



Introduction

Background: Recent methods like DreamBooth [2] can fine-tune diffusion models to only output images of a specific identity. [1] proposed to use PCA to build a latent space called Weights2Weights ($w2w$) from a set of weights of fine-tuned diffusion models. Users can manipulate, interpolate or sample weights from $w2w$ space to generate new diffusion models which that encodes novel and consistent identities.

Motivation: PCA is a *linear method* and may not capture the complex relationships between the weights. Our Insight is *VAEs can be used to construct a more expressive, informative latent space*.

Task statement: Given a set of LoRA weights from different identity-specific diffusion models fine-tuned using DreamBooth [2], we aim to learn a latent space representation of the weights using a VAE.

Dataset: 60k+ fine-tuned LoRA weights used in [1].

Evaluation Metric: Quantitative comparison with [1] in *subject inversion* task using ID scores.

Method

Construct weights manifold: [1] applies PCA on a LoRA weights dataset and models $w2w$ as a linear combination of PCA bases. In comparison, we propose to train a VAE model on the same dataset and use its latent space as $w2w++$.

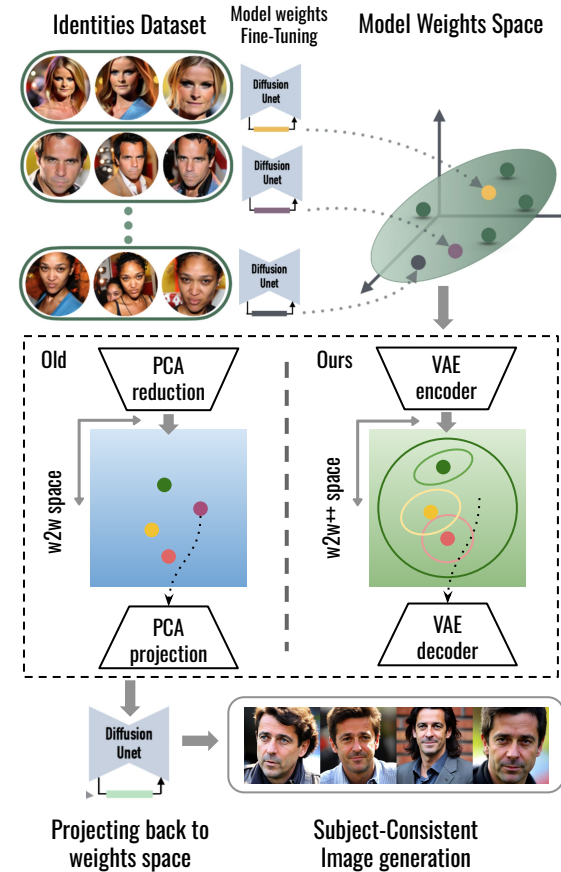
W2W++ VAE: Encodes an 1-d weight vector into a latent distribution, then decodes model weights from samples drawn from this latent distribution. We apply KL Weight Annealing [3] to stabilize the training.

Downstream tasks benefited by $w2w++$ space

- **Sampling:** Sample a latent vector from $N(0, I)$ and pass it through the decoder, yielding a new model.
- **Interpolation:** Interpolation between two latent embeddings can blend two different subjects, resulting in fancy visualizations.

- **Subject Inversion:** Given an identity image, invert it into $w2w++$ space. Motivated by [1], we fine-tune a diffusion model by only optimizing the latent vector that is passed into VAE decoder to generate the inverted model.

Pipeline



References

- [1] Amil Dravid, Yossi Gandelsman, Kuan-Chieh Wang, Rameen Abdal, Gordon Wetzstein, Alexei A. Efros, and Kfir Aberman. Interpreting the Weight Space of Customized Diffusion Models. *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.
- [2] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.
- [3] Samuel R. Bowman, Luke Vilnis, Oriol Vinyals, Andrew M. Dai, Rafal Jozefowicz, and Samy Bengio. Generating Sentences from a Continuous Space. *arXiv preprint arXiv: 1511.06349*, 2015.

Results

Reconstruction



Sampling



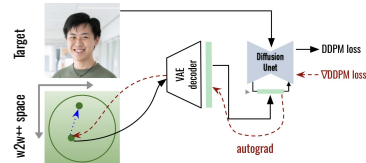
Interpolation



Baseline Inversion:



$w2w++$ Inversion:



Visualizing latent $w2w++$ space

