
GANSpace: Discovering Interpretable GAN Controls

Erik Härkönen^{1,2}

Aaron Hertzmann²

Jaakko Lehtinen^{1,3}

Sylvain Paris²

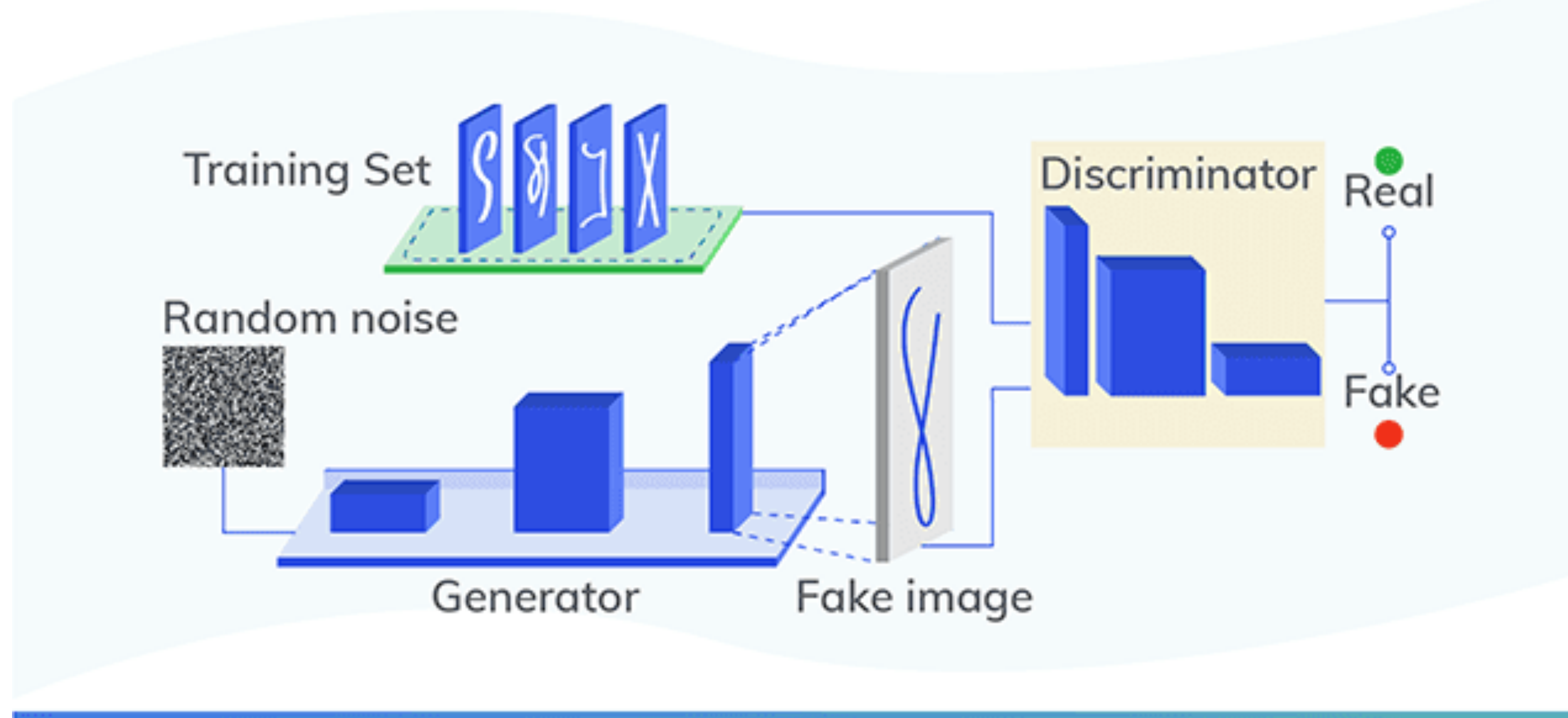
¹Aalto University ²Adobe Research ³NVIDIA

Slides by Nolan Dey

<https://arxiv.org/pdf/2004.02546.pdf>

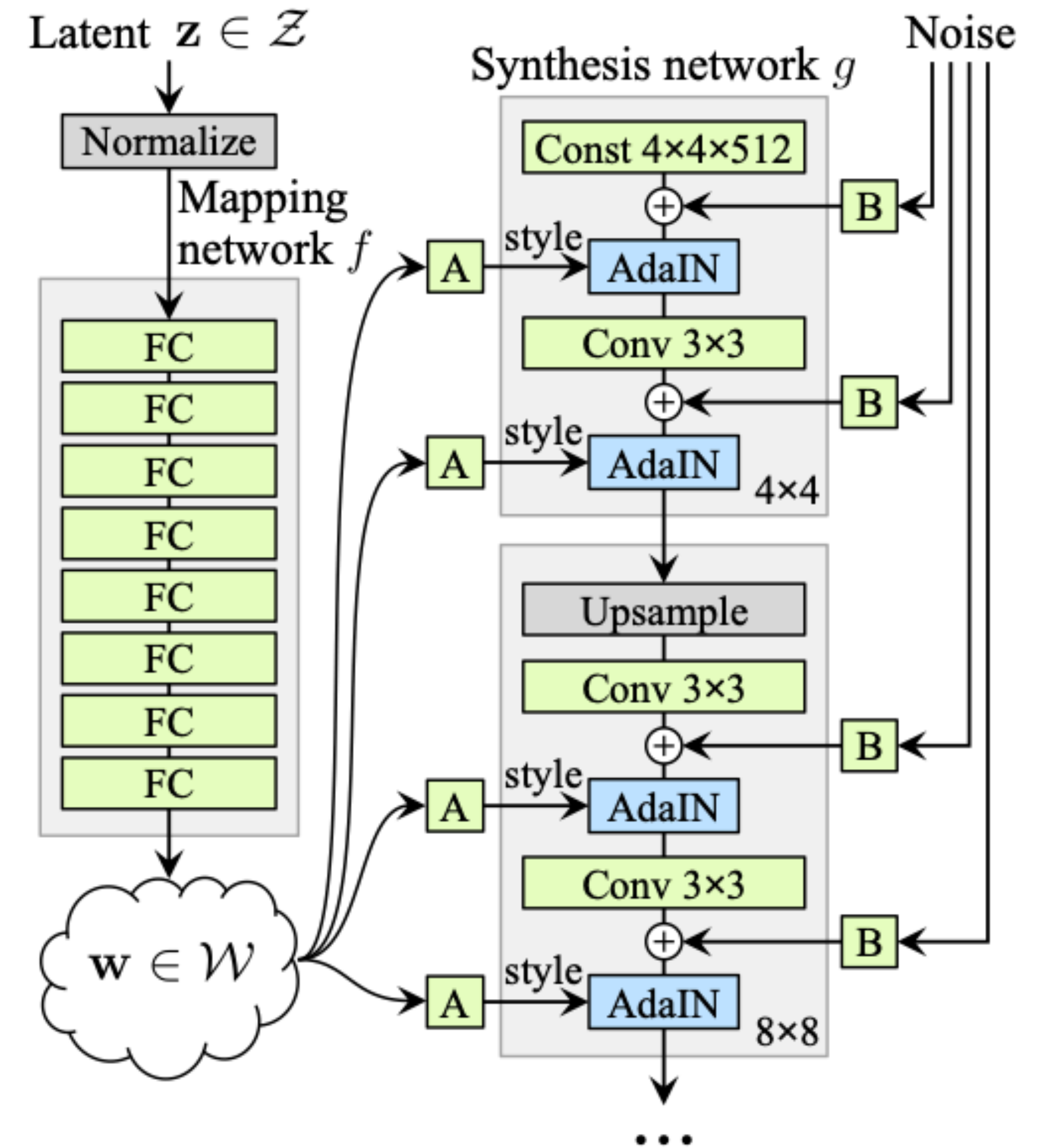
What is a GAN?

Generative Adversarial Networks



StyleGAN

- Sample random points in Z
- Forward pass to get points in W
- Compute PCA in W
- Edit PCA component values are forward pass to control generated output



<https://arxiv.org/pdf/1812.04948.pdf>

<https://arxiv.org/pdf/2004.02546.pdf>

BigGAN

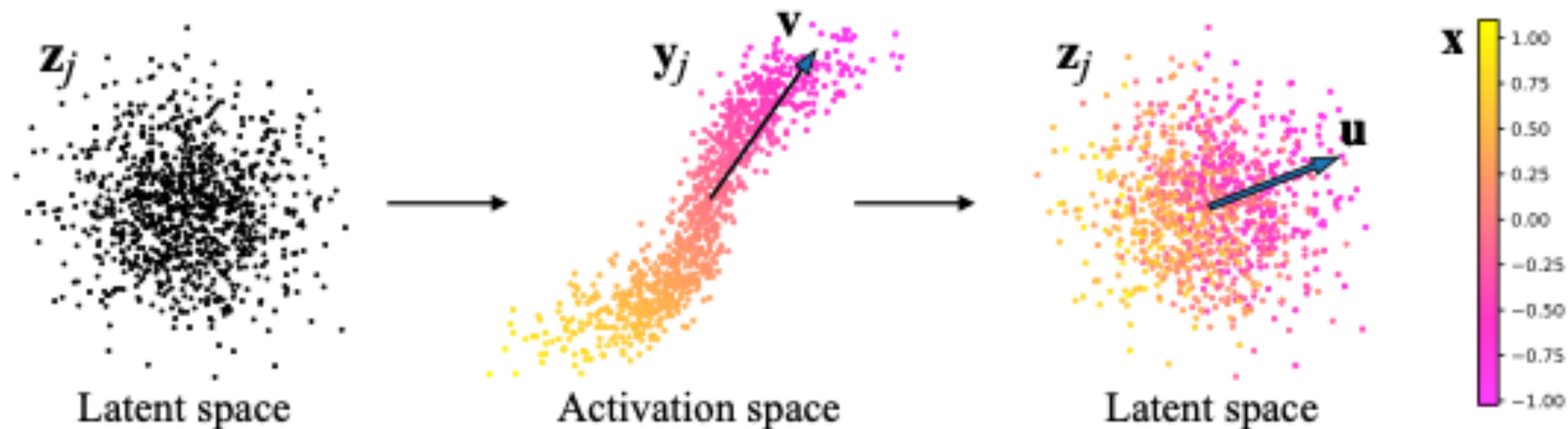


Figure 2: 2D Illustration of identifying a principal activation direction for BigGAN. Random latent vectors z_j are sampled, and converted to activations y_j . The PCA direction v is computed from the samples, and PCA coordinates x_j computed, shown here by color-coding. Finally, back in the latent space, the direction u is computed by regression from z_j to x_j .

<https://arxiv.org/pdf/2004.02546.pdf>

Discovering Interpretable GAN Controls

Erik Härkönen^{1,2}, Aaron Hertzmann², Jaakko Lehtinen^{1,3}, Sylvain Paris²

¹Aalto University ²Adobe Research ³NVIDIA

https://www.youtube.com/watch?v=jdTICDa_eAI&feature=youtu.be

Findings and results

- Earlier layers gave more useful controls than later layers
- disallowed combinations
 - wrinkles cannot be added to child's face
 - makeup and lipstick have little to no effect on male faces
- some entanglements were still observed
 - making a car more sporty causes a more open road background
 - more family car appears in woods or city streets
 - rotating dogs causes their mouths to open
- PCA components matched the edit directions from the GAN steerability paper

<https://arxiv.org/pdf/2004.02546.pdf>

Discussion Topics

- Do generative models learning similar image representations as discriminative networks?
- GANSpace for data augmentation
- Could we use GANSpace to develop natural adversarial edits?