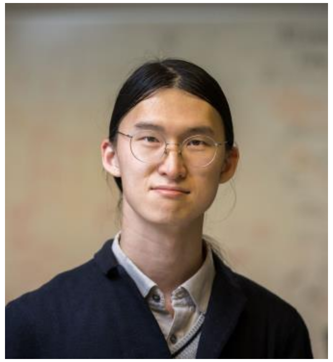
The Geometry of Deep Generative Models and Its Applications

Binxu Wang, Carlos R. Ponce

ICLR, 2021

Washington University in St.Louis



Deep Generative Models

Generator map:

$$G: \mathbb{R}^d \to \mathcal{I}, z \mapsto x.\ \mathcal{I} := \mathbb{R}^{3 \times H \times W}$$

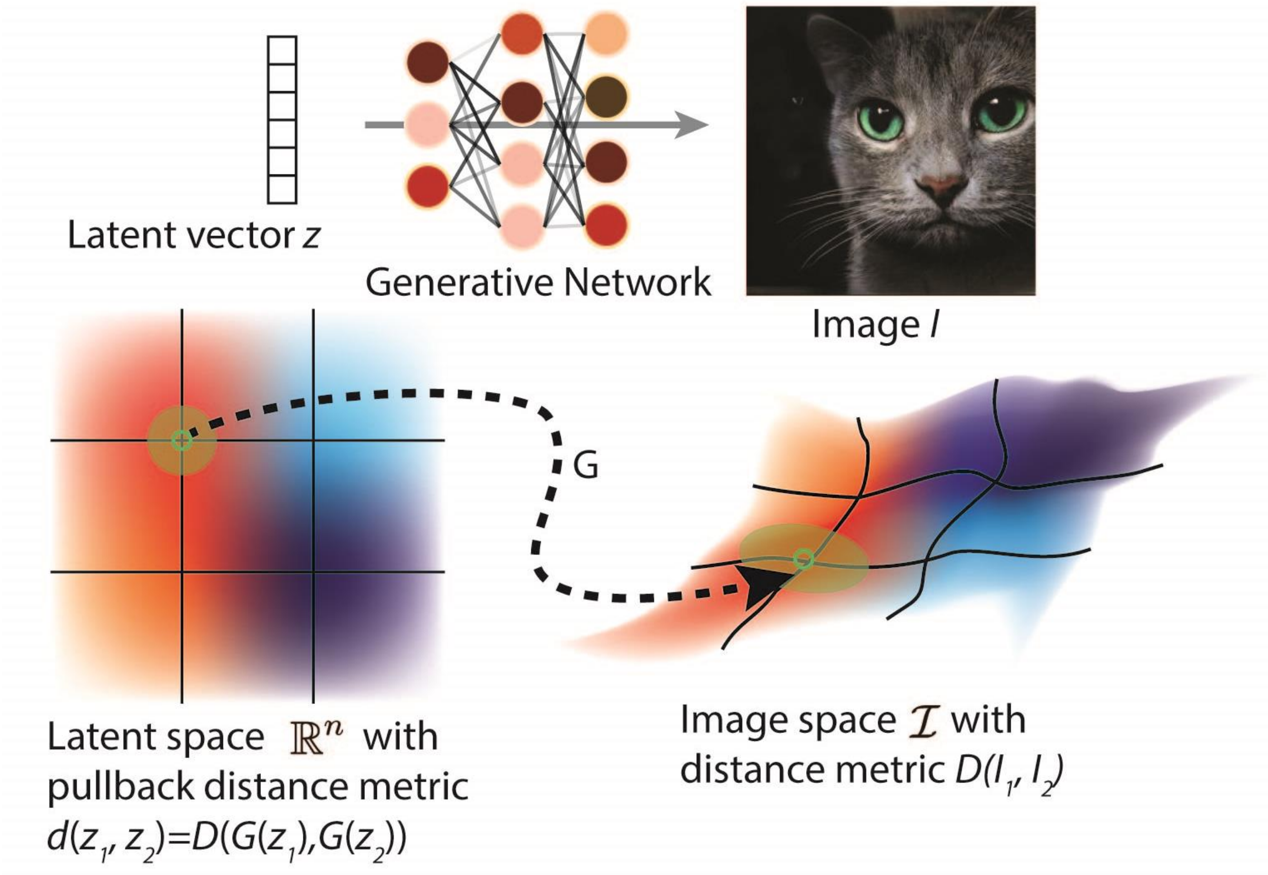
The structure of the latent (input) space requires a clearer understanding.

BigGAN 2018

StyleGAN2 2020

Formulation

- Deep Generative Models parametrize a manifold in the space of samples (e.g. images)
  - $G: \mathbb{R}^d \to \mathcal{I}, z \mapsto I.\ \mathcal{I} \coloneqq \mathbb{R}^{3\times H\times W}$
- We define the Riemannian geometry of the manifold by pulling back distance in image space.
  - *Image space metric,* $D: \mathcal{I} \times \mathcal{I} \to \mathbb{R}_+$
  - *Latent space metric,* $d: \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+, d(z_1, z_2) \coloneqq D(G(z_1), G(z_2))$



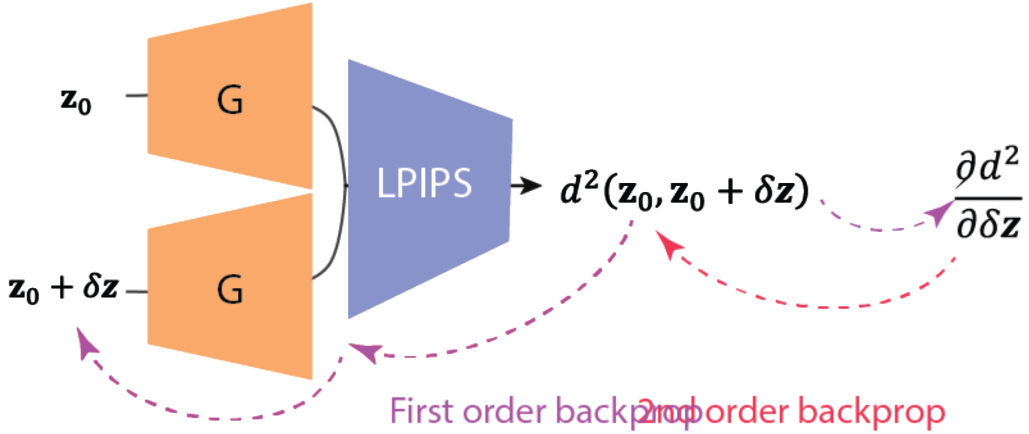
Compute the Riemannian Metric Tensor

$$H(\mathbf{z}_0) = \frac{1}{2}\frac{\partial^2 d^2(\mathbf{z}_0, \mathbf{z})}{\partial \mathbf{z}^2}\Big|_{\mathbf{z}=\mathbf{z}_0}$$

$$d^2(\mathbf{z}_0, \mathbf{z}_0 + \delta\mathbf{z}) \approx \delta\mathbf{z}^T H(\mathbf{z}_0)\delta\mathbf{z}$$

- It encodes a local notion of distance

- Numerical Method
  - 2nd order auto-differentiation
  - Applying Lanczos iteration to Hessian vector product (HVP) operator, to compute top eigenpairs.



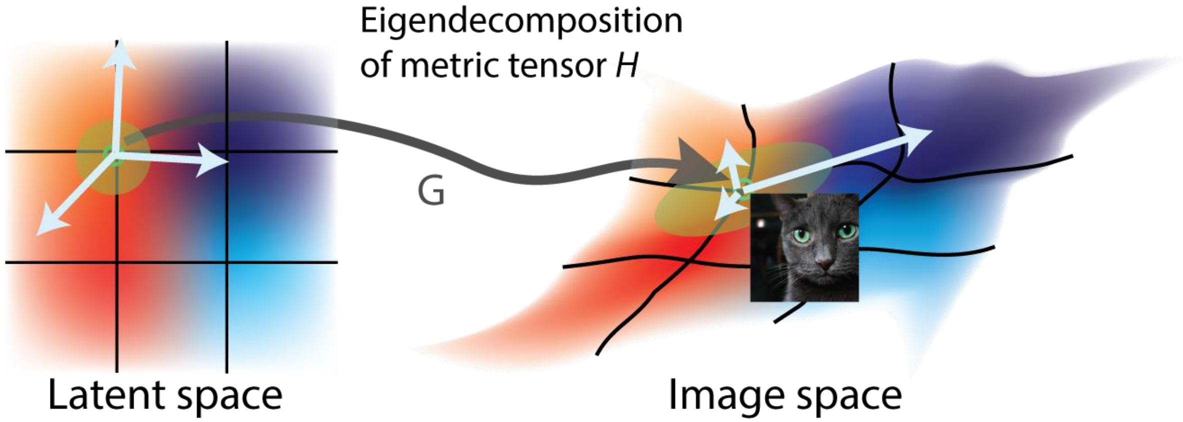
Code available at: https://github.com/Animadversio/GAN-Geometry

Lanczos Iteration on HVP

- Define Hessian vector product operator $HVP: v \mapsto Hv$
  - Forward HVP: Finite differencing to compute directional derivative of gradient
$$H_{\boldsymbol{z_0}}\boldsymbol{v} = \boldsymbol{v} \cdot \partial_{\boldsymbol{z}} g(\boldsymbol{z}) \approx \frac{g(\boldsymbol{z}_0 + \epsilon\boldsymbol{v}) - g(\boldsymbol{z}_0 - \epsilon\boldsymbol{v})}{2\epsilon\|\boldsymbol{v}\|}$$
  - Backward HVP: Use Jacobian vector product in auto-diff packages.
$$H_{\boldsymbol{z}_0}\boldsymbol{v} = \partial_{\boldsymbol{z}}(\boldsymbol{v}^T g(\boldsymbol{z}))$$

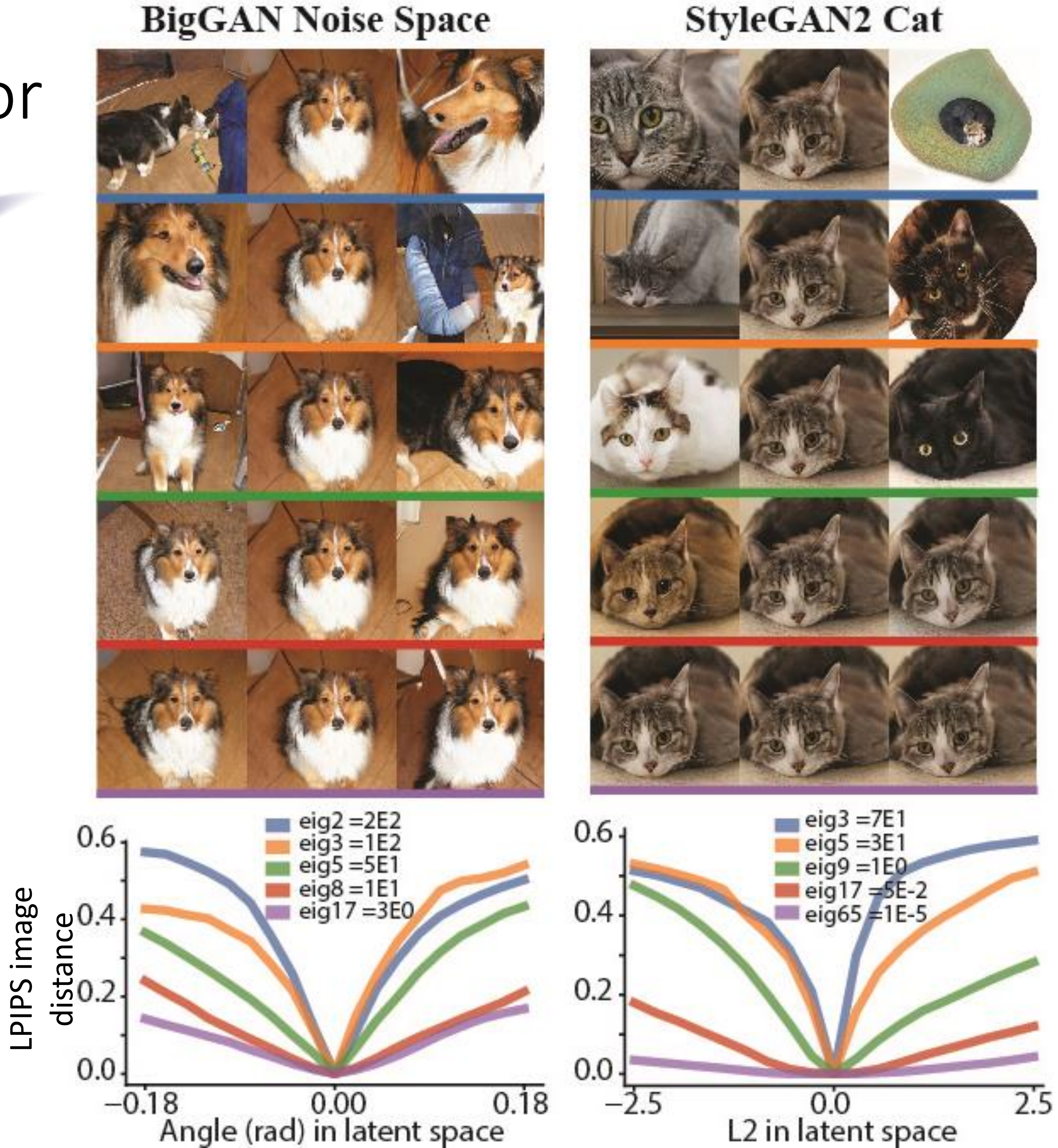
- Then call the Lanczos algorithm in ARPACK wrapped by scipy.

Eigen structure of Metric Tensor

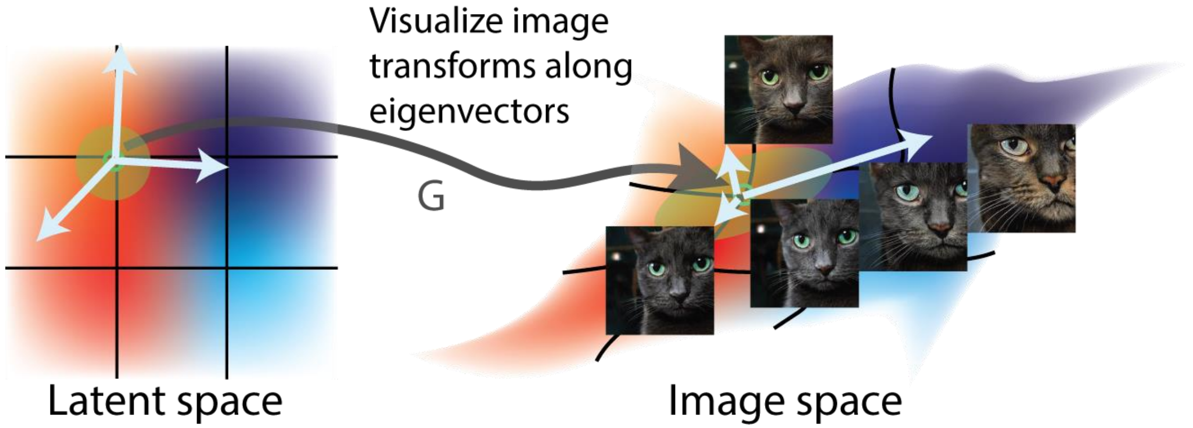


$$H = U\Lambda U^T = \sum_i \lambda_i u_i u_i^T$$

- Eigenvalues λ_i, rate of image change along u_i
- Observation
  - Rate of change varies dramatically along the spectra (4-10 orders of magnitude)
  - Different parts of spectrum tend to encode different image transforms.



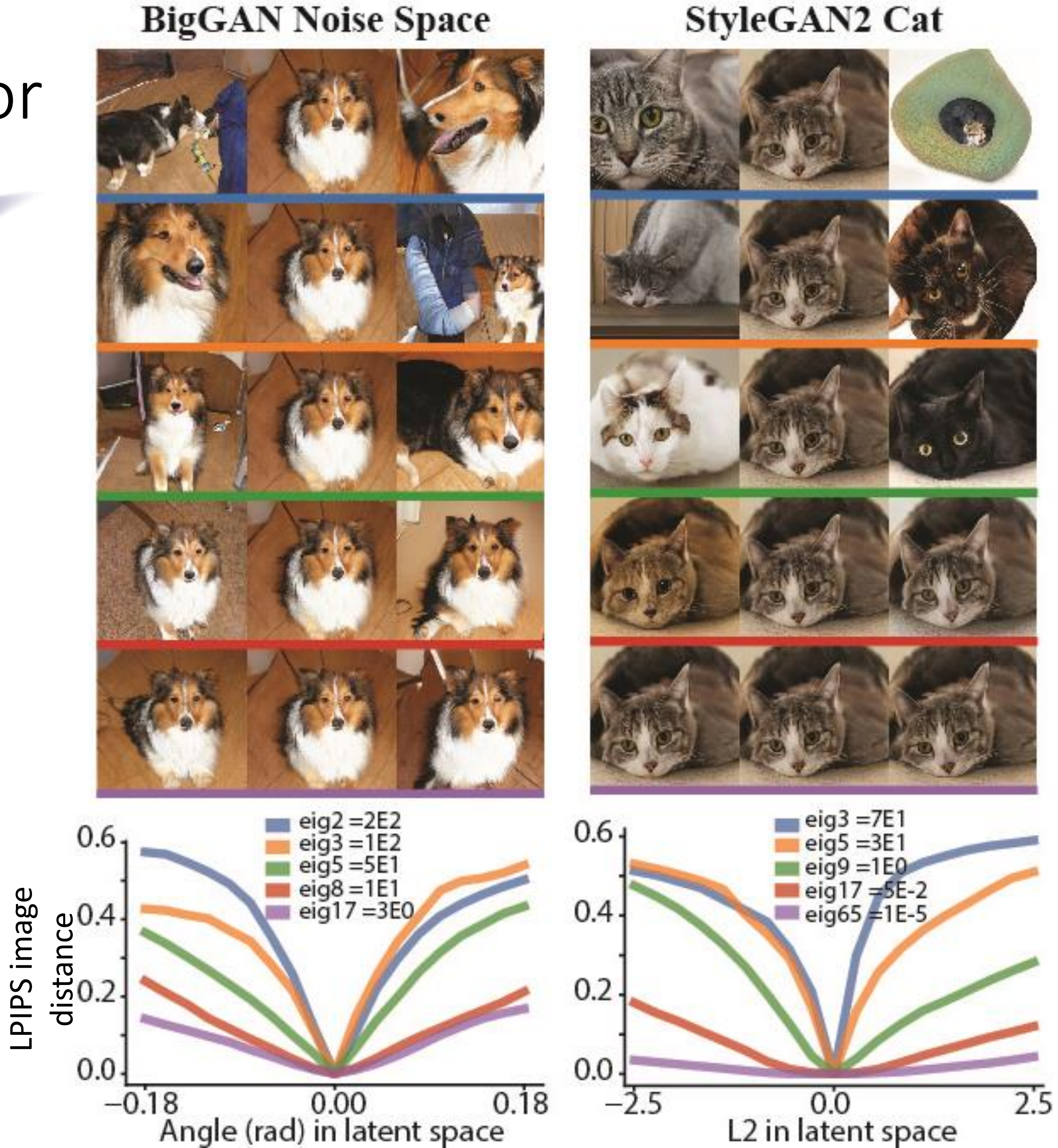
Eigen structure of Metric Tensor



$$H = U\Lambda U^T = \sum_i \lambda_i u_i u_i^T$$

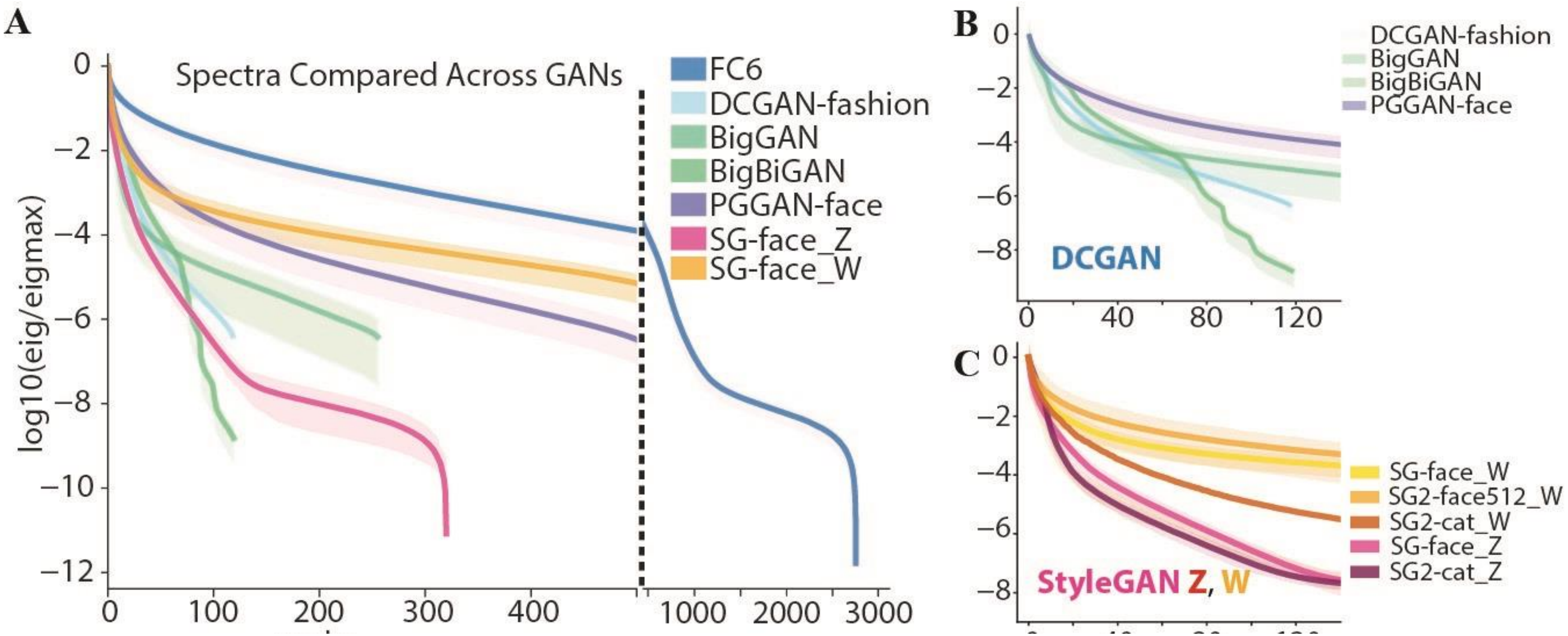
- Eigenvalues λ_i, rate of image change along u_i

- Observation
  - Rate of change varies dramatically along the spectra (4-10 orders of magnitude)
  - Different parts of spectrum tend to encode different image transforms.



Anisotropy in Most GANs

- Eigenvalues span orders of magnitude (4-10)
- Weight shuffled GANs are less anisotropic.

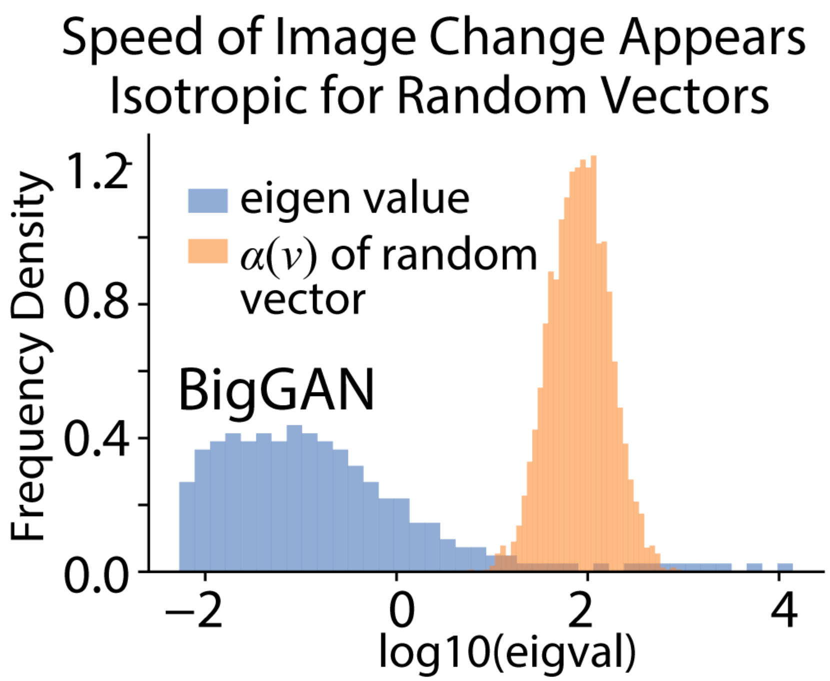


"Illusion of Isotropy" with Random Direction

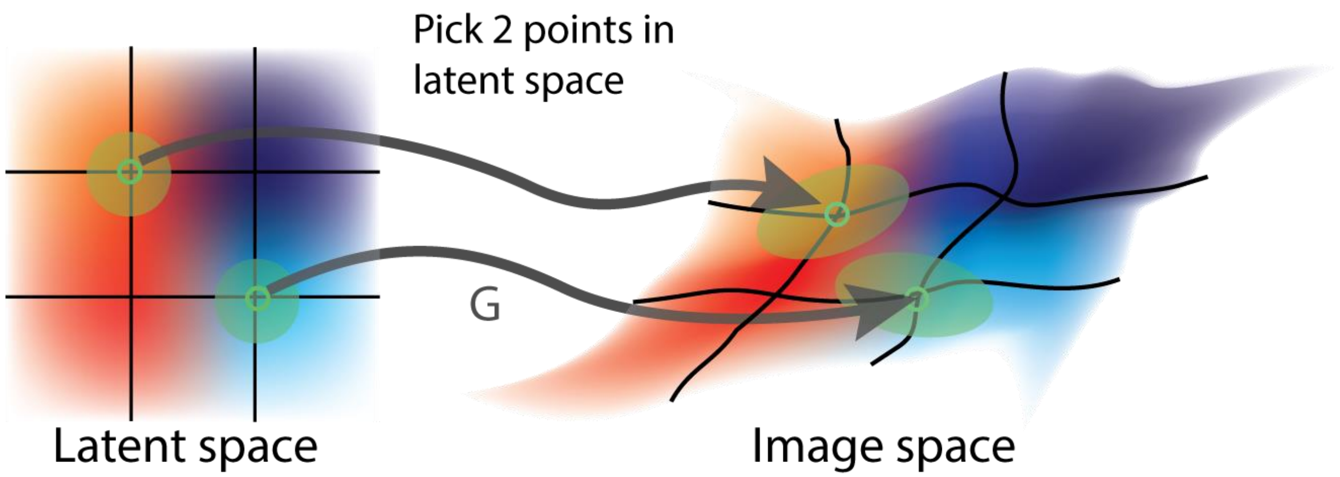
- Interpolating at random directions doesn't feel as anisotropic.
- We prove that rate of image change along random directions is much smaller than the total span of the eigenvalues.

$$Var[\alpha(v)] = \frac{2}{n+2} Var[\lambda]$$

$$v \sim \mathcal{N}(0, I_n), \alpha(v) = \frac{v^T H v}{v^T v}$$



Homogeneity: Metric Tensors are consistent across space

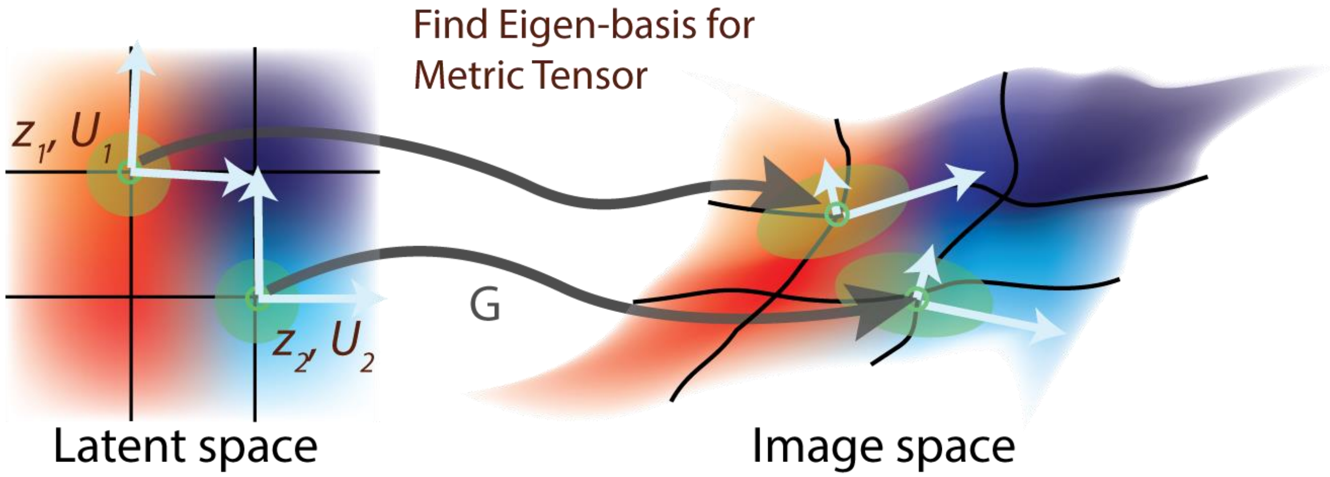


- Consistency Measure:

$$H_1 = U_1 \Lambda U_1^T, H_2 = U_2 \Lambda_2 U_2^T$$
$$\Lambda_{12} = diag(U_1^T H_2 U_1), \Lambda_2 = diag(U_2^T H_2 U_2)$$
$$C = corr(\log \Lambda_{12}, \log \Lambda_2)$$

- Results:
  - Metric tensors are similar across space.
  - Hessian structure is “semi-global” in the latent space.
  - Homogeneity observed in all the GANs.

Homogeneity: Metric Tensors are consistent across space

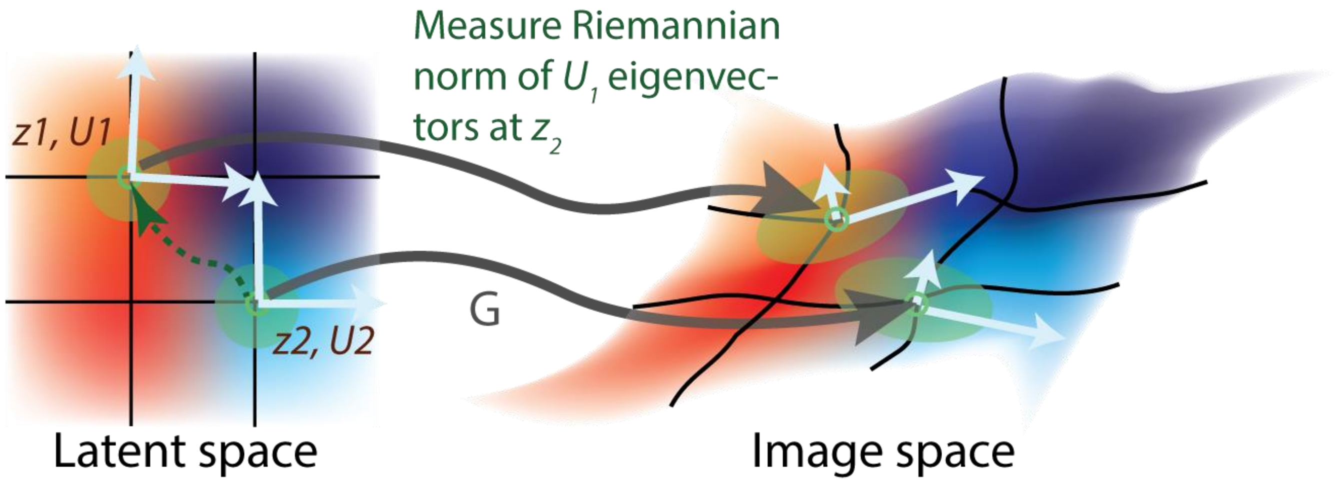


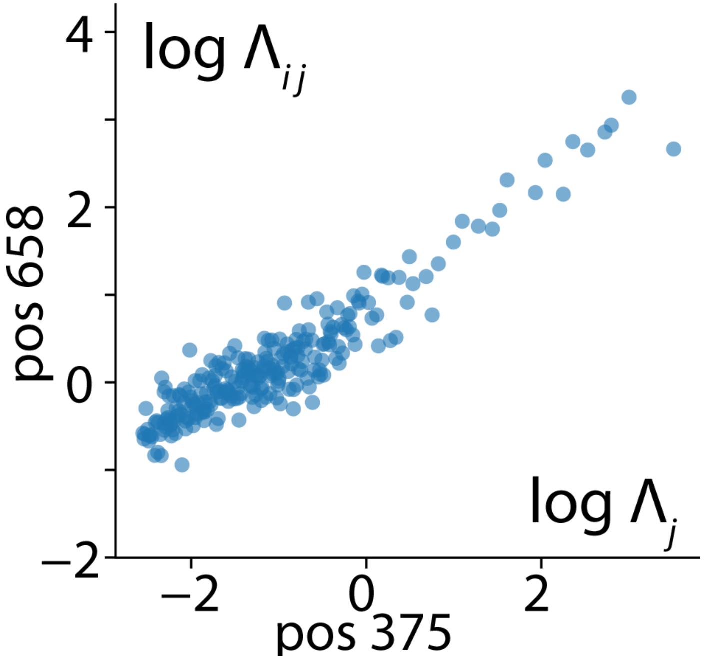
- Consistency Measure:

$$H_1 = U_1 \Lambda U_1^T, H_2 = U_2 \Lambda_2 U_2^T$$
$$\Lambda_{12} = diag(U_1^T H_2 U_1), \Lambda_2 = diag(U_2^T H_2 U_2)$$
$$C = corr(\log \Lambda_{12}, \log \Lambda_2)$$

- Results:
  - Metric tensors are similar across space.
  - Hessian structure is “semi-global” in the latent space.
  - Homogeneity observed in all the GANs.

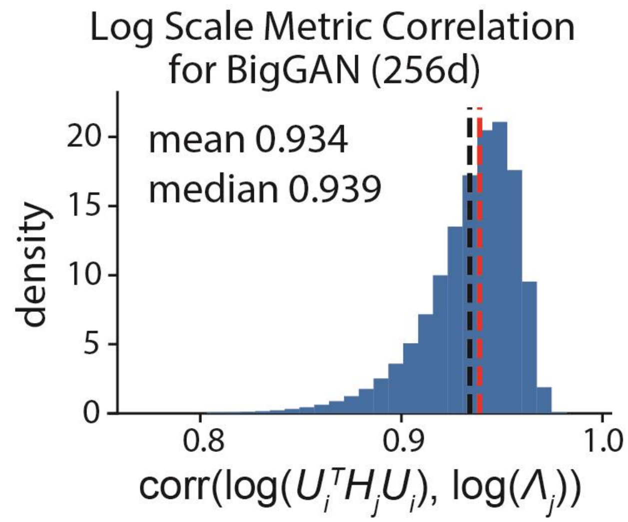
Homogeneity: Metric Tensors are consistent across space





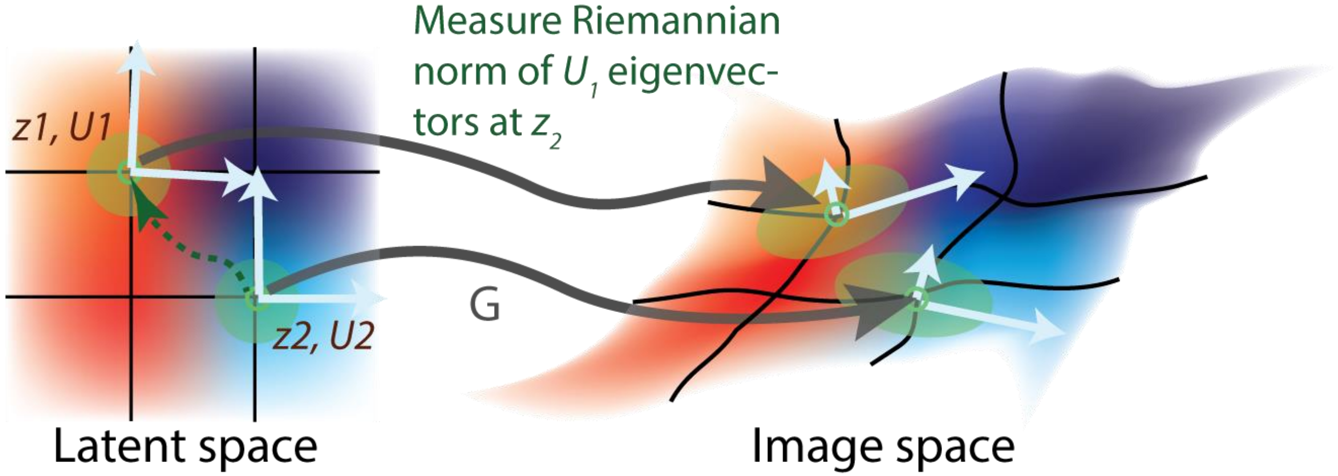
- Consistency Measure:

$$H_1 = U_1 \Lambda U_1^T, H_2 = U_2 \Lambda_2 U_2^T$$
$$\Lambda_{12} = diag(U_1^T H_2 U_1), \Lambda_2 = diag(U_2^T H_2 U_2)$$
$$C = corr(\log \Lambda_{12}, \log \Lambda_2)$$



- Results:
  - Metric tensors are similar across space.
  - Hessian structure is “semi-global” in the latent space.
  - Homogeneity observed in all the GANs.

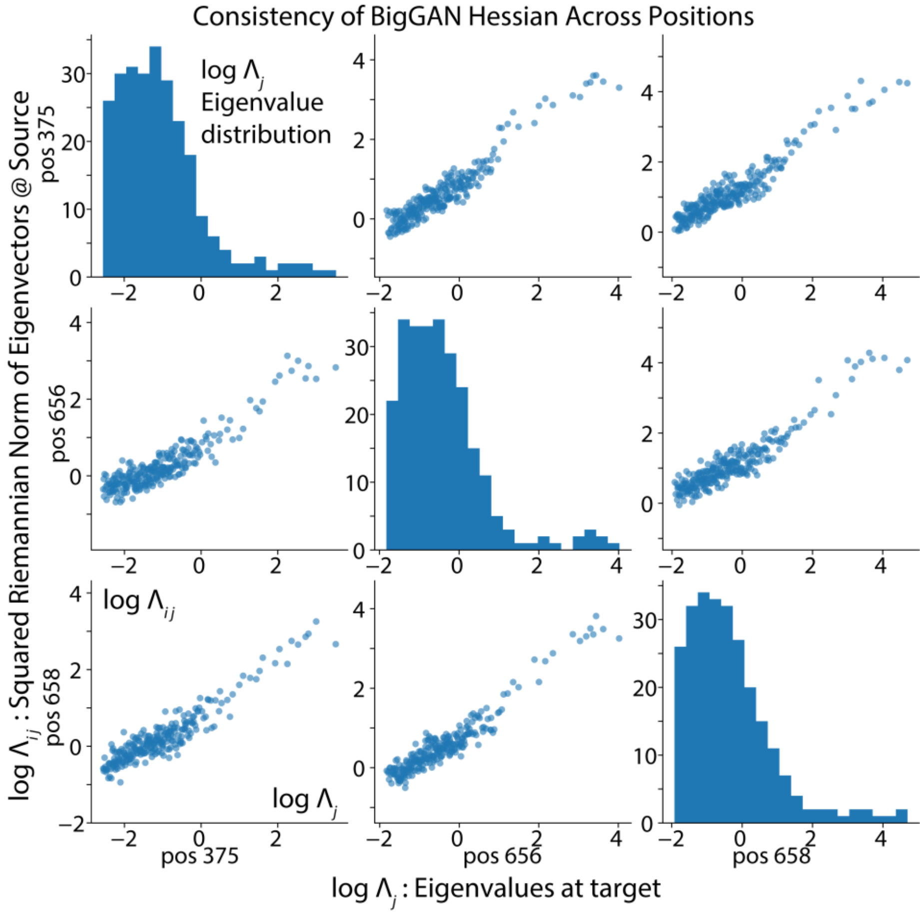
Homogeneity: Metric Tensors are consistent across space

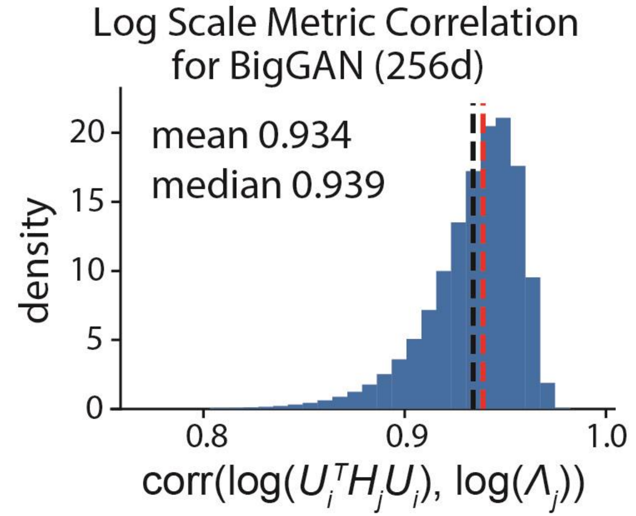


- Consistency Measure:

$$H_1 = U_1 \Lambda U_1^T, H_2 = U_2 \Lambda_2 U_2^T$$
$$\Lambda_{12} = diag(U_1^T H_2 U_1), \Lambda_2 = diag(U_2^T H_2 U_2)$$
$$C = corr(\log \Lambda_{12}, \log \Lambda_2)$$

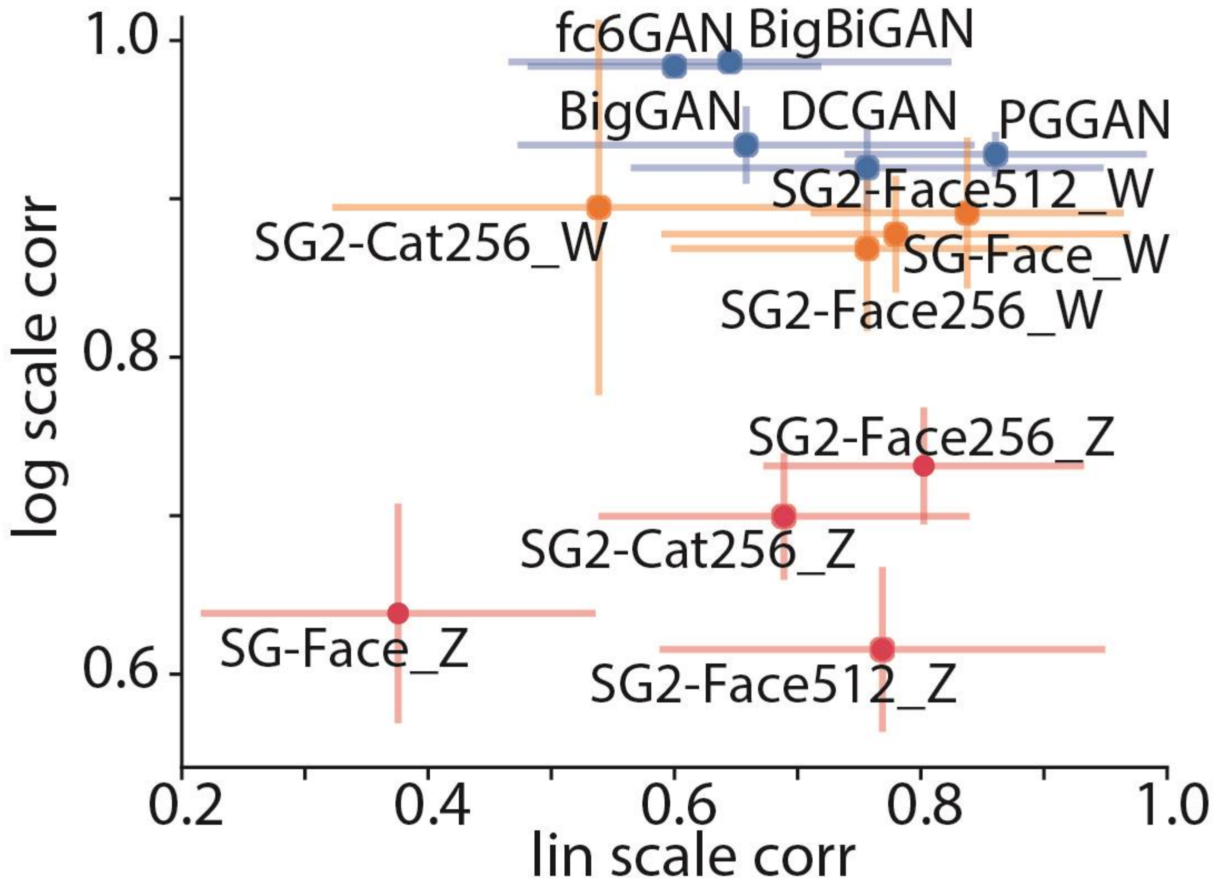
- Results:
  - Metric tensors are similar across space.
  - Hessian structure is “semi-global” in the latent space.
  - Homogeneity observed in all the GANs.





Homogeneity: Most GANs Exhibit Homogeneous geometry

- This “global” Hessian structure is observed in the latent space of many Generators.
- This Hessian consistency is disrupted in weight shuffled Generators.



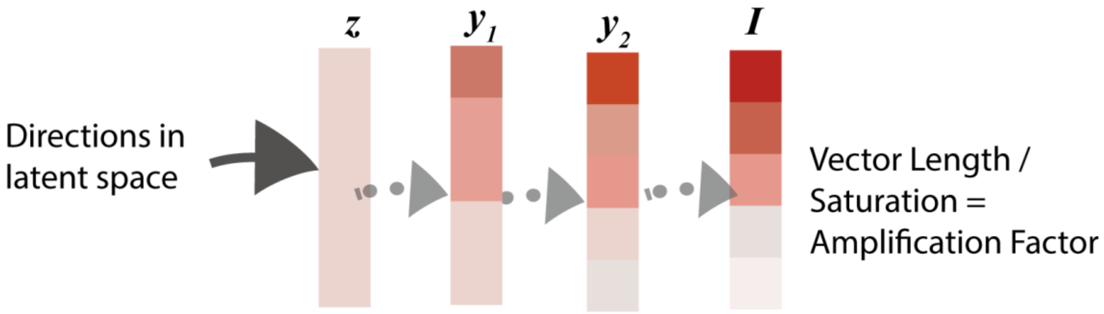
Alignment through layers

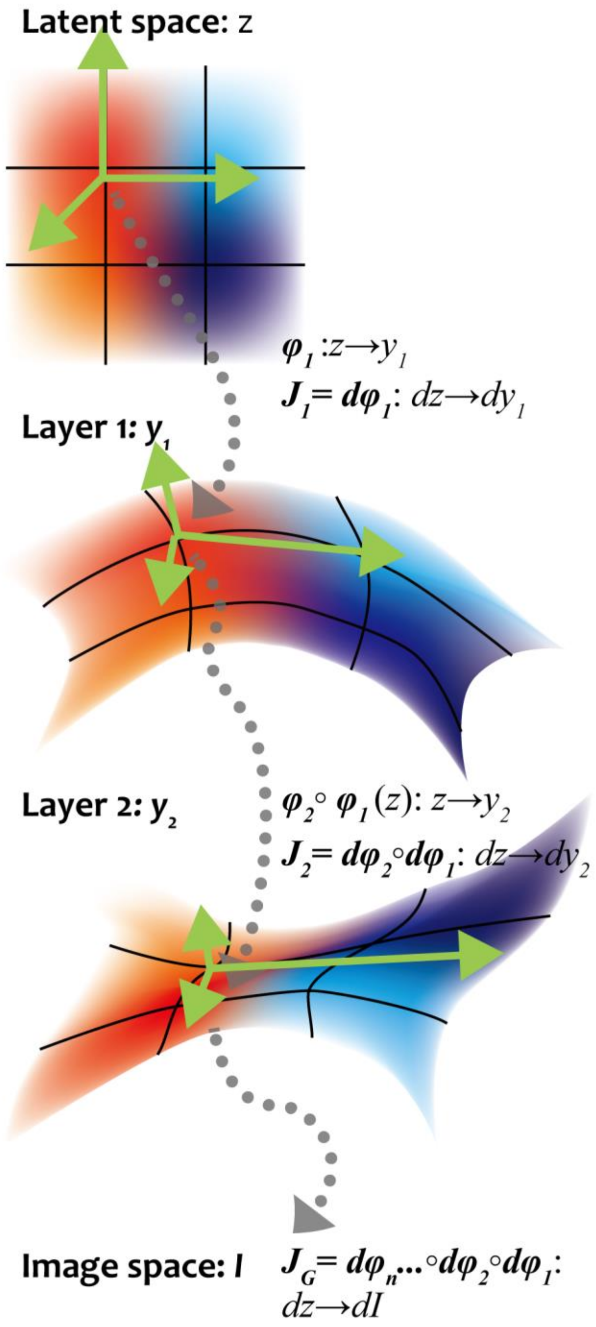
- Riemannian metric tensor can be computed for any representation space, induced by L2 distance in that space.

$$\boldsymbol{y} = \phi(\boldsymbol{z})$$

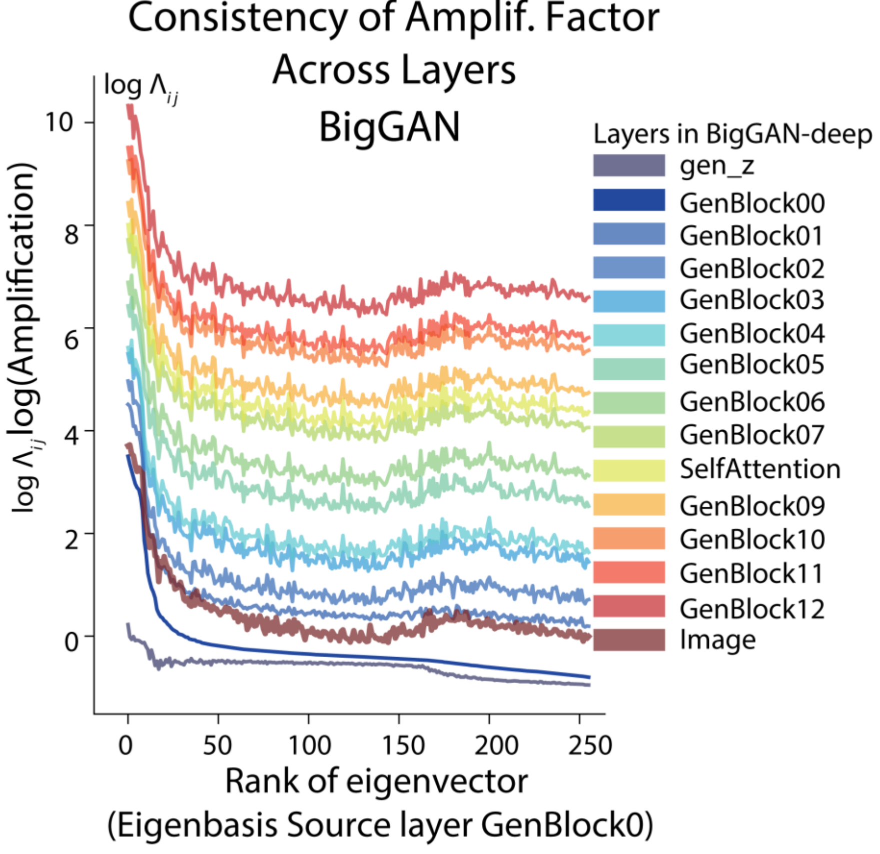
$$H(z_0) = \partial_z^2 \|\phi(\boldsymbol{z}) - \phi(\boldsymbol{z}_0)\|_2^2 = J^T J, \qquad J = \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{z}}\Big|_{\boldsymbol{z}_0}$$

- We could compute Riemannian metric tensor at the same vector $\boldsymbol{z}_0$ at different layer.
- Alignment measures the amplification effect of different layers.

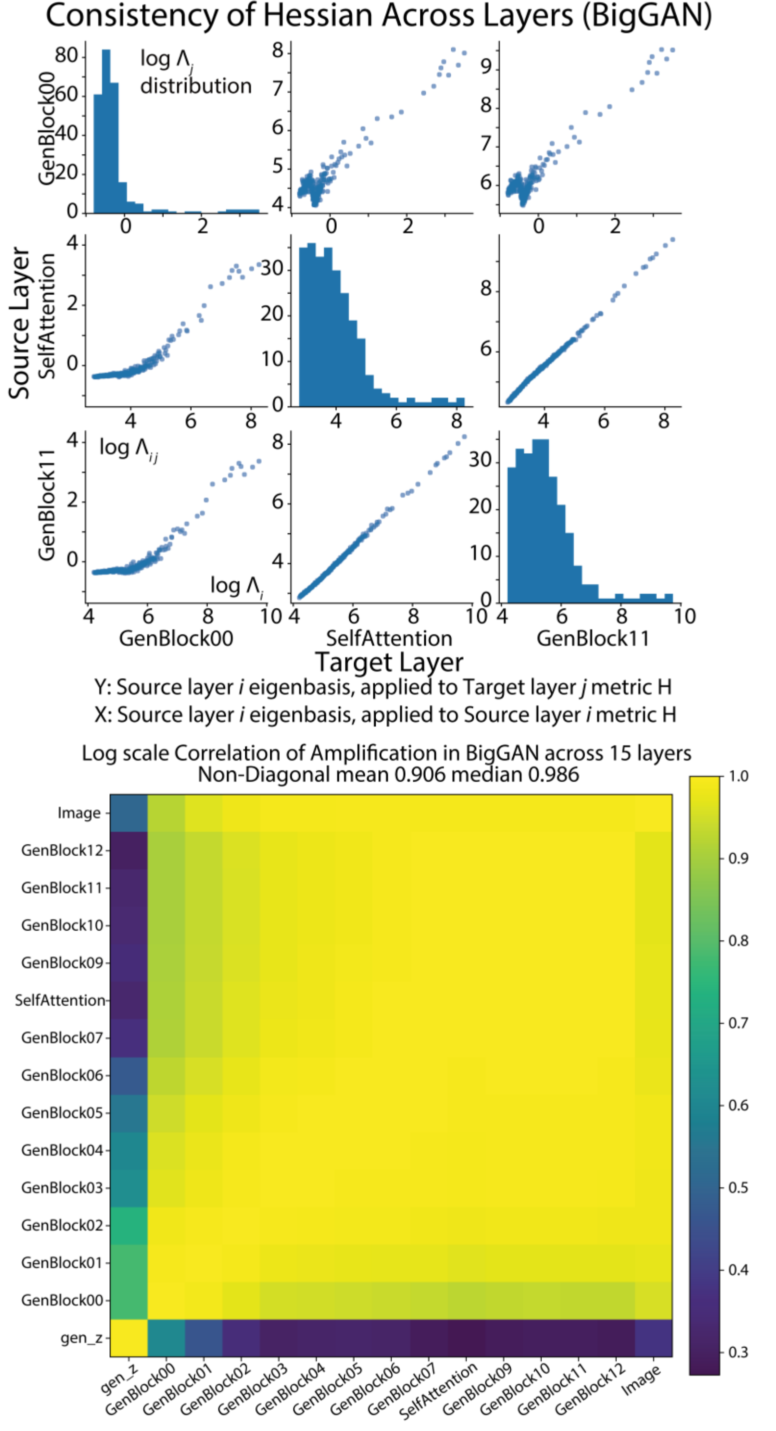




Alignment through layers



Median metric tensor correlation across 14 layers 0.994

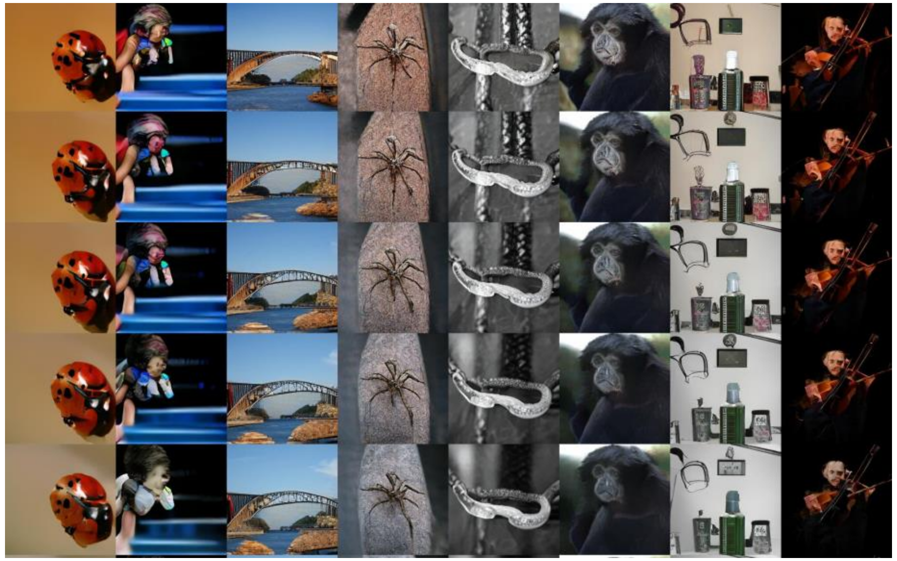


Implication: Null space

- Bottom eigenspace $\boldsymbol{u}_i$ induce "negligible" rate of change.
- It's effectively a null space!
- Implication
    1. Angle in the latent space not necessarily represents similarity.
    2. Each "interpretable" direction is associated with an approximately equivalent subspace of directions.
    $$\boldsymbol{v} \sim \boldsymbol{v} + \mu \boldsymbol{u}_i$$
    3. Exploration should not be directed in the null space.
    4. GAN is probably compressible.

Noise space Eigen 76-80 from center $\boldsymbol{z}_0 = 0$

Noise space Eigen 40 on references z_0



Implication: Extreme images in top eig space

- Noise vector is usually sampled on a (truncated) Gaussian sphere.

$$z \sim \mathcal{N}(0, I_d)$$
$$\|z\| \approx \sqrt{d}$$

  - z has approximately same norm.
  - z with the same norm have same prior density.
- Vectors within the top eigenspaces are usually too far away (in image space) from the center image, so usually extreme and unrealistic.
  - Assuming homogeneity, image space distance could be approximated by

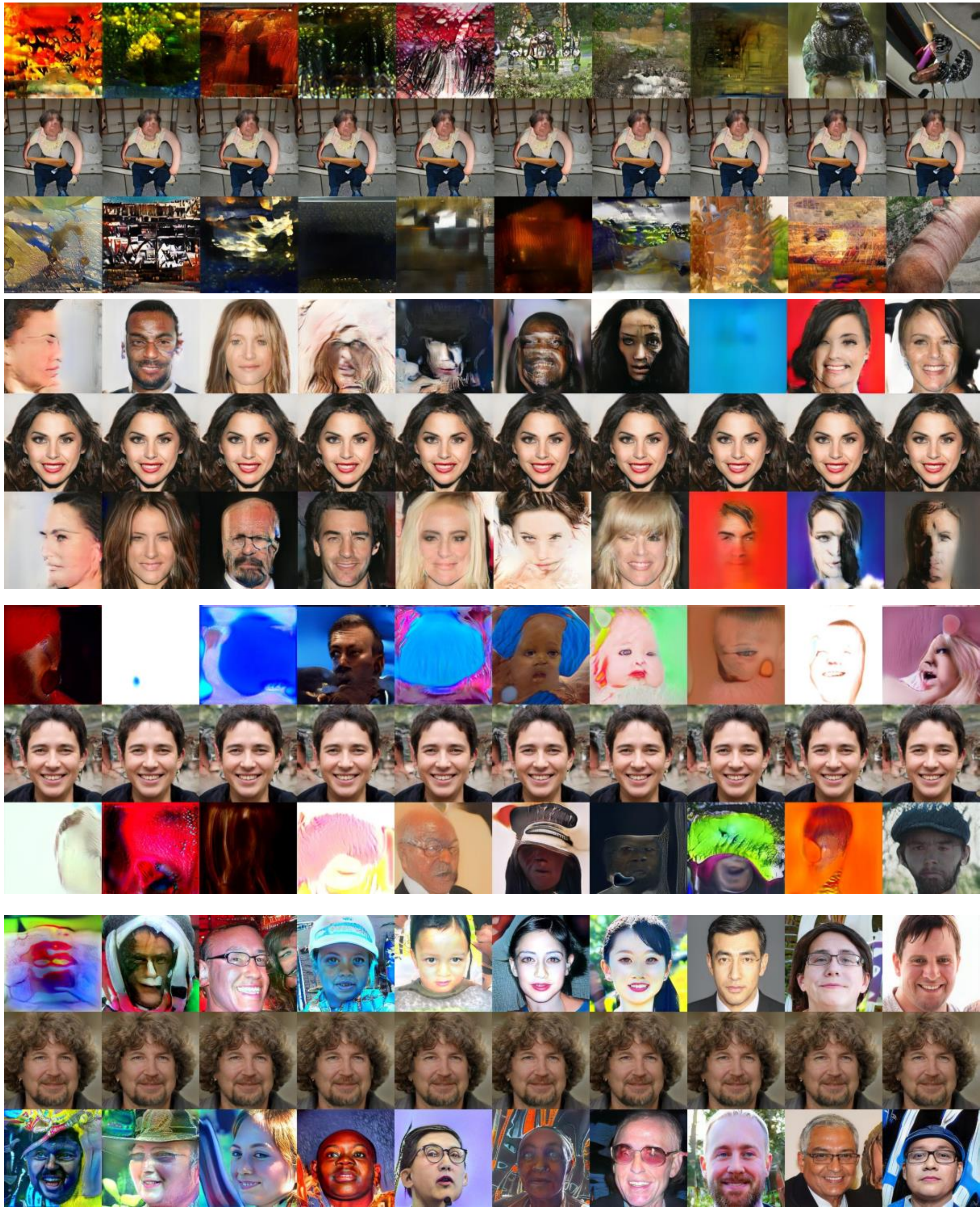
$$d^2\big(G(\mathbf{0}), G(z)\big) \approx z^T H(s) z$$

BigGAN

PGGAN

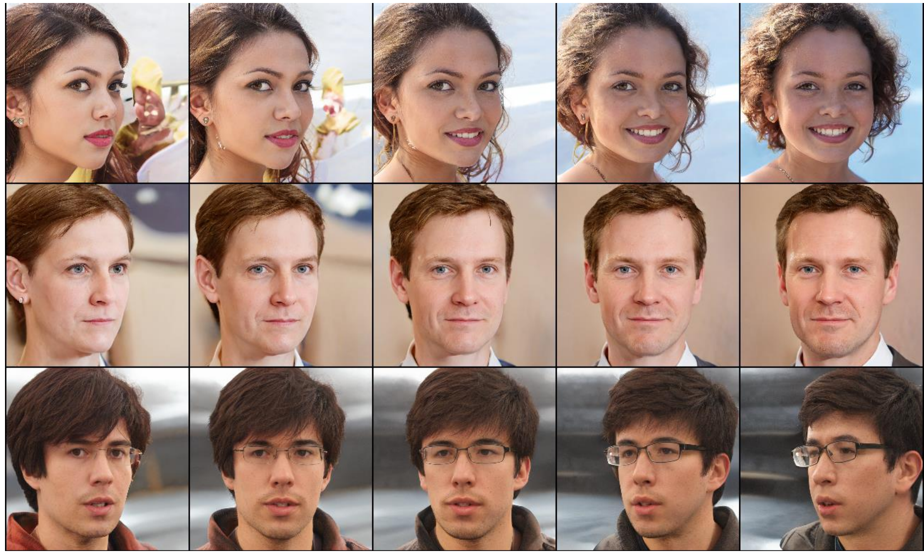
StyleGAN1

StyleGAN2



Applications of Geometric Knowledge

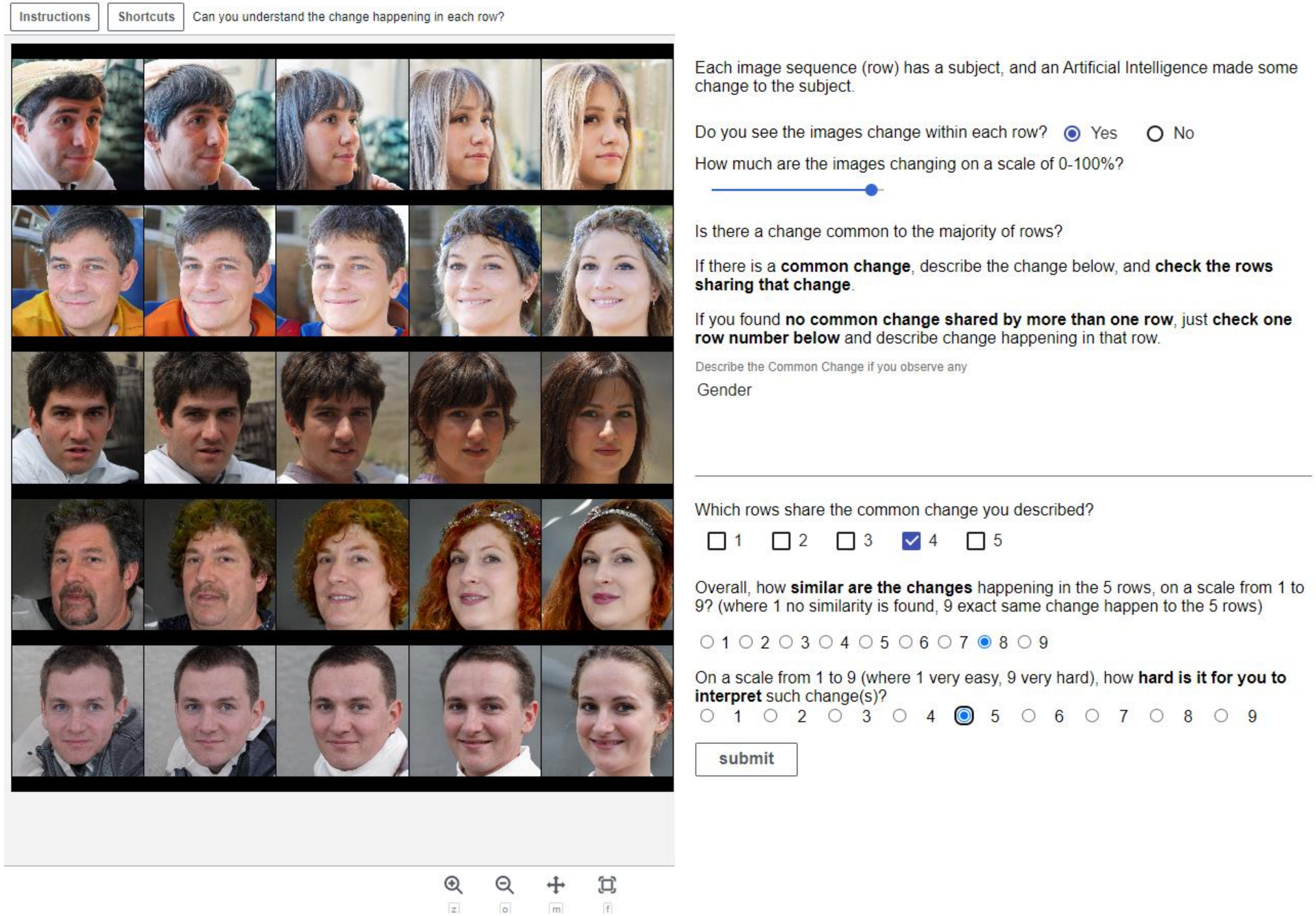
- Unify previous methods for unsupervised discovery of interpretable axes.
- Accelerate optimization on image manifold, both
  - Gradient based: ADAM
  - Gradient free: CMAES



StyleGAN2 Face eig 0

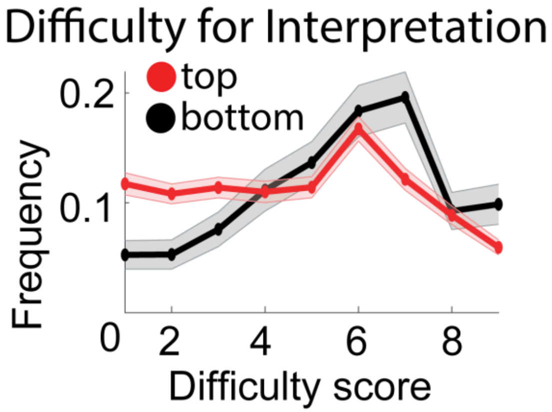
Previous work on unsupervised interpretability
Ramesh et al., 2018;
Härkönen et al., 2020;
Shen & Zhou, 2020;
Voynov & Babenko, 2020;
Peebles et al., 2020

Mturk User Study:



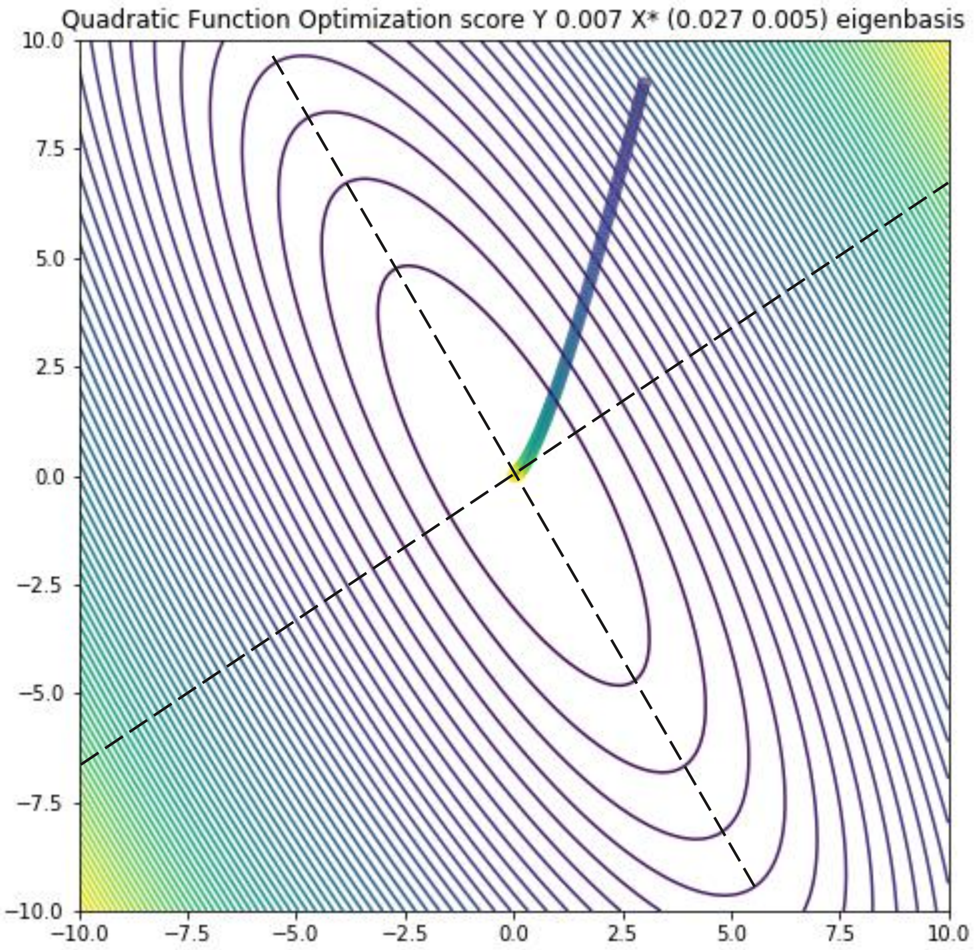
Subjects answer a series of questions

- See any change?
- How much change (0-100)
- Is there a common change?
  - Free text description of change
  - Which rows share the change
- How similar are the changes among 5 rows (1-9)
- How hard it is to interpret (1-9)



Accelerate Gradient-Based Optimization

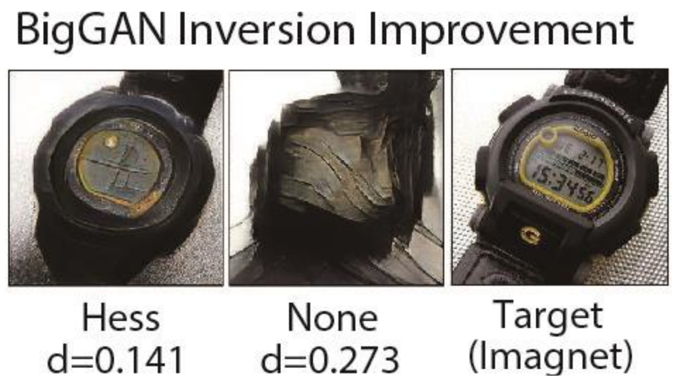
- Adam optimizer approximates a diagonal Hessian online.
- Eigenframe rotation makes Hessian each point more diagonal, accelerating Adam optimizers.
    - Original
    $$\mathrm{Adam}(z)$$
    - Hessian rotated
    $$\mathrm{Adam}(y), y := U^T z, \bar{H} = U\Lambda U^T$$

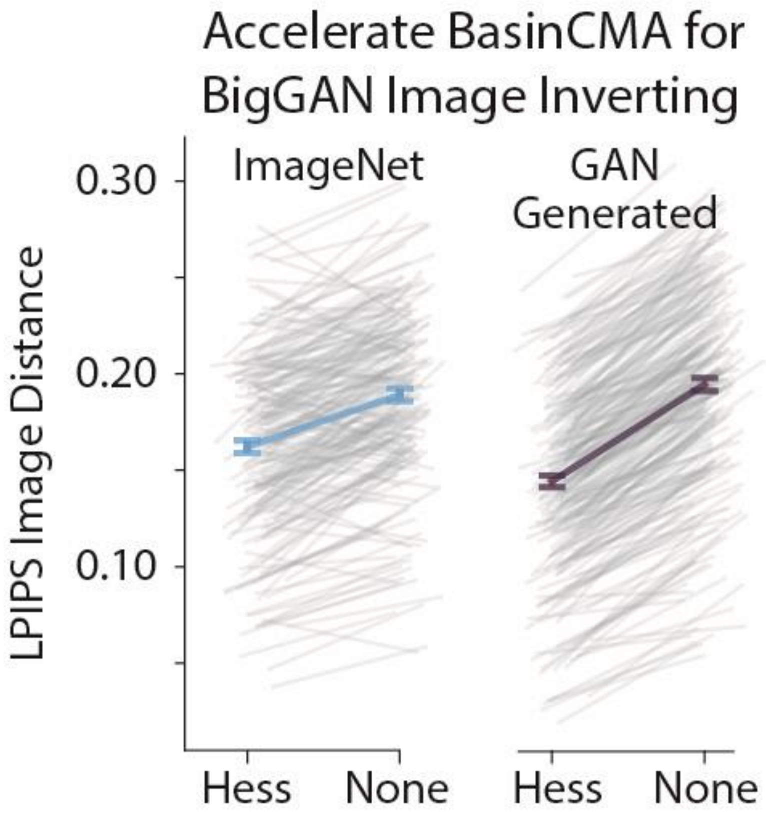


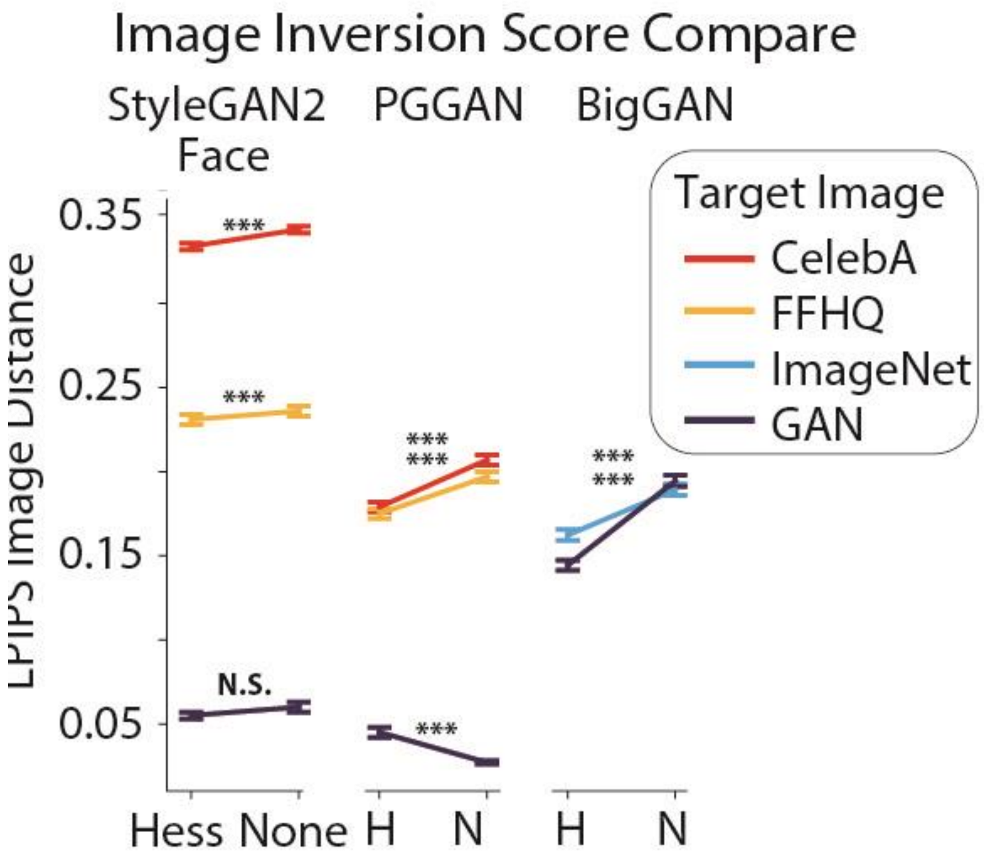
Improvement on GAN inversion

GAN inversion solve this optimization:

$$z^* = \arg\min_z d(G(z), I^*)$$







Accelerate Gradient-Free Optimization

- CMA-ES optimizer works by sampling and updating gaussian $\mathcal{N}(\mu_t, \Sigma_t)$
- Pre-computed metric tensor H could inform Σ, $\Sigma = AA^T$
  - Original CMA

$$\boldsymbol{z}_j^{t+1} = \bar{\boldsymbol{z}}^t + A\boldsymbol{y_j}, \boldsymbol{y_j} \sim N\left(0, I^d\right)$$
$$\text{Update } A \text{ online}$$

  - Hess-CMA

$$\boldsymbol{z}_j^{t+1} = \bar{\boldsymbol{z}}^t + U_r \Lambda_r^{-\alpha} \boldsymbol{y_j}, \boldsymbol{y_j} \sim N\left(0, I^d\right)$$
$$H = U \Lambda U^T,$$
$$U_r = [u_1, u_2 \dots u_r]$$
$$\Lambda_r = \mathrm{diag}(\lambda_1, \lambda_2 \dots \lambda_r)$$

  - α scales the step sizes w.r.t. global geometry.
  - r Cut off uninformative axes.

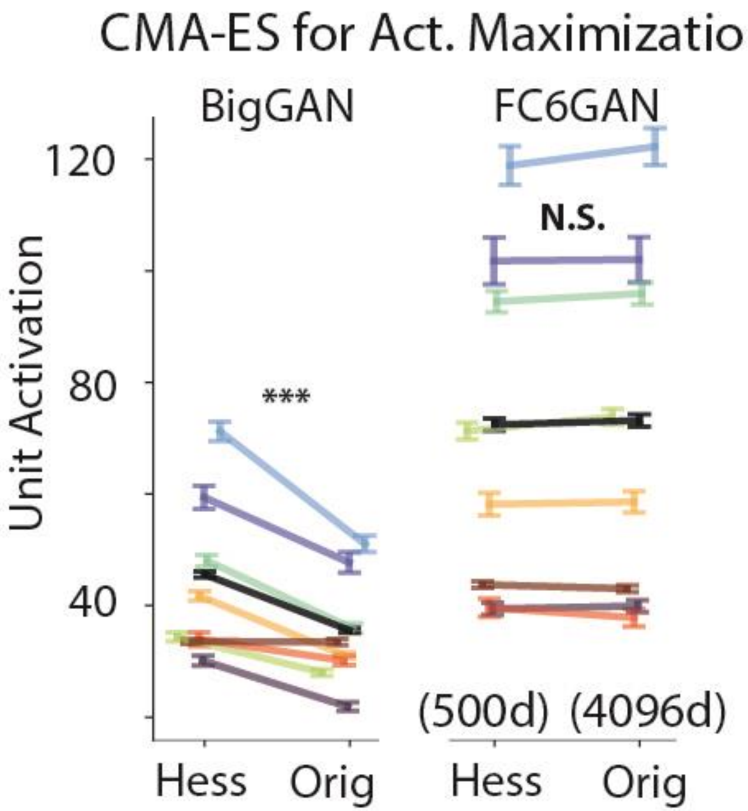
Improvement on Activation Maximization

Activation maximization solve this optimization:

$$z^* = \arg\max_z \phi(G(z))$$

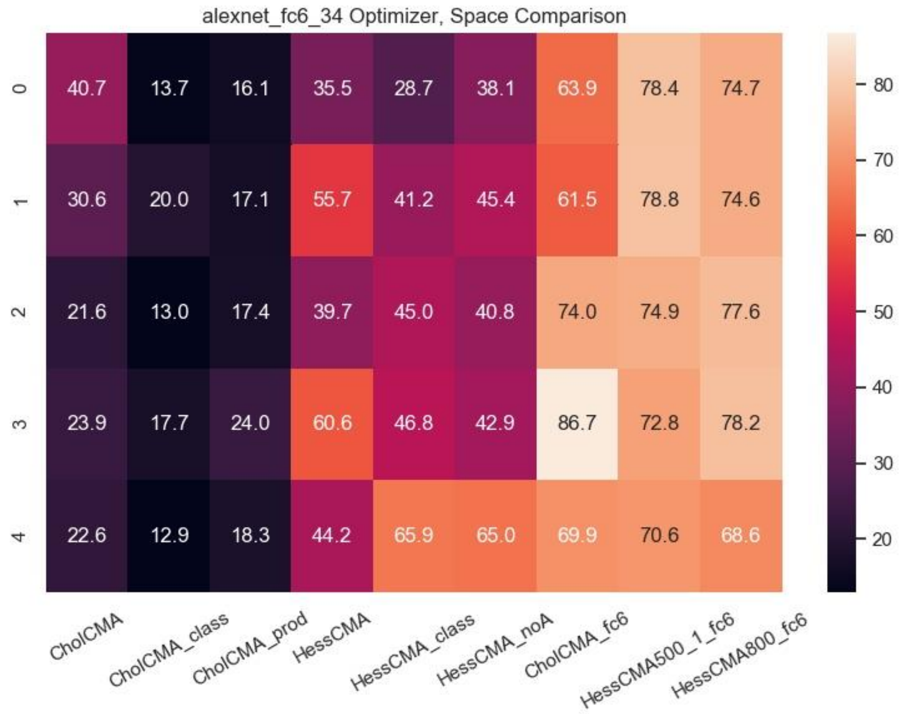
ϕ is a function over image space (e.g. a unit in CNN like AlexNet).

Useful for black box attack and feature visualization

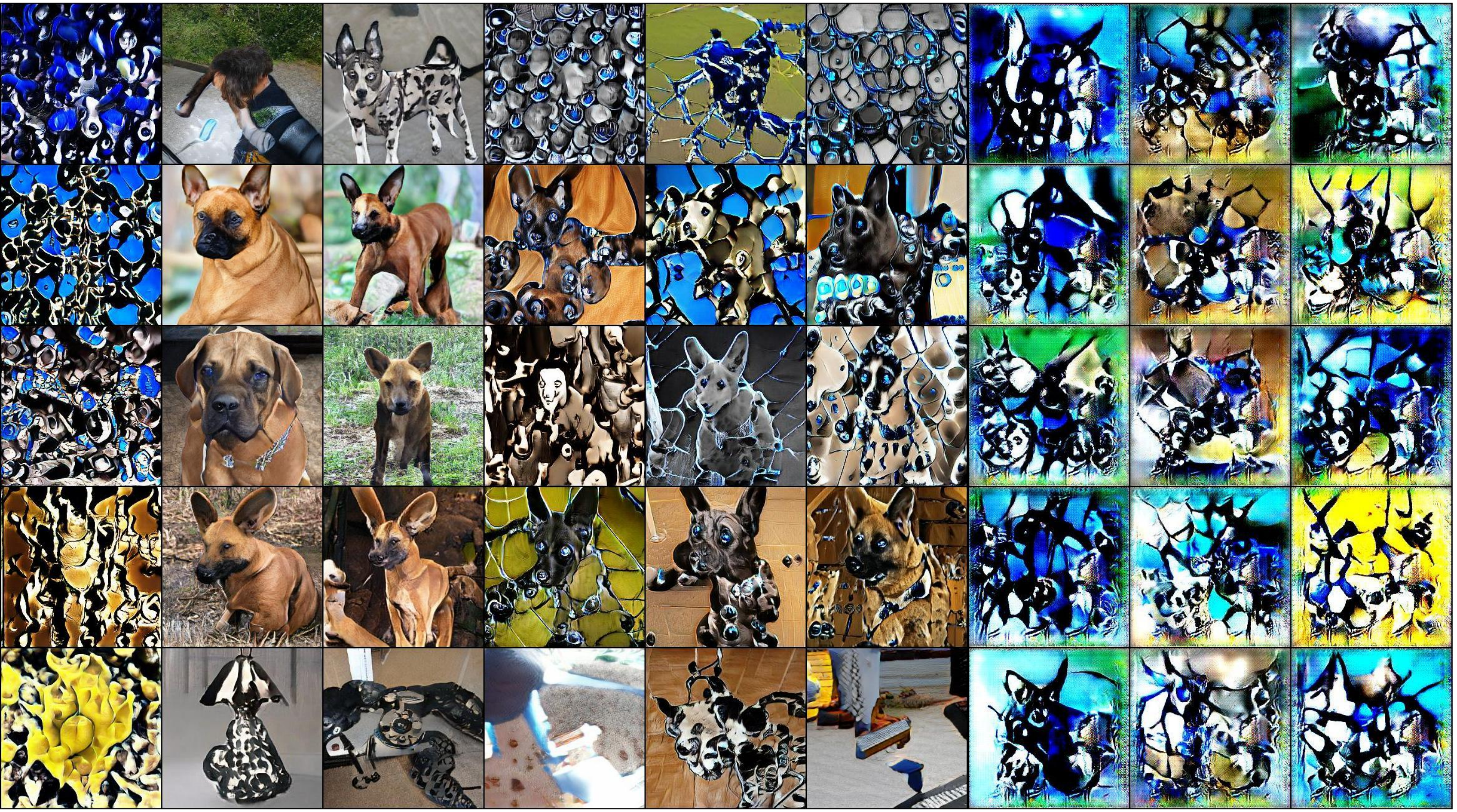


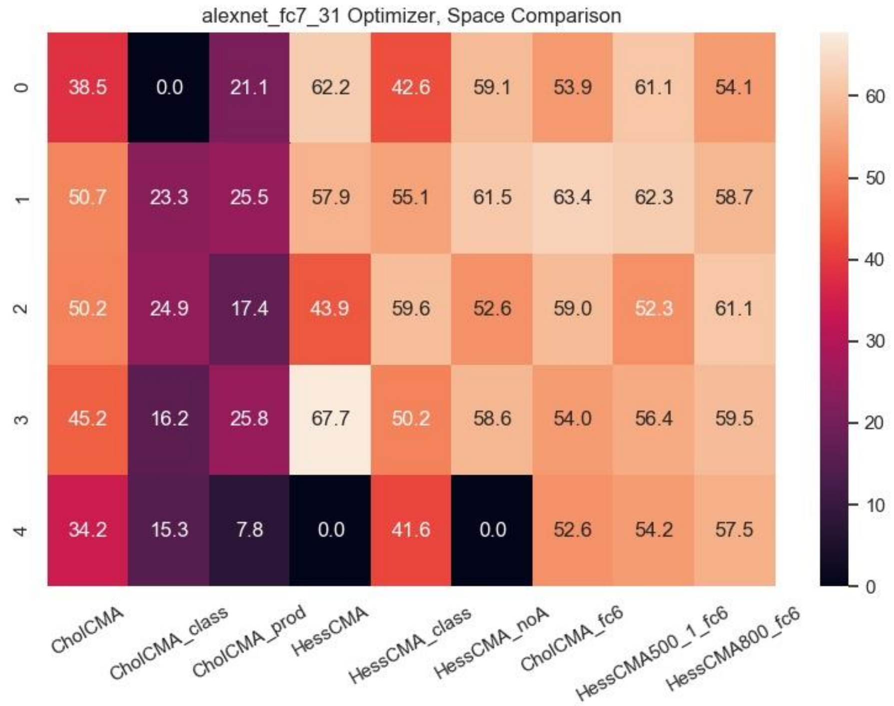
Color = Layer in Alex-Net

Improvement on Activation Maximization



Improvement on Activation Maximization





Discussion: Nonlinear PCA of Natural Image Manifolds

- Different GAN models trained on human face datasets, the top eigen spaces contain axes that represent similar semantics.
- Maybe those are the major nonlinear axes in the true manifold of all faces.

Summary of Contributions

- We propose an architecture-agnostic way to efficiently measure the geometric structure of deep generative models.
- We characterize the common geometric features (i.e. anisotropy & homogeneity & alignment) across multiple ($n = 8$) modern generative models.
- Top eigenspaces usually encode interpretable transforms.
- Global metric tensor is helpful for optimization on the GAN manifold with or without gradient.

Acknowledgement

Advisors

- Carlos Ponce
- Tim Holy

Friends and colleagues that read and provide suggestions on the manuscript

- Zhengdao Chen, *NYU Courant*
- Yunyi Shen, *U Wisc Madison->UToronto*
- Hao Sun, *CUHK->Cambridge*
- Lingwei Kong, *Arizona State*
- Yuxiu Shao, *LNC2, ENS Paris*
- Kaining Zhang, *Wash U in St. Louis*

Washington University in St.Louis

Cognitive, Computational and Systems Neuroscience Pathway (CCSN)

McDonnell Center for Systems Neuroscience