# FAVAE-Effective Frequency Aware Latent Tokenizer

**Anonymous Author(s)**
Affiliation
Address
email

## Abstract

*Latent generative models have shown remarkable progress in high-fidelity image synthesis, typically using a two-stage training process that involves compressing images into latent embeddings via learned tokenizers in the first stage. The quality of generation strongly depends on how expressive and well-optimized these latent embeddings are. While various methods have been proposed to learn effective latent representations, the reconstructed images often lack realism, particularly in textured regions with sharp transitions, due to loss of fine details governed by high frequencies. We conduct a detailed frequency decomposition of existing state-of-the-art (SOTA) latent tokenizers and show that conventional objectives inherently prioritize low-frequency reconstruction, often at the expense of high-frequency fidelity. Our analysis reveals these latent tokenizers exhibit a bias toward low-frequency information, when jointly optimized, leading to over-smoothed outputs and visual artifacts that diminish perceptual quality. To address this, we propose a wavelet-based, frequency-aware variational autoencoder (FA-VAE) framework that explicitly decouples the optimization of low- and high-frequency components. This decoupling enables improved reconstruction of fine textures while preserving global structure. Our approach bridges the fidelity gap in current latent tokenizers and emphasizes the importance of frequency-aware optimization for realistic image representation, with broader implications for applications in content creation, neural rendering, and medical imaging.*

## 1 Introduction

Latent generative modeling [1, 2, 3, 4] has emerged as a cornerstone of modern content creation, with recent advances demonstrating remarkable capabilities in synthesizing high-fidelity visual content. These models typically operate in compressed latent spaces learned via autoencoders such as VAEs [5, 6], where generation quality is directly influenced by the expressiveness of these latent embeddings. While increasing latent dimensionality can enhance representational power, this often leads to diminishing visual outputs and increased computational cost. VAVAE [7] addresses this by aligning the latent space with pre-trained vision foundation models to further improve convergence and generative realism. However, despite such improvements, generated outputs or even the reconstructed outputs often lack sharp textures and fine details, particularly in regions dominated by the high-frequency information such as text on images. This limits perceptual realism and results in overly smooth outputs. Prior works [8, 9, 10] have explored architectural and spectral enhancements to inject high-frequency signals, but a systematic frequency-level analysis of how these current latent tokenizers influence reconstruction quality, especially in differentiating low- vs. high-frequency information remains absent.

In this work, we conduct a comprehensive frequency-based analysis of the reconstruction behavior of state-of-the-art latent tokenizers used in generative pipelines [7, 6, 11, 12, 13], focusing on latent diffusion and autoregressive models, which represent the current forefront of high-quality

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

visual generation. These models typically follow a two-stage pipeline: (1) learning quantized or non-quantized latent embeddings using a latent tokenizer like VAE or its variant, and (2) training a probabilistic generative model in the latent embedding space. Our analysis reveals a consistent bias during the first stage of learning latent embeddings: while low-frequency components are on par well reconstructed, high-frequency signals such as textures are poorly preserved. This low frequency bias over high frequencies in optimization contributes significantly to perceptual degradation.

To address this fidelity gap, we propose a novel **Frequency-aware VAE (FA-VAE)** latent tokenizer that explicitly decouples and separately optimizes low- and high-frequency components using wavelet decomposition. Our method learns distinct latent embeddings for low-high frequency subbands and later fuses them into a unified latent space representation. This simple design allows for better preservation of both global structure and fine details. Figure 1 presents a visual comparison between our method and the recent state-of-the-art VAVAE, highlighting improved preservation of fine details and sharp structures.

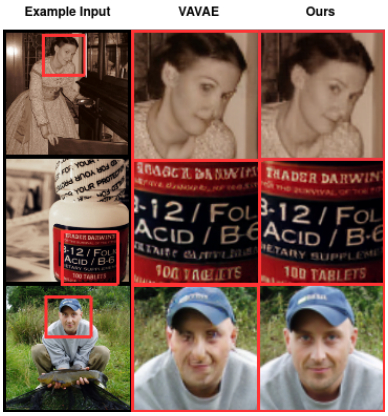


**Figure 1:** Visual comparison of reconstructions. From left to right: example original image, VAVAE reconstruction, and our approach (FA-VAE). The red color highlighted regions emphasize areas rich in textures, edges, and text. Our method better preserves high-frequency details and sharp structures, resulting in reconstructions visually closer to the input.

**Our main contributions are summarized as follows:**

- ✓ We provide a frequency-based analysis of latent tokenizers used in latent generative pipelines, showing that existing VAE variants disproportionately emphasize low-frequency components at the expense of fine detail.
- ✓ We introduce **FA-VAE**, a frequency-aware VAE framework that decouples optimization of low- and high-frequency subbands using wavelet decomposition, resulting in frequency aware expressive latent representations.

# 2 Method

## 2.1 Frequency Evaluation of Latent Embeddings

Latent tokenization models typically begin with a Variational Autoencoder (VAE) or its tokenizer variants, which learn compact embeddings $\mathbf{z} \in \mathcal{Z}$ from input data $\mathbf{x} \in \mathcal{X}$. Reconstruction quality is often measured by the pixel-level error.

$$\mathcal{L}_{\text{rec}} = \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2, \quad \hat{\mathbf{x}} = D(E(\mathbf{x})),$$

where $E$ and $D$ are the encoder and decoder.

To study fidelity across frequency bands, we apply a discrete wavelet transform (DWT) with Haar filter to decompose signals into low- and high-frequency bands with one decomposition level:

$$\mathcal{W}(\mathbf{x}) = (\mathbf{x}_L, \mathbf{x}_H), \quad \mathcal{W}(\hat{\mathbf{x}}) = (\hat{\mathbf{x}}_L, \hat{\mathbf{x}}_H),$$

Then, we compute the frequency-aware reconstruction losses as follows:

$$\mathcal{L}_L = \|\mathbf{x}_L - \hat{\mathbf{x}}_L\|_2^2, \quad \mathcal{L}_H = \|\mathbf{x}_H - \hat{\mathbf{x}}_H\|_2^2.$$

Beyond pixel errors, perceptual similarity is captured by LPIPS [14], and distributional alignment by reconstruction FID (rFID) [15], computed from Inception features [16]. Together, combining frequency-aware losses $(\mathcal{L}_L, \mathcal{L}_H)$, perceptual similarity (LPIPS), and rFID provides a more complete evaluation of reconstruction fidelity and expressiveness of the learned latent embeddings.

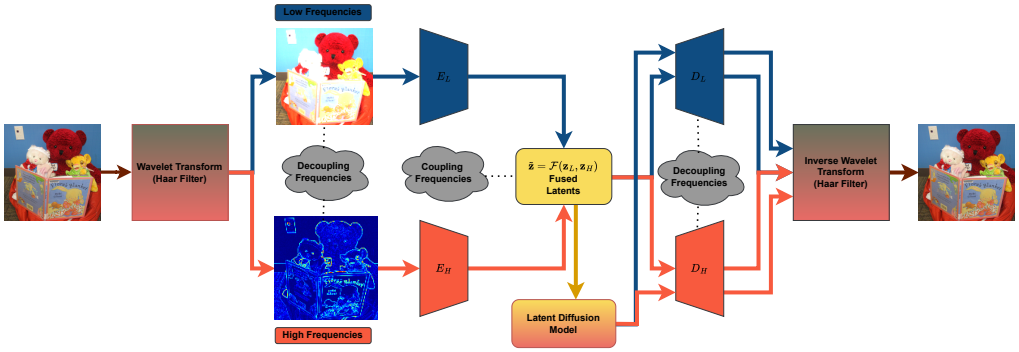


**Figure 2: FA-VAE.** The Overall pipeline of FA-VAE latent tokenizer.

### 2.2 Frequency-Aware VAE (FA-VAE)

The overview of FA-VAE is shown in Fig. 2. To enhance the fidelity of latent embeddings, we extend the standard VAE formulation by decoupling the input image into low- and high-frequency components, which are then learned independently. Given an input image $\mathbf{x} \in \mathcal{X}$, we apply a discrete wavelet transform $\mathcal{W}(\cdot)$ to obtain:

$$\mathcal{W}(\mathbf{x}) = (\mathbf{x}_L, \mathbf{x}_H),$$

where $\mathbf{x}_L$ and $\mathbf{x}_H$ denote the low- and high-frequency representations, respectively. We use the Haar filter for wavelet transformation, and apply the normalization strategy proposed in [17] to normalize both low and high frequency components. We employ separate encoder–decoder pairs $(E_L, D_L)$ and $(E_H, D_H)$ to learn frequency-specific latent embeddings:

$$\mathbf{z}_L = E_L(\mathbf{x}_L), \quad \mathbf{z}_H = E_H(\mathbf{x}_H).$$

**Low-Frequency Objective.** To learn low-frequency latent embeddings, we adopt a VA-VAE-style objective [7], incorporating a vision foundation alignment loss $\mathcal{L}^L_{\text{VF}}$ [7], adversarial regularization $\mathcal{L}^L_{\text{GAN}}$ inspired by VQGAN [18], and an additional perceptual loss $\mathcal{L}^L_{\text{LPIPS}}$ [14] to improve visual quality. The total low-frequency objective is defined as:

$$\mathcal{L}_{\text{low}} = \mathcal{L}^L_{\text{rec}} + \beta \cdot \mathcal{L}^L_{\text{KL}} + \lambda_{\text{VF}} \cdot \mathcal{L}^L_{\text{VF}} + \lambda_{\text{GAN}} \cdot \mathcal{L}^L_{\text{GAN}} + \lambda_{\text{LPIPS}} \cdot \mathcal{L}^L_{\text{LPIPS}}$$

where:

$$\mathcal{L}^L_{\text{rec}} = \mathbb{E}_{q(\mathbf{z}_L|\mathbf{x}_L)} \left[ \|\mathbf{x}_L - D_L(\mathbf{z}_L)\|_2^2 \right],$$
$$\mathcal{L}^L_{\text{KL}} = D_{\text{KL}} \left( q(\mathbf{z}_L \mid \mathbf{x}_L) \,\|\, p(\mathbf{z}_L) \right),$$

with $q(\mathbf{z}_L \mid \mathbf{x}_L)$ being the encoder's approximate posterior and $p(\mathbf{z}_L) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, the standard gaussian prior. The vision foundation loss $\mathcal{L}^L_{\text{VF}}$ aligns the latent codes with the feature space of a pretrained foundation model (e.g., DINOv2 [19]). The adversarial loss $\mathcal{L}^L_{\text{GAN}}$ introduces a discriminator to distinguish real from reconstructed low-frequency inputs. Finally, the perceptual loss $\mathcal{L}^L_{\text{LPIPS}}$ encourages perceptual similarity between input and reconstruction using features from a pretrained Inception network [16]. All these loss components are commonly used in modern VAE frameworks.

**High-Frequency Objective.** In contrast, high-frequency components are trained without supervision from pretrained models, as they tend to be biased toward low-frequency content. We use a lightweight VAE objective focused on reconstructing fine-scale details, along with adversarial regularization:

$$\mathcal{L}_{\text{high}} = \mathcal{L}^H_{\text{rec}} + \beta \cdot \mathcal{L}^H_{\text{KL}} + \mathcal{L}^H_{\text{GAN}},$$
$$\mathcal{L}^H_{\text{rec}} = \mathbb{E}_{q(\mathbf{z}_H|\mathbf{x}_H)} \left[ \|\mathbf{x}_H - D_H(\mathbf{z}_H)\|_1 \right],$$
$$\mathcal{L}^H_{\text{KL}} = D_{\text{KL}} \left( q(\mathbf{z}_H \mid \mathbf{x}_H) \,\|\, p(\mathbf{z}_H) \right)$$

**At inference**, both decoders reconstruct the respective frequency bands, and the final image is synthesized via inverse wavelet transform. This frequency-aware latent tokenization via FA-VAE yields more expressive embeddings by preserving details across the full spectrum of spatial frequencies.

$$\hat{\mathbf{x}}_L = D_L(\mathbf{z}_L), \quad \hat{\mathbf{x}}_H = D_H(\mathbf{z}_H), \quad \hat{\mathbf{x}} = \mathcal{W}^{-1}(\hat{\mathbf{x}}_L, \hat{\mathbf{x}}_H).$$

| Model (Tokenizer) | Tokenizer Config | Recon. Loss | Low Freq. Loss | High Freq. Loss | LPIPS | rFID |
|---|---|---|---|---|---|---|
| DC-AE | f32c32 | 0.0194 | 0.0484 | 0.0097 | 0.1580 | 0.7450 |
| DC-AE | f64c128 | 0.0207 | 0.0527 | 0.0100 | 0.1667 | 0.7623 |
| DC-AE | f128c512 | 0.0225 | 0.0581 | 0.0106 | 0.1805 | 0.7912 |
| SD-VAE | f16c16 | 0.0180 | 0.0422 | 0.0100 | 0.1743 | 0.6213 |
| KL-VAE | f16c16 | 0.0148 | 0.0326 | 0.0089 | 0.1355 | 0.5318 |
| MS-VQ-VAE | f16c32v4096 | 0.0195 | 0.0549 | 0.0076 | 0.1890 | 0.6981 |
| RQ-VAE | f32c256v16384 | 0.0416 | 0.1219 | 0.0149 | 0.2712 | 0.9095 |
| TiTok-VQ-VAE | f256x2c64v8192 | 0.0226 | 0.0561 | 0.0114 | 0.2082 | 0.7550 |
| TiTok-VQ-VAE | f256x4c64v8192 | 0.0339 | 0.0947 | 0.0137 | 0.2657 | 0.8823 |
| TiTok-VQ-VAE | f256x8c64v8192 | 0.0450 | 0.1344 | 0.0152 | 0.2949 | 0.9416 |
| TiTok-VAE | f256x1c16 | 0.0332 | 0.0923 | 0.0134 | 0.2232 | 0.8640 |
| TiTok-VAE | f256x2c16 | 0.0461 | 0.1380 | 0.0155 | 0.2682 | 0.9187 |
| TiTok-VAE | f256x4c16 | 0.0617 | 0.1961 | 0.0170 | 0.3182 | 0.9761 |
| VQ-VAE | f16c256v1024 | 0.0492 | 0.1478 | 0.0163 | 0.3064 | 0.9102 |
| VQ-VAE | f16c256v16384 | 0.0438 | 0.1262 | 0.0164 | 0.2784 | 0.8826 |
| **VA-VAE** | **f16c32** | **0.0105** | **0.0200** | **0.0074** | **0.0975** | **0.4884** |
| **FA-VAE (Ours)** | **f16c32** | **0.0044** | **0.0114** | **0.0020** | **0.0940** | **0.4156** |

**Table 1:** Quantitative comparison of latent tokenizers. Configurations: **f** = latent spatial resolution, **c** = latent dimensionality, **v** = vocabulary size (for quantized models). Lower is better.

# 3 Experiments

To evaluate the reconstruction performance of latent tokenizers in both the spatial and spectral (frequency) domains, we consider a range of widely adopted visual tokenizers commonly used in modern latent generative models. These tokenizers are based on different variants of autoencoding architectures, each aiming to learn compact and informative latent embedding representations of the input data distribution. For a comprehensive analysis, we include recent representative tokenizers based on standard Autoencoders [20], Variational Autoencoders (VAEs) [21, 6, 22, 23, 7], and Vector Quantized Autoencoders (VQ-VAEs) [23, 18, 11]. We also include our proposed FA-VAE, which explicitly incorporates frequency-awareness into the latent embedding learning process. This selection enables a fair comparison across diverse tokenization strategies with varying latent dimensionalities (c) and latent feature resolutions (f) as shown in Table 1 with diverse tokenization configurations. We use the metrics discussed in method section 2.1 for our evaluation.

Our analysis in Table 1 shows that KL-regularized VAEs with carefully designed parameterizations consistently outperform both standard VAEs and VQ-VAE baselines across all reconstruction metrics, achieving efficient performance at higher latent compression rates due to the absence of quantization artifacts. Their effectiveness is strongly tied to latent parameterization quality, as seen in recent autoregressive models [24] and VA-VAE [7], which further align latent spaces with foundation models (e.g., DINOv2 [19]) for improved perceptual quality. However, such models jointly optimize frequency components, leading to trade-offs and suboptimal preservation of low and high-frequency details (Especially high frequency details). In contrast, our FA-VAE explicitly decouples low- and high-frequency bands, enabling specialized representation learning and more precise reconstructions of both global structures and fine details. FA-VAE nearly halves the reconstruction loss of the strongest baseline while achieving the best overall performance across both low and high frequency reconstructions, demonstrating the benefits of frequency-aware modeling. The consistent gap between low- and high-frequency reconstructions across models further underscores the general difficulty of capturing fine details effectively. The other metrics like LPIPS and rFID showcases the perceptual quality of reconstructions of FA-VAE compared with other baselines.

# 4 Conclusion

In this work, we investigate frequency awareness in the latent embeddings of latent tokenizer models. We find that jointly optimizing low- and high-frequency components leads to a frequency bias favoring low-frequencies thereby degrading the reconstruction of fine-grained, high-frequency details and overall perceptual quality. To address this limitation, we introduce **FA-VAE**, a frequency-aware VAE framework that explicitly decouples low and high-frequency components using wavelet decomposition, processes them independently, and fuses them into a unified latent representation. FA-VAE achieves state-of-the-art performance across frequency-aware reconstruction metrics, demonstrating improved fidelity and perceptual quality of the learned latent embeddings.

# References


[1] Andreas Blattmann, Robin Rombach, Huan Ling, Tim Dockhorn, Seung Wook Kim, Sanja Fidler, and Karsten Kreis. Align your latents: High-resolution video synthesis with latent diffusion models. In *Proc. CVPR 2023*, 2023.

[2] Yuan Dong, Qi Zuo, Xiaodong Gu, Weihao Yuan, Zhengyi Zhao, Zilong Dong, Liefeng Bo, and Qixing Huang. Gpld3d: Latent diffusion of 3d shape generative models by enforcing geometric and physical priors. In *Proc. CVPR 2024*, page 56–66, 2024.

[3] Simian Luo, Yiqin Tan, Longbo Huang, Jian Li, and Hang Zhao. Latent consistency models: Synthesizing high-resolution images with few-step inference. In *Proc. ICLR 2024*, 2024.

[4] Cai Zhou, Xiyuan Wang, and Muhan Zhang. Unifying generation and prediction on graphs with latent graph diffusion. In *Proc. NeurIPS 2024*, 2024.

[5] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[6] Tianhong Li, Yonglong Tian, He Li, Mingyang Deng, and Kaiming He. Autoregressive image generation without vector quantization. *arXiv preprint arXiv:2406.11838*, 2024.

[7] Jingfeng Yao, Bin Yang, and Xinggang Wang. Reconstruction vs. generation: Taming optimization dilemma in latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025.

[8] Hao Phung, Quan Dao, and Anh Tran. Wavelet diffusion models are fast and scalable image generators. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10199–10208, June 2023.

[9] Yueying Li, Hanbin Zhao, Jiaqing Zhou, Guozhi Xu, Tianlei Hu, Gang Chen, and Haobo Wang. Fedsr: Frequency-aware enhancement for diffusion-based image super-resolution. In *ICLR 2025 (withdrawn)*, 2024. Amplitude and high-frequency enhancement modules for diffusion SR.

[10] Florentin Guth, Simon Coste, Valentin De Bortoli, and Stéphane Mallat. Wavelet score-based generative modeling. In *NeurIPS*, 2022. Multi-scale diffusion over wavelet coefficients, linear time complexity.

[11] Keyu Tian, Yi Jiang, Zehuan Yuan, Bingyue Peng, and Liwei Wang. Visual autoregressive modeling: Scalable image generation via next-scale prediction. In *Advances in Neural Information Processing Systems*, volume 37, pages 84839–84865, 2024.

[12] Enze Xie, Junsong Chen, Junyu Chen, Han Cai, Haotian Tang, Yujun Lin, Zhekai Zhang, Muyang Li, Ligeng Zhu, Yao Lu, and Song Han. Sana: Efficient high-resolution image synthesis with linear diffusion transformer, 2024.

[13] Enze Xie, Junsong Chen, Yuyang Zhao, Jincheng Yu, Ligeng Zhu, Yujun Lin, Zhekai Zhang, Muyang Li, Junyu Chen, Han Cai, et al. Sana 1.5: Efficient scaling of training-time and inference-time compute in linear diffusion transformer, 2025.

[14] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 586–595, 2018.

[15] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 6626–6637, 2017.

[16] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826, 2016.

[17] Colm Mulcahy. Image compression using the haar wavelet transform. *Spelman Science and Mathematics Journal*, 1(1):22–31, 1997.

[18] Patrick et al. Esser et al. Taming transformers for high-resolution image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12873–12883, June 2021.

[19] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.

[20] Junyu Chen, Han Cai, Junsong Chen, Enze Xie, Shang Yang, Haotian Tang, Muyang Li, Yao Lu, and Song Han. Deep compression autoencoder for efficient high-resolution diffusion models. *arXiv preprint arXiv:2410.10733*, 2024.

[21] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10684–10695, June 2022.

[22] Doyup Lee, Chiheon Kim, Saehoon Kim, Minsu Cho, and Wook-Shin Han. Autoregressive image generation using residual quantization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11523–11532, June 2022.

[23] Qihang Yu, Mark Weber, Xueqing Deng, Xiaohui Shen, Daniel Cremers, and Liang-Chieh Chen. An image is worth 32 tokens for reconstruction and generation. In *Advances in Neural Information Processing Systems*, volume 37, pages 128940–128966, 2024.

[24] Hao Chen, Yujin Han, Fangyi Chen, Xiang Li, Yidong Wang, Jindong Wang, Ze Wang, Zicheng Liu, Difan Zou, and Bhiksha Raj. Masked autoencoders are effective tokenizers for diffusion models. In *Forty-second International Conference on Machine Learning*, 2025.


**Figure 3:** Qualitative reconstructions using **FA-VAE** on ImageNet 256×256.

# A Architecture and Implementation Details

The architecture of our Frequency-Aware Variational Au- toencoder (FA-VAE) consists of two independent encoder- decoder pairs, designed to separately process low- and high-frequency components of the input data. The encoder- decoder pair architecture is inspired by VAVAE [7]. We further follow a similar hyperparameter setup to [21, 7] for implementing our latent reconstruction module. To support distributed training across multiple nodes, we scale the learning rate and global batch size to $1 \times 10^{-4}$ and 256, respectively, following the MAR setup [6]. We experiment with two $f16$ tokenizers for low and high frequency subbands: one trained without visual alignment VF loss for high frequencies and one with VF loss using DINOv2 [19] for low frequencies. Here, $f$ represents the latent spatial resolution factor and $d$ the latent dimensionality. Following [7], we set the VF loss hyperparameters to $m_1 = 0.5$, $m_2 = 0.25$, and $w_{\text{hyper}} = 0.1$. Figure 3 shows the qualitative reconstructions of FA-VAE.

# NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: See Sec.1.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [NA]

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: See Appendix A.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [NA]

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [NA]

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [NA]

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [NA]

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: See Abstract.

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [No]