# Anytime Sampling for Autoregressive Models via Ordered Autoencoding

Yilun Xu, Yang Song, Sahaj Garg, Linyuan Gong, Rui Shu, Aditya Grover, Stefano Ermon

ICLR 2021

Massachusetts Institute of Technology

Stanford University

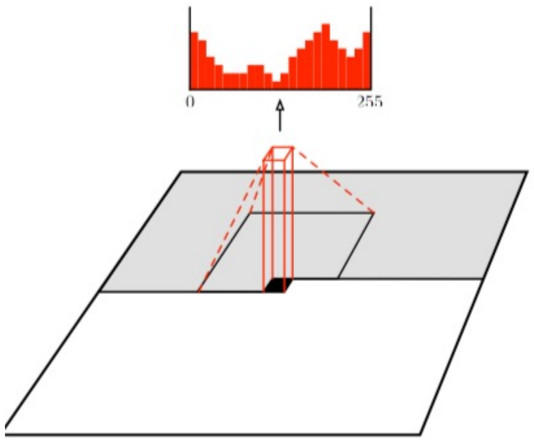
https://github.com/Newbeeer/Anytime-Auto-Regressive-Model

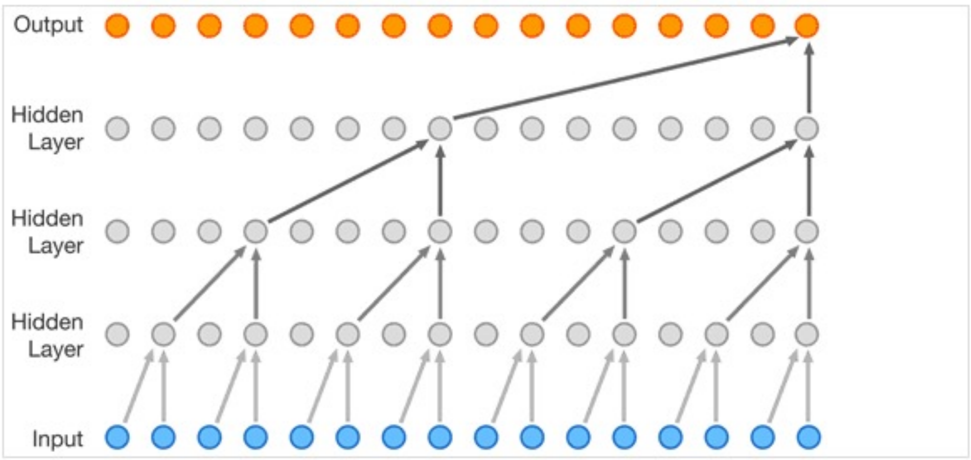
# An overview of auto-regressive models

$$p_\theta(\mathbf{x}) = \prod_{i=1}^{D} p_\theta(x_i \mid x_1, \cdots, x_{i-1})$$

MLE objective: $\mathcal{L} = \mathbb{E}_{p_d(\mathbf{x})}[\log p_\theta(\mathbf{x})]$

- Very powerful / large capacity
- Slow/Unnatural inference for traditional auto-regressive model
  - PixelCNN, PixelCNN++, Image-transformer, Gated-PixelCNN, WaveNet, ViT, etc



# Autoregressive on the spectral space

- Training autoregressive model on spectral space, e.g. latent space projected by the PCA matrix
  - Lossless
  - Support ***anytime sampling :***
    - **Trade-off on the fly to accommodate instantaneous resource** ***(Adaptive !)***
    - **Progressively generation**

☹ ***Inferior likelihoods / sample quality.***

# Learning the spectral space by Ordered Auto-encoder

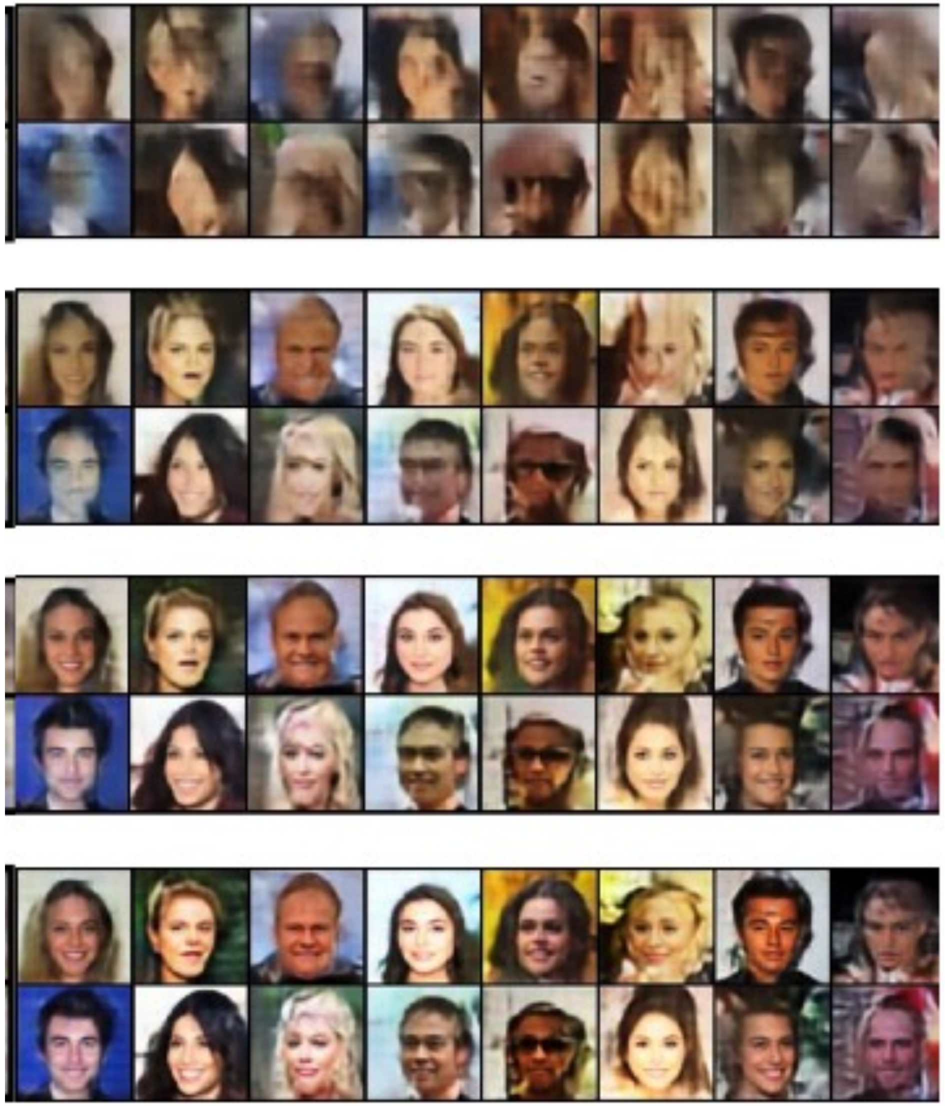
**Objective to induce the spectral space:**

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{K}\sum_{j=1}^{K}\|\mathbf{x}_i - d_\phi(e_\theta(\mathbf{x}_i)_{\leq j})_{\leq j}\|_2^2$$

- Recover PCA if we pick proper MLPs as encoder / decoder

- A coarse-to-fine order!

$$I(e_\theta(x)_i; x|e_\theta(x)_{\leq i-1}) \geq I(e_\theta(x)_{i-1}; x|e_\theta(x)_{\leq i-2})$$



# Ordered VQ-VAE

Apply the **ordered autoencoder** framework to the vector quantized variational autoencoder **(VQ-VAE)**

$$\frac{1}{K}\sum_{i=1}^{K}\left[\left\|d_\phi(\mathbf{z}^d_{\leq i})-\mathbf{x}\right\|_2^2+\left\|\mathbf{sg}[e_\theta(\mathbf{x})_{\leq i}]-\mathbf{z}^d_{\leq i}\right\|_F^2+\beta\left\|e_\theta(\mathbf{x})_{\leq i}-\mathbf{sg}[\mathbf{z}^d_{\leq i}]\right\|_F^2\right]$$

- Discrete latent codes
- Channel-Wise Quantization

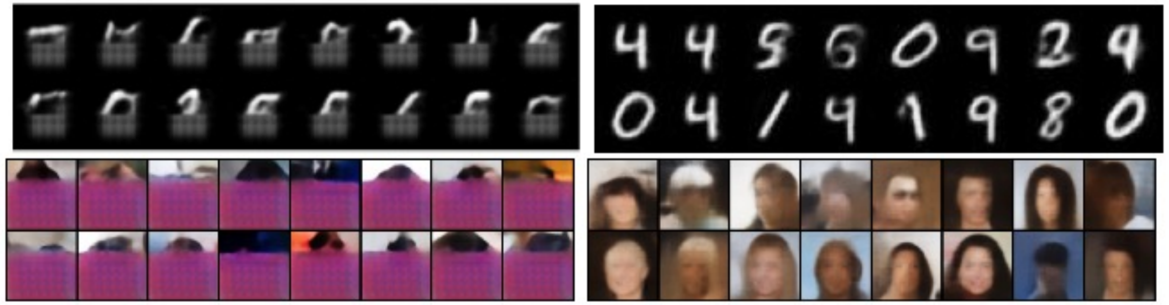
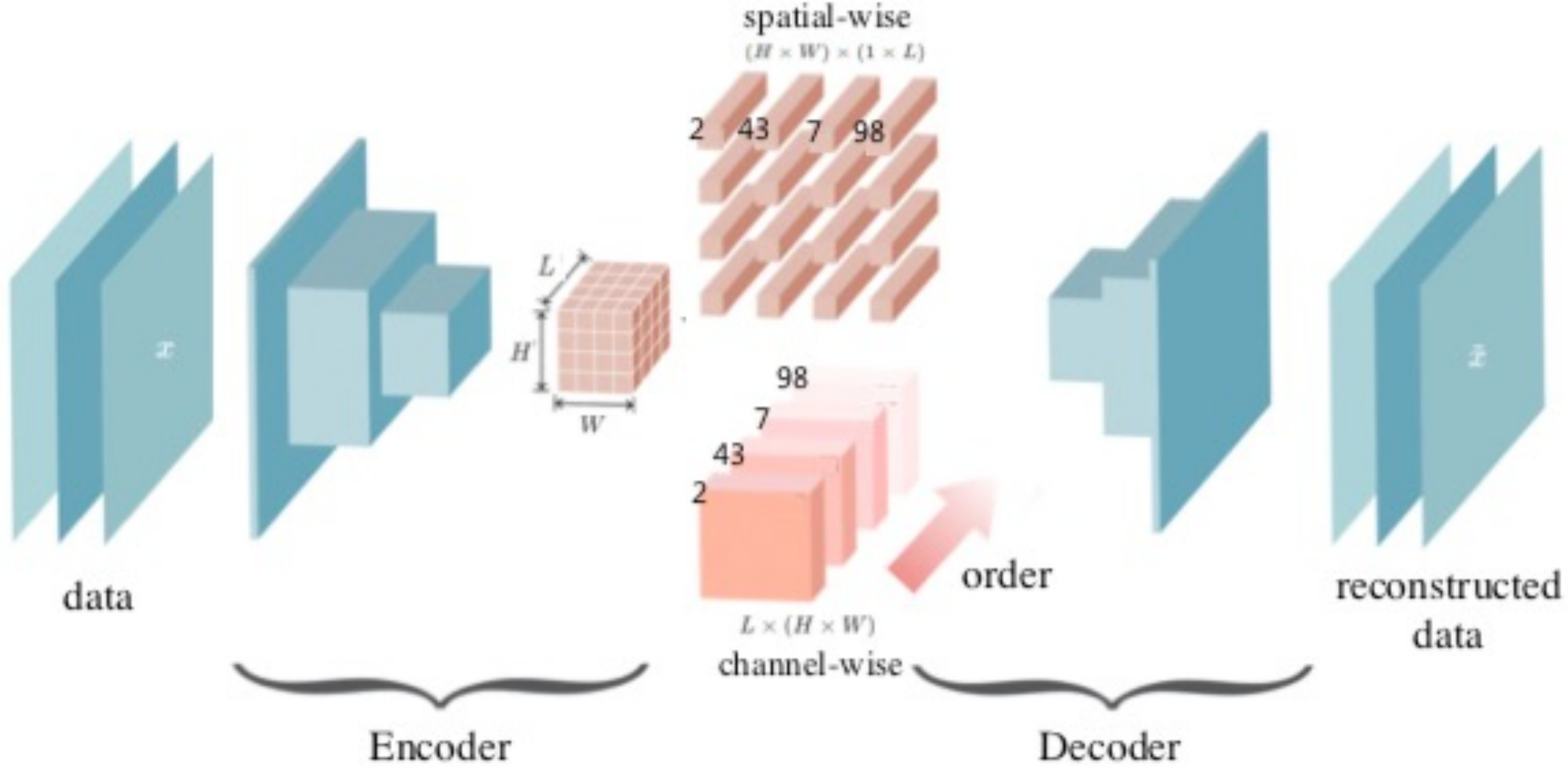


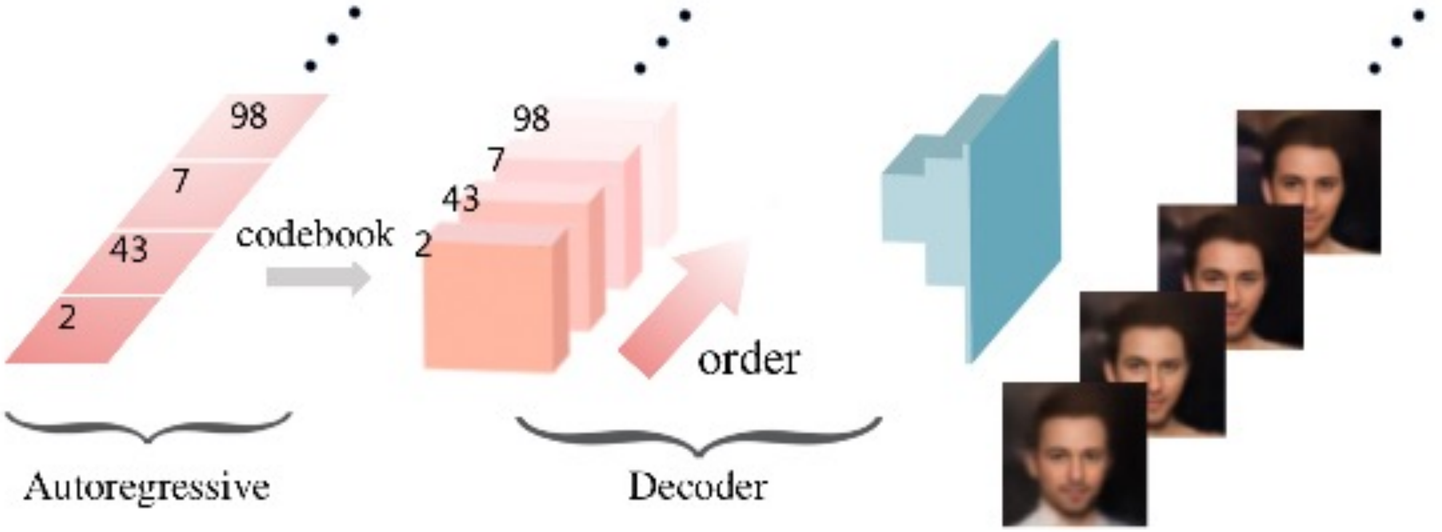
Figure 2: MNIST (**top**) and CelebA (**bottom**) samples generated with $^1/_4$ of the code length. **Left:** Spatial-wise quantization. **Right:** Channel-wise quantization.

# Ordered VQ-VAE

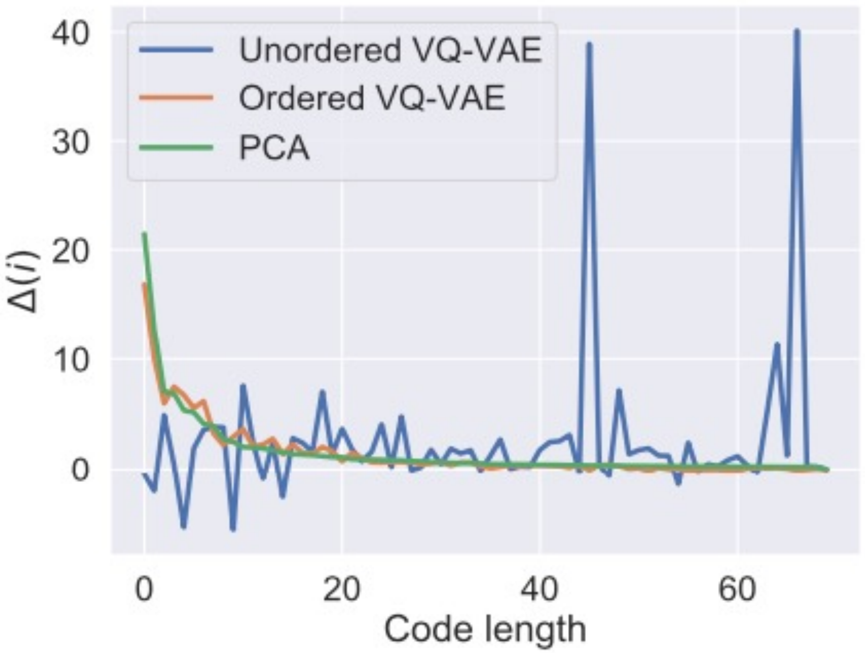


# Anytime Sampling

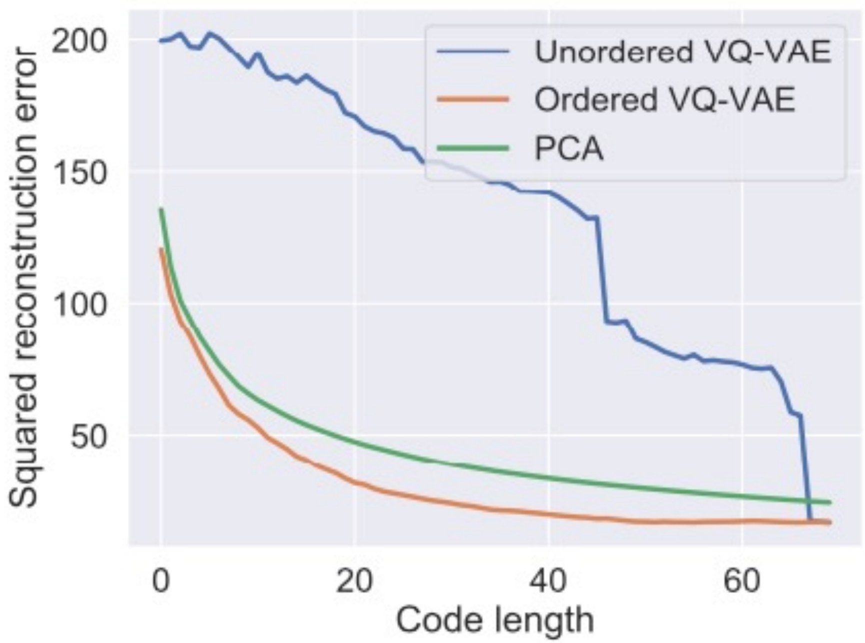
- Training auto-regressive model on the learned spectral space:
  **Ex-post density estimation** on the discrete codes
- We use **Transformer** as the auto-regressive model



# Spectral codes: ordering like PCA!




(a)

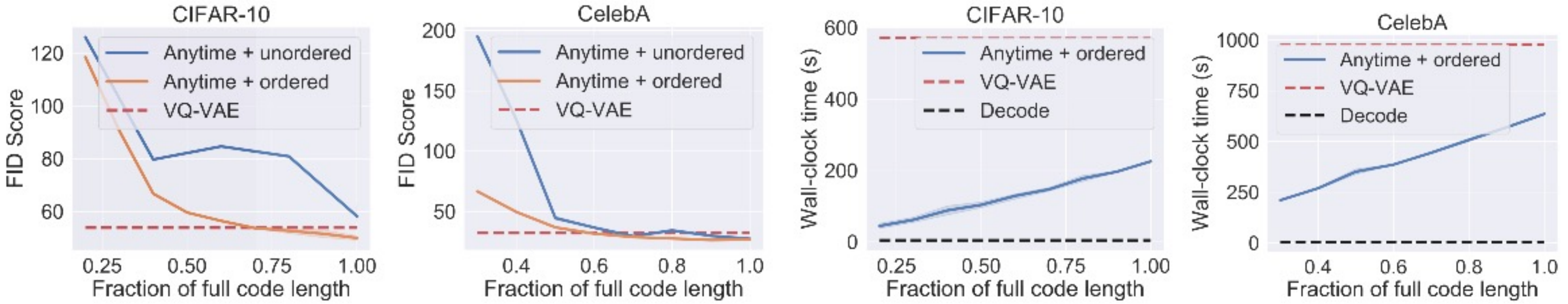



(b)

Figure 3: Ordered vs. unordered codes on CIFAR-10.

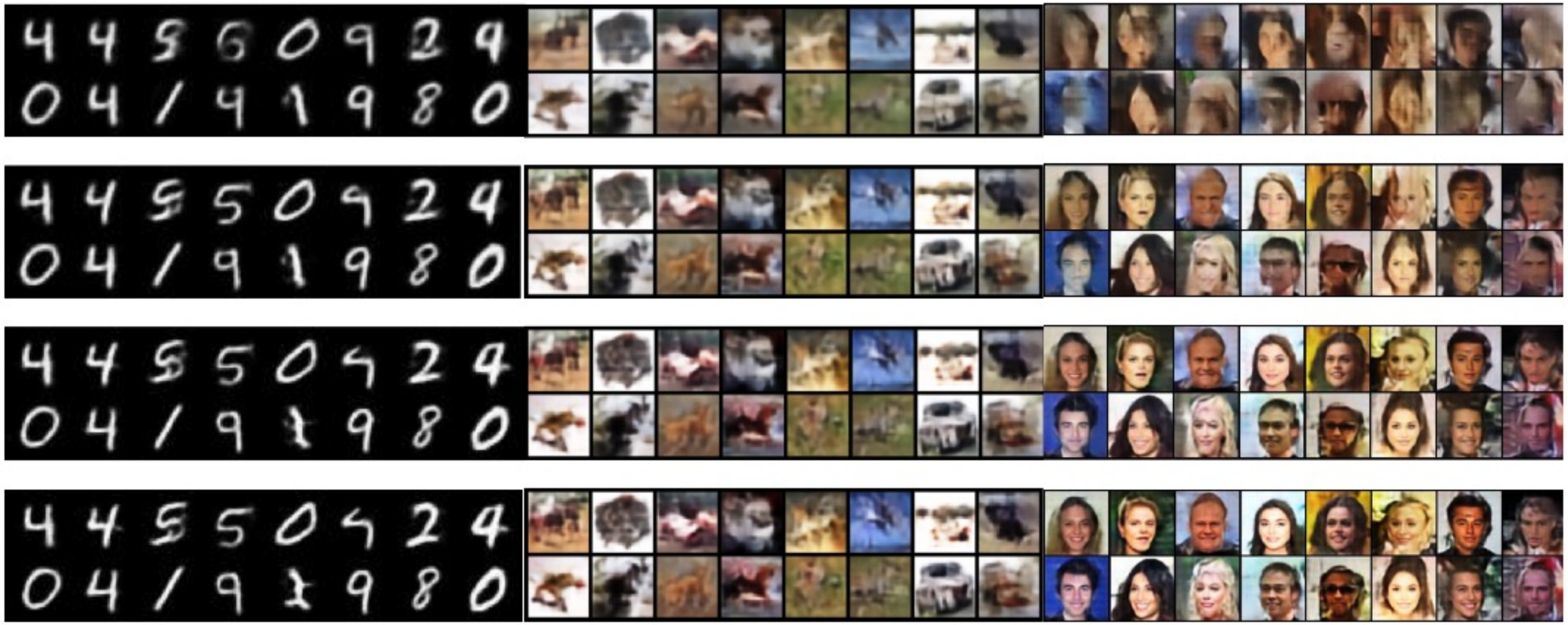
# Image generation

- **Strictly superior anytime sampling performance !**
- **Linearly increasing inference time**



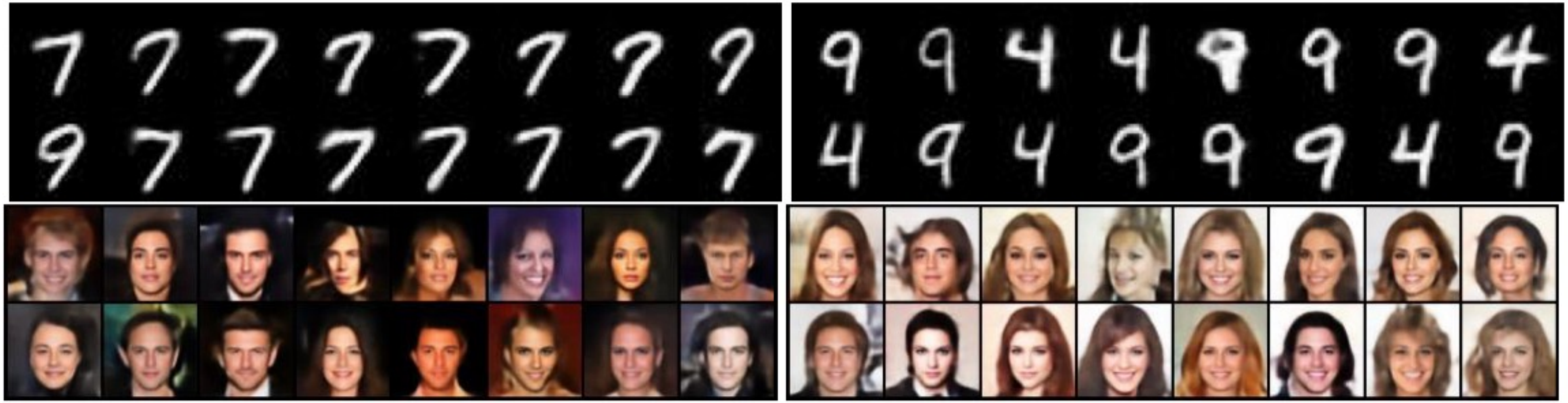

(a) Image quality (b) Inference speed

# Image generation



(a) MNIST (b) CIFAR-10 (c) CelebA

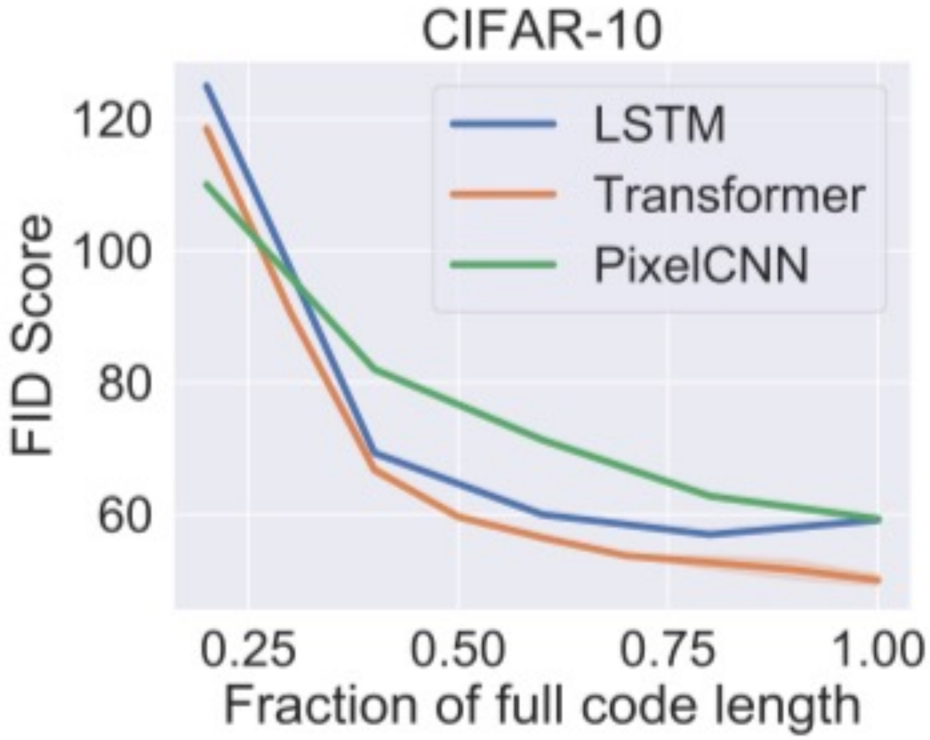
# Samples on the same priority code



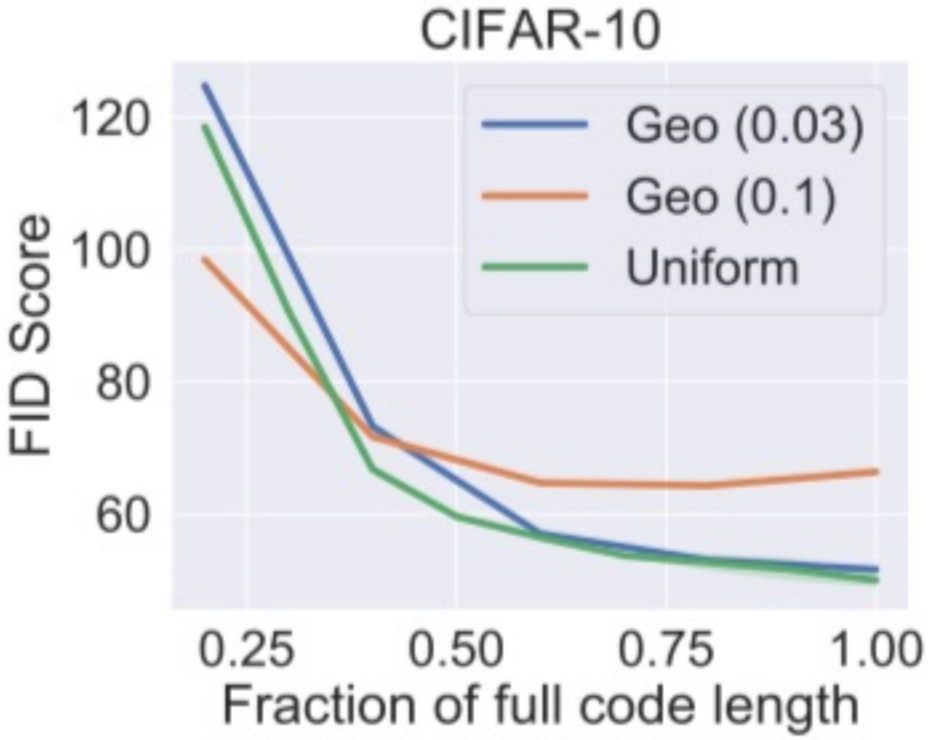
# Ablations

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{K}\sum_{j=1}^{K}\|\mathbf{x}_i - d_\phi(e_\theta(\mathbf{x}_i)_{\leq j})_{\leq j}\|_2^2$$

**Architectures:**




**Sampling Distributions:**



# Summary

- Propose an anytime sampling framework that can naturally trade-off computation resources with sample quality.

- With both theoretical arguments and empirical evidence, we show that ordered autoencoders can induce a valid ordering that facilitates anytime sampling.

- The sample quality of the anytime sampler degrades gracefully as we gradually reduce the code length.