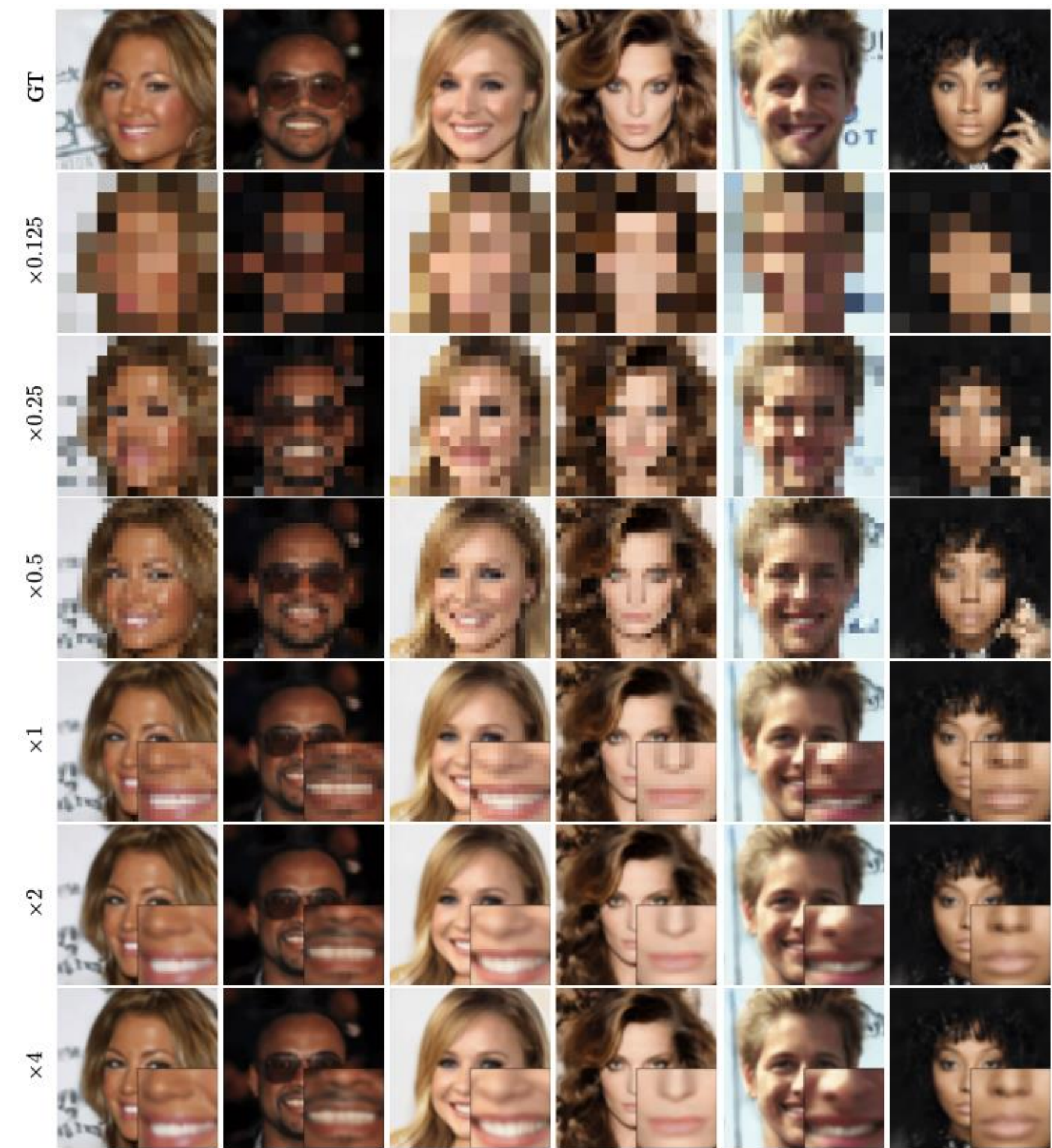
Table 3: Parameter efficiency of LDMI.



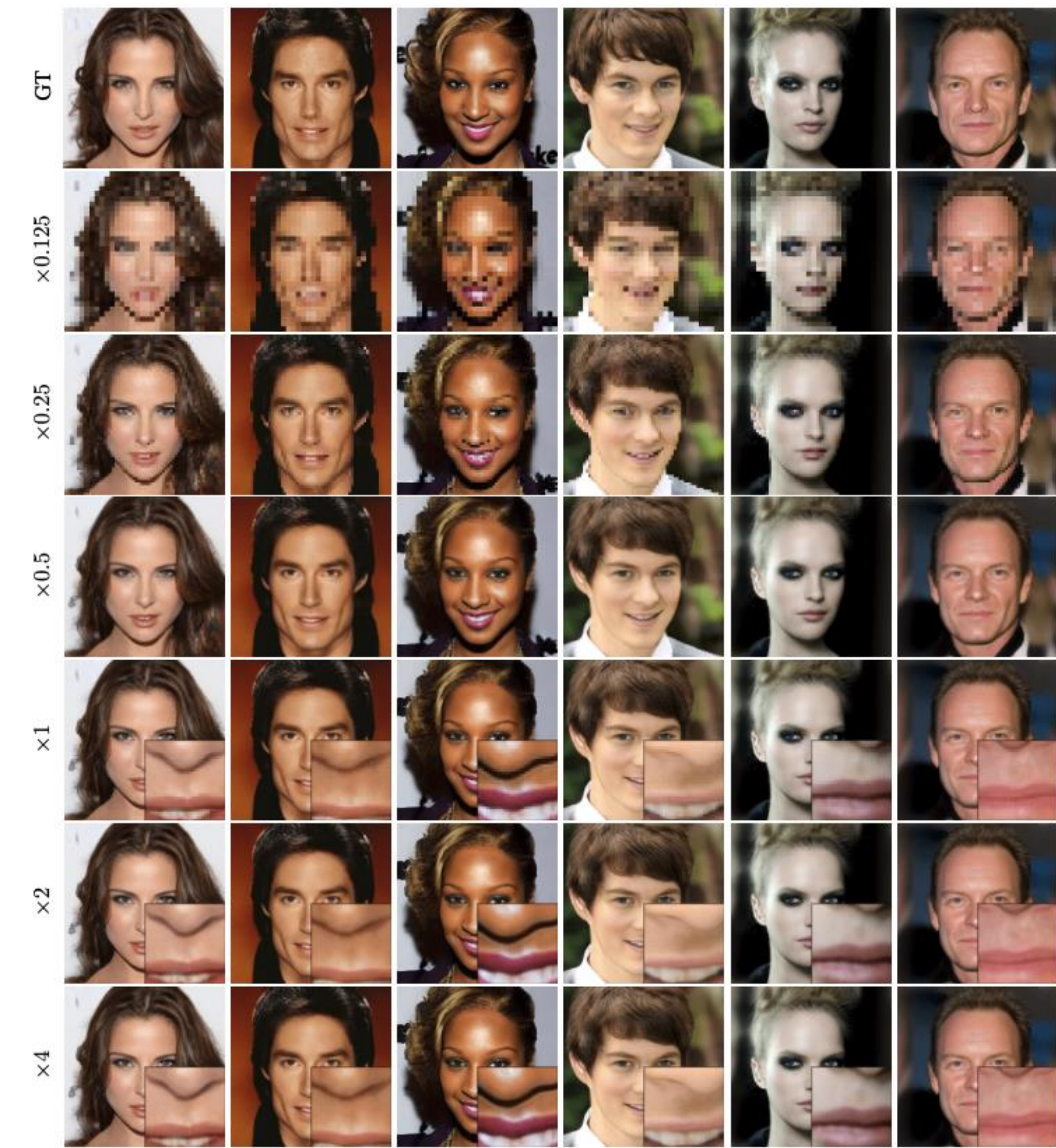
Experiments

Reconstruction

CelebA-HQ (64x64)



CelebA-HQ (256x256)



Experiments

Data completion

VAMoH



LDMI



Conclusion

Thanks to using **Latent Diffusion** and a novel **Transformer-based hypernetwork**, **LDMI** enhances

- Resolution-agnostic generation.
- Resolution-agnostic reconstruction.

While:

- ✓ Being scalable.
- ✓ Being efficient in parameter usage.
- ✓ Working with multiple data modalities.
- ✓ Allowing for generation of **bigger INRs** and more complex data.

Further details

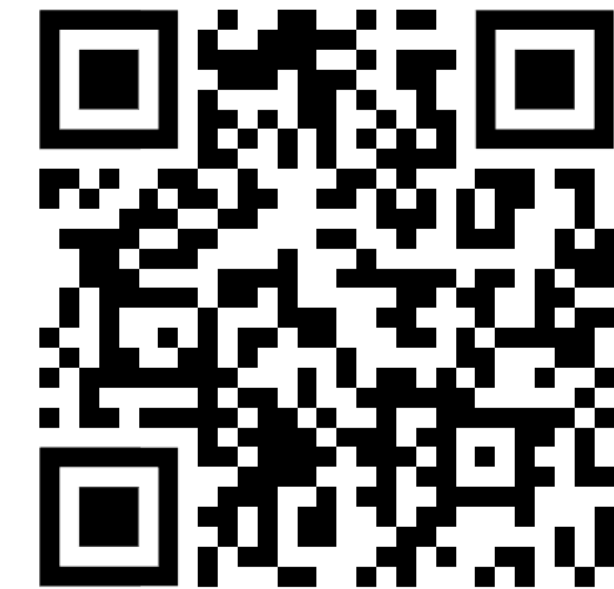
HYPER-TRANSFORMING LATENT DIFFUSION MODELS

Ignacio Peis*
Technical University of Denmark

Batuhan Koyuncu
Saarland University

Isabel Valera
Saarland University

Jes Frellsen
Technical University of Denmark



[\[Paper\]](#)

[1] Peis et al., 2025

Further details

VARIATIONAL MIXTURE OF HYPERGENERATORS FOR LEARNING DISTRIBUTIONS OVER FUNCTIONS

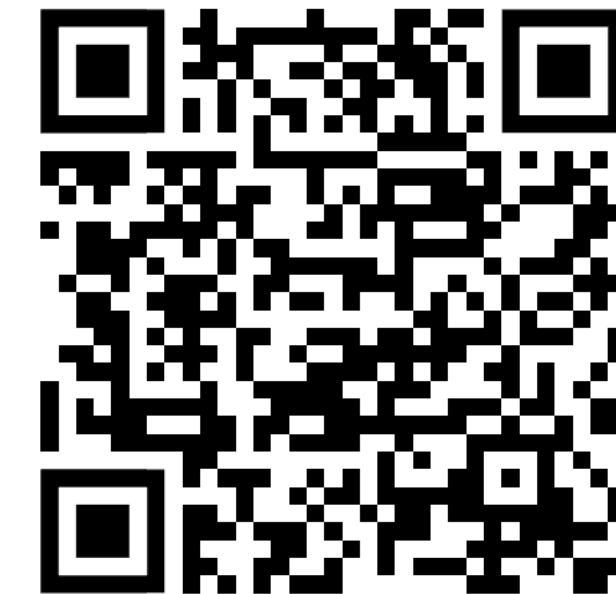
Batuhan Koyuncu*
Saarland University
Saarbrücken, Germany

Pablo Sánchez-Martín
Max Planck Institute for Intelligent Systems
Tübingen, Germany

Ignacio Peis
Universidad Carlos III de Madrid
Madrid, Spain

Pablo M. Olmos
Universidad Carlos III de Madrid
Madrid, Spain

Isabel Valera
Saarland University
Saarbrücken, Germany



[\[Paper\]](#)

[9] Koyuncu et al., 2023

References

- 📄 [1] [Peis, I., Koyuncu, B., Valera, I. & Frellsen, J. \(2025\). Hyper-Transforming Latent Diffusion Models. In International Conference on Machine Learning.](#)
- 📄 [2] Sitzmann, V., Martel, J., Bergman, A., Lindell, D., & Wetzstein, G. (2020). Implicit neural representations with periodic activation functions. *Advances in Neural Information Processing Systems*, 33, 7462-7473.
- 📄 [3] Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., & Geiger, A. (2019). Occupancy networks: Learning 3d reconstruction in function space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4460-4470).
- 📄 [4] Sitzmann, V., Zollhöfer, M., & Wetzstein, G. (2019). Scene representation networks: Continuous 3d-structure-aware neural scene representations. *Advances in Neural Information Processing Systems*, 32.
- 📄 [5] Ha, David, Andrew M. Dai, and Quoc V. Le. "HyperNetworks." International Conference on Learning Representations. 2017.
- 📄 [6] Dupont, E., Whye Teh, Y.; Doucet, A.. (2022). Generative Models as Distributions of Functions. *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics, in Proceedings of Machine Learning Research* 151:2989-3015.
- 📄 [7] Dupont, E., Kim, H., Eslami, S. A., Rezende, D. J., & Rosenbaum, D. (2022, June). From data to functa: Your data point is a function and you can treat it like one. In *International Conference on Machine Learning* (pp. 5694-5725). PMLR.

References

- 📄 [8] Bauer, M., Dupont, E., Brock, A., Rosenbaum, D., Schwarz, J. R., & Kim, H. (2023). Spatial functa: Scaling functa to imagenet classification and generation. arXiv preprint arXiv:2302.03130.
- 📄 [9] [Koyuncu, B., Sanchez-Martin, P., Peis, I., Olmos, P. M., & Valera, I. \(2023\). Variational Mixture of HyperGenerators for Learning Distributions Over Functions. In Proceedings of the 40th International Conference on Machine Learning, 2023.](#)
- 📄 [10] Chen, Yinbo, and Xiaolong Wang. "Transformers as meta-learners for implicit neural representations." European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022.
- 📄 [11] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10684-10695).
- 📄 [12]: Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., & Poole, B. Score-Based Generative Modeling through Stochastic Differential Equations. In International Conference on Learning Representations.
- 📄 [13]: Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in neural information processing systems, 33, 6840-6851.
- 📄 [14]: Song, J., Meng, C., & Ermon, S. Denoising Diffusion Implicit Models. In International Conference on Learning Representations.

Thank you!



ipeaz@dtu.dk

