Finding the Global Semantic Representation in GAN through Fréchet Mean

Jaewoong Choi, Geonho Hwang, Hyunsoo Cho, Myungjoo Kang

KOREA INSTITUTE FOR ADVANCED STUDY

SEOUL NATIONAL UNIVERSITY
NUMERICAL COMPUTING & IMAGE ANALYSIS LAB

ICLR 2023

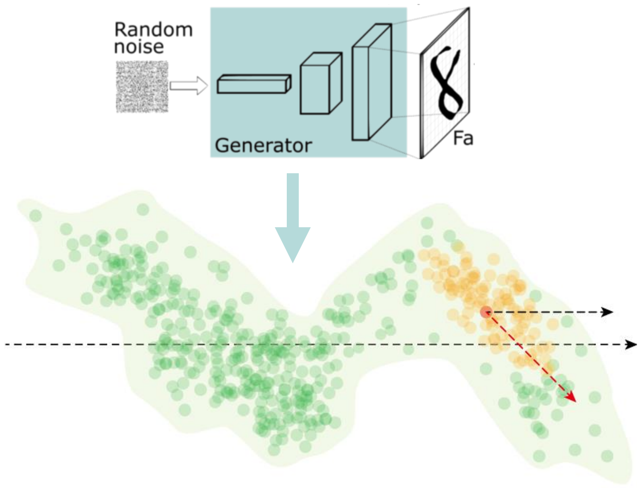
Preliminaries

Disentangled Latent Perturbation

- Consider a GAN-generator $G : \mathcal{Z} \longrightarrow \mathcal{X}$ and an image transformation T.
- The **disentangled latent perturbation, i.e., *Semantic Basis*, $v_t(z)$** on $\mathcal{Z}$ for T is defined as follows:

$$G(z + v_t(z)) = T(x) \qquad \text{where } x = G(z).$$

- $\boldsymbol{v_t(z)}$ is **global** if $v_t(z) = v_t$ for all $z \in \mathcal{Z}$, and is **local** otherwise.



Latent Perturbation [1]

Disentangled Semantic Variation [2]. Red box denotes the original image.

[1] Choi, Jaewoong, et al. "Do not escape from the manifold: Discovering the local coordinates on the latent space of GANs." *ICLR*, 2022.
[2] Härkönen, Erik, et al. "Ganspace: Discovering interpretable gan controls." *NeurIPS, 2020.*

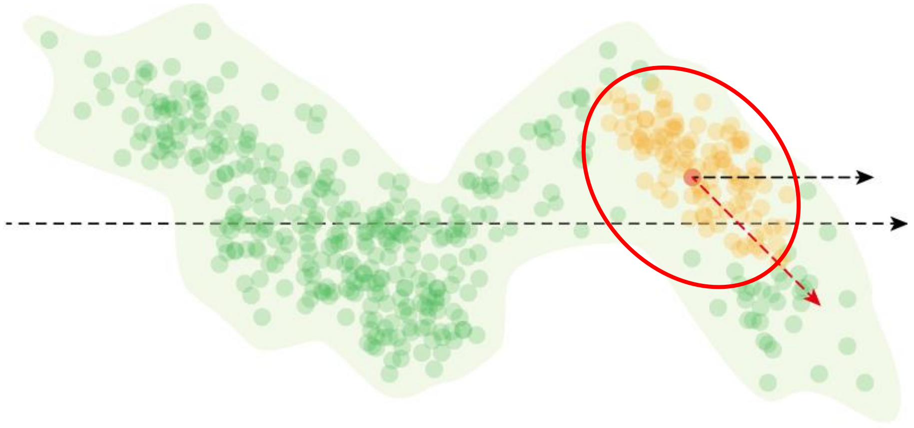
Preliminaries

Understanding Semantics via Latent Manifold

- Divide a pretrained GAN into two subnetworks based on the intermediate latent space $\tilde{W}$

$$\underbrace{\mathcal{Z} = \mathbb{R}^n}_{\text{Input Noise Space}} \xrightarrow{f\,:\,\text{Mapping network}} \underbrace{\boxed{\tilde{\mathcal{W}}}}_{\text{Target Latent Space}} \longrightarrow \underbrace{\mathcal{X}}_{\text{Image Space}},$$

- [1] suggested **Local Semantic Basis** by estimating $\mathcal{W} = f(\mathcal{Z})$ with the **lower-dimensional local approximation** motivated by manifold hypothesis.



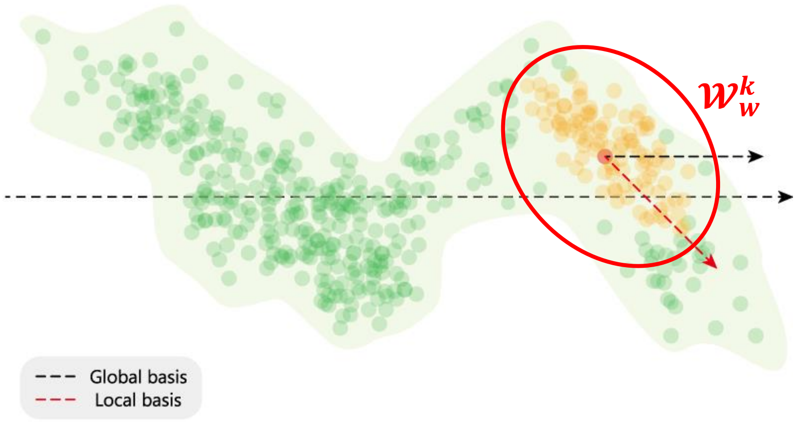
Black-box latent space $\mathcal{W} = f(\mathcal{Z})$ [1]

[1] Choi, Jaewoong, et al. "Do not escape from the manifold: Discovering the local coordinates on the latent space of GANs." *ICLR*, 2022.

Preliminaries

Finding Local (Semantic) Basis

- Discover the local approximation through the **Singular Value Decomposition** of df_z.



Black-box latent space $\mathcal{W} = f(\mathcal{Z})$

SVD of Subnetwork Jacobian

$$df_{\mathbf{z}}(\mathbf{u}_i^{\mathbf{z}}) = \sigma_i^{\mathbf{z}} \cdot \mathbf{v}_i^{\mathbf{w}} \text{ for } \forall i,$$

$$\text{Local Basis}(\mathbf{w} = f(\mathbf{z})) = \{\mathbf{v}_i^{\mathbf{w}}\}_{1 \leq i \leq n}.$$

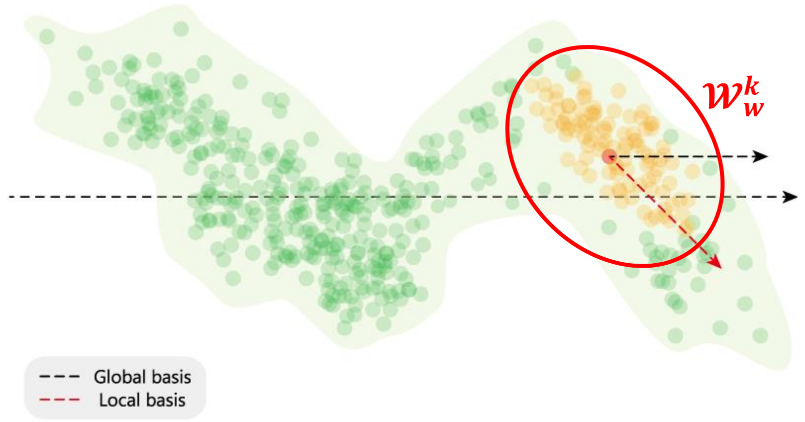
$$\Rightarrow \mathcal{W}_{\mathbf{w}}^k = \left\{ f\left(\mathbf{z} + \sum_i t_i \cdot \mathbf{u}_i^{\mathbf{z}}\right) \mid t_i \in (-\epsilon_i, \epsilon_i), \text{ for } 1 \leq i \leq k \right\}$$

[1] Choi, Jaewoong, et al. "Do not escape from the manifold: Discovering the local coordinates on the latent space of GANs." *ICLR*, 2022.

Preliminaries

Finding Local (Semantic) Basis

- Discover the local approximation through the **Singular Value Decomposition** of df_z.



Black-box latent space $\mathcal{W} = f(\mathcal{Z})$

$$df_{\mathbf{z}}(\mathbf{u}_i^{\mathbf{z}}) = \sigma_i^{\mathbf{z}} \cdot \mathbf{v}_i^{\mathbf{w}} \text{ for } \forall i,$$

$$\text{Local Basis}(\mathbf{w} = f(\mathbf{z})) = \{\mathbf{v}_i^{\mathbf{w}}\}_{1 \leq i \leq n}.$$

Semantic Basis

$$\Rightarrow \mathcal{W}_{\mathbf{w}}^k = \left\{ f\left(\mathbf{z} + \sum_i t_i \cdot \mathbf{u}_i^{\mathbf{z}}\right) \mid t_i \in (-\epsilon_i, \epsilon_i), \text{ for } 1 \leq i \leq k \right\}$$

- The top-k Local Basis spans the tangent space of k-dim approximating manifold:

$$T_{\mathbf{w}} \mathcal{W}_{\mathbf{w}}^k = \text{span}\{\mathbf{v}_i^{\mathbf{w}} : 1 \leq i \leq k\}.$$

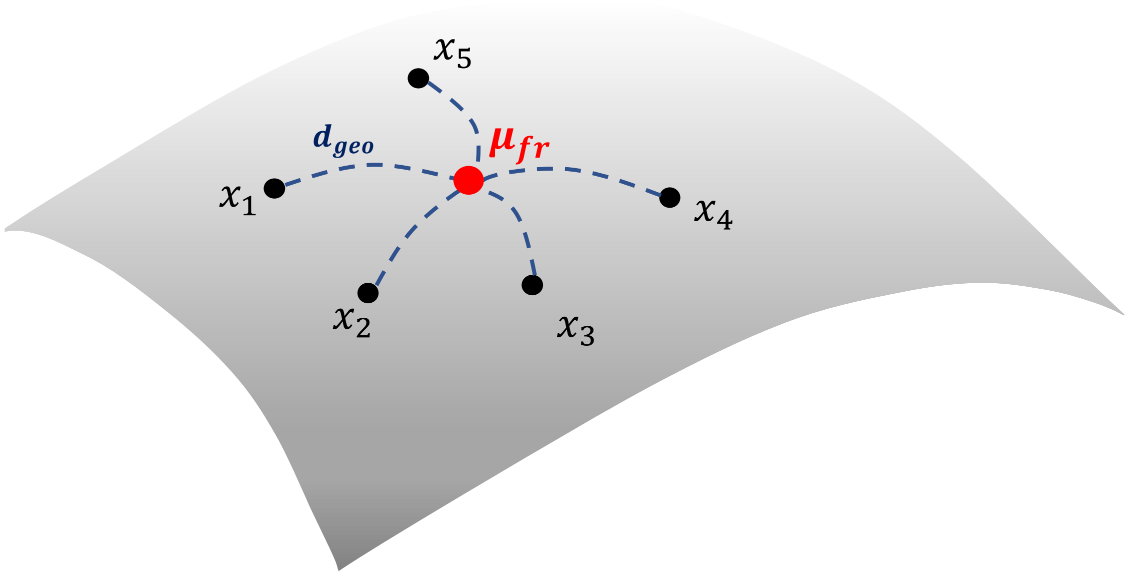
Semantic Subspace

[1] Choi, Jaewoong, et al. "Do not escape from the manifold: Discovering the local coordinates on the latent space of GANs." *ICLR*, 2022.

Fréchet Basis

Finding Global (Semantic) Basis

- **Discover the Global Semantic Basis by averaging the Local Semantic Basis**
  - **Fréchet Mean $\boldsymbol{\mu_{fr}}$** on the metric space X
    - For $x_1, x_2, \ldots, x_n \in X$, $\mu_{fr} = \underset{\mu \in X}{\arg\min} \sum_{1 \le i \le n} d(\mu, x_i)^2$.



Fréchet Mean $\boldsymbol{\mu_{fr}}$ on Riemannian Manifold

Fréchet Basis

Finding Global Semantic Subspace

- Discover **Global Semantic Subspace $\mathcal{S}_S$** by Fréchet Mean on **Grassmannian manifold $Gr(d_{\mathcal{W}}, \mathbb{R}^{d_{\mathcal{W}}})$**

  I. **Sample the intrinsic tangent space $T_{w_i}\mathcal{W}_{w_i}^{d_{w_i}}$ at each $w_i \in \mathcal{W}$**

  II. **Embed each tangent space to $Gr(d_{\mathcal{W}}, \mathbb{R}^{d_{\mathcal{W}}})$ by matching its dimension to $d_{\mathcal{W}}$**

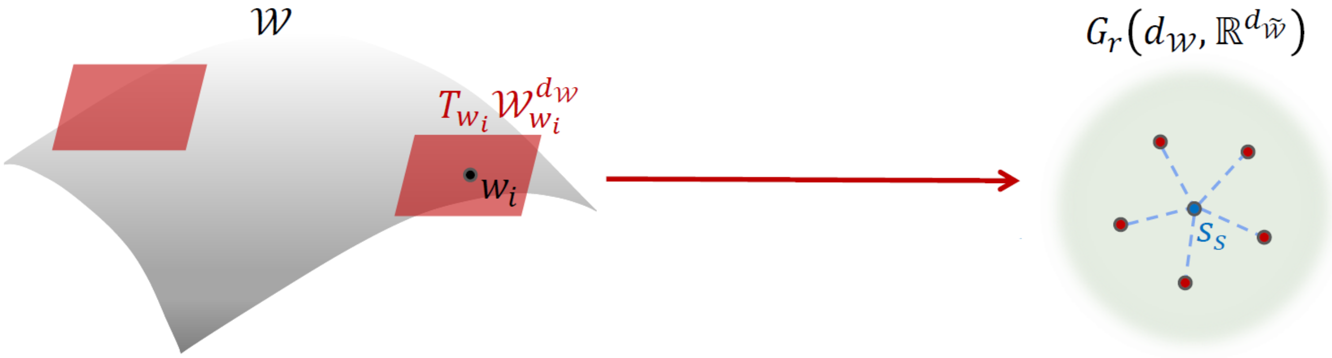
  III. **Solve the following optimization problem:**

$$\mathcal{S}_s = \underset{\mu \in Gr(d_{\mathcal{W}}, \mathbb{R}^{d_{\tilde{\mathcal{W}}}})}{\arg\min} \sum_{1 \le i \le n} d_{geo}\left(\mu,\, T_{\mathbf{w}_i}\mathcal{W}_{\mathbf{w}_i}^{d_{\mathcal{W}}}\right)^2$$

Geodesic distance

$$d_{\text{geo}}(W, W') = \left(\sum_{i=1}^{k} \theta_i^2\right)^{1/2}$$

$$\theta_i = \cos^{-1}(\sigma_i(M_W^\top M_{W'}))$$



Finding Fréchet Basis

- Choose the optimal **Global Semantic Basis of $\mathcal{S}_S$** by Fréchet Mean on **Special Orthogonal Group**

- **Why Special Orthogonal Group $\boldsymbol{SO(d_W)}$?**
  - Let the columns of $M_S, M'_S \in \mathbb{R}^{d_{\tilde{W}} \times d_W}$ be the **two distinct orthonormal basis** of the Semantic Subspace $\boldsymbol{\mathcal{S}_S}$
  - Then, there exists an orthogonal matrix O s.t. $\boldsymbol{M'_S = M_S O}$.

  ➡ **Finding the optimal basis of $\mathcal{S}_S$ is equivalent to finding the orthogonal matrix $\boldsymbol{O}$.**

  - **Orthogonal Group is disconnected.**
  - Our task is independent of the flipping ((-1)-multiplication) of each basis component.

  ➡ **From $\boldsymbol{O(d_W)}$ to $\boldsymbol{SO(d_W)}$**

Finding Fréchet Basis

- The proposed scheme for Basis Refinement is as follows:
  I. **Project each local semantic basis to the $\mathcal{S}_S$**
  II. **Project these projected local semantic basis to $SO(d_{\mathcal{W}})$**

$$P_{so}(X) = U \ \mathrm{diag}\left(1, 1, \ldots, 1, \det\left(P_o(X)\right)\right) V^\top$$

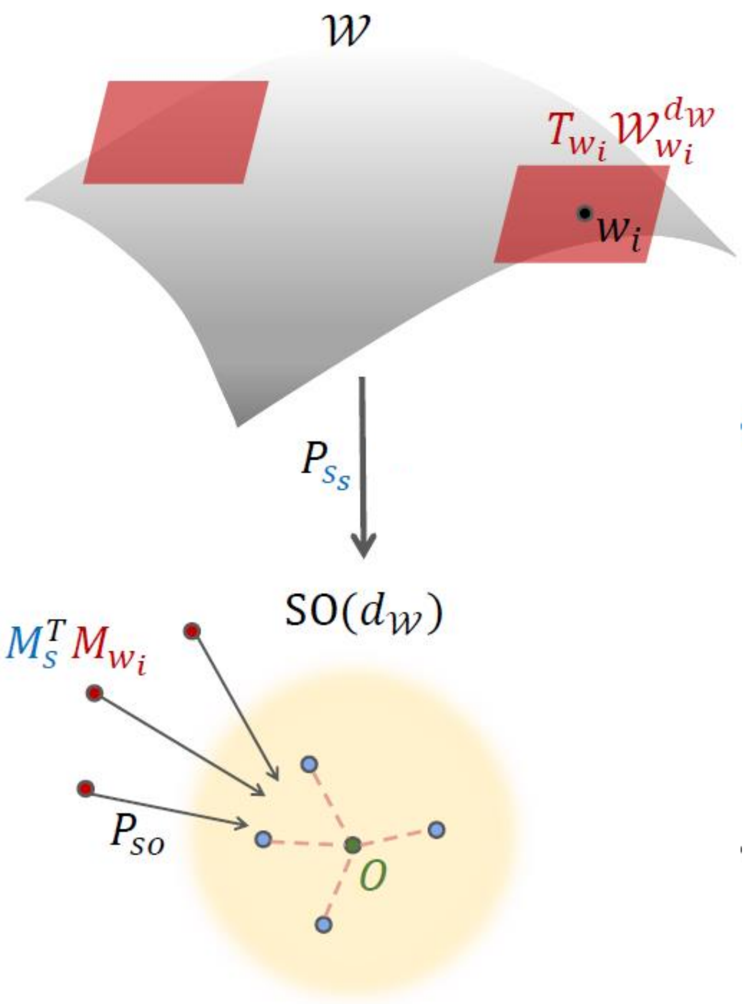
$$\text{where} \ \ P_o(X) = UV^\top.$$

  III. **Find the Fréchet mean O in $SO(d_{\mathcal{W}})$ and**

Embed back to the ambient space.

$$\mathcal{B}_s = M_{\mathcal{S}} O \quad \text{where} \ \ O = \underset{\mu \in SO(d_{\mathcal{W}})}{\arg\min} \sum_{1 \le i \le n} d\left(\mu, \ P_{so}\left(M_{\mathcal{S}}^\top M_{\mathbf{w}_i}\right)\right)^2 .$$

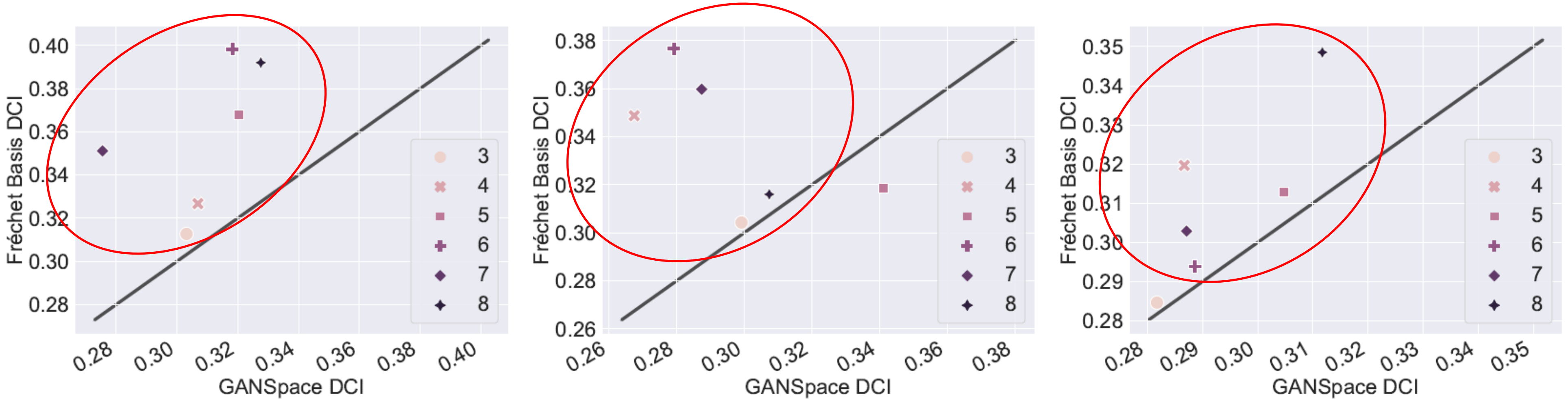
Fréchet Basis



Experiments

Quantitative Comparison of Semantic Factorization.

- Compare **DCI scores (↑) of Fréchet Basis and GANSpace** representations of the same latent space
  - DCI is a supervised disentanglement metric that assesses the axis-wise alignment of semantics.
- **In most latent spaces, the latent space achieves a higher DCI score in the Fréchet Basis representation.**



(a) StyleGAN1-FFHQ (b) StyleGAN2-e-FFHQ (c) StyleGAN2-FFHQ

Figure 4: **Quantitative Comparison of Semantic Factorization.** DCI score (↑) is a supervised disentanglement metric that evaluates the axis-wise alignment of semantics. Fréchet basis outperforms GANSpace when a point is located *above* the black line. (The black line illustrates $y = x$.)

Experiments

Quantitative Comparison of Image Fidelity

- Compare **FID scores (↓) of Fréchet Basis and GANSpace** under the latent perturbation

  - A large FID score indicates that the distribution of the traversed image is far from that of the natural training images, which implies the image collapse.

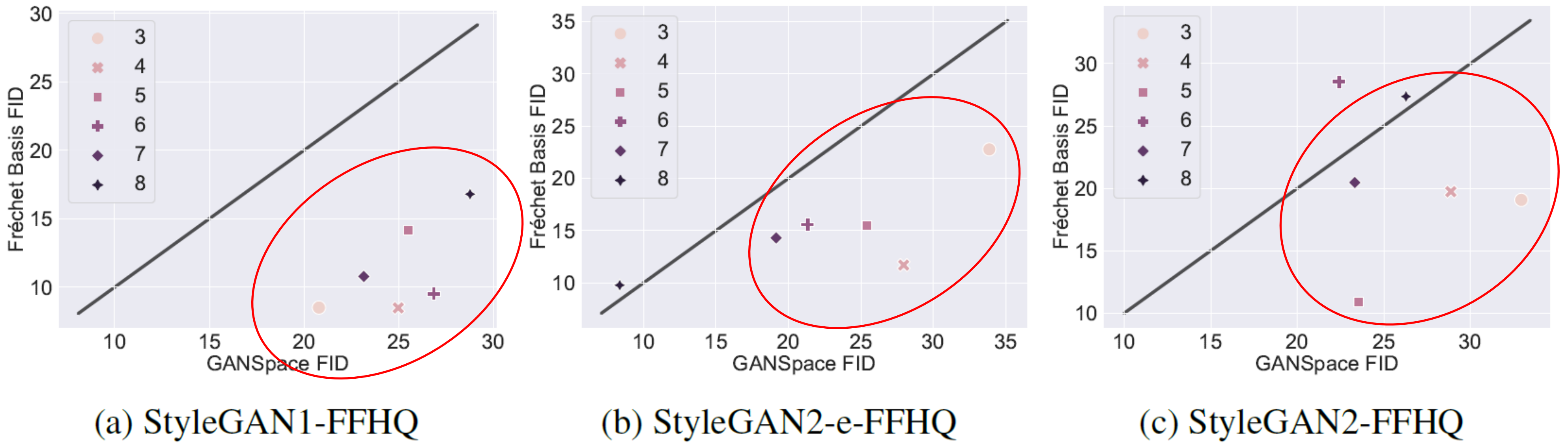
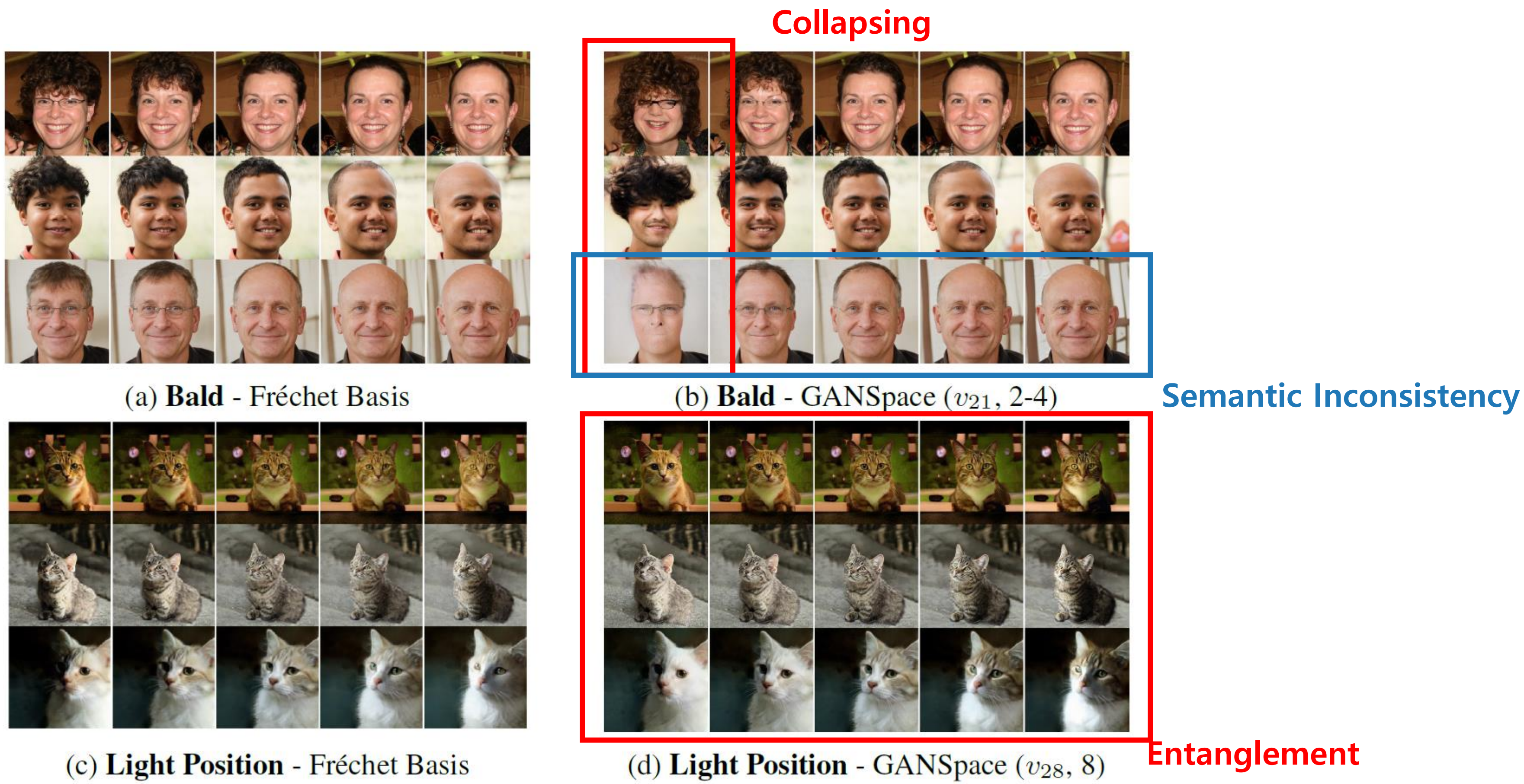


Figure 5: **Quantitative Comparison of Robustness.** The image fidelity under the latent traversal is evaluated by FID (↓). Fréchet basis performs better when a point is placed *below* the black line.

Fréchet Basis as Global Semantic Perturbation

- Compare Fréchet Basis with GANSpace of the highest cosine similarity.



(a) **Bald** - Fréchet Basis

(b) **Bald** - GANSpace (v_{21}, 2-4)

(c) **Light Position** - Fréchet Basis

(d) **Light Position** - GANSpace (v_{28}, 8)

Ours (Fréchet Basis)

GANSpace

Thank you!