Towards a Conditional Generative Model for Trajectory Modeling in Medical Imaging

Anita Kriz

ECSE 556

Table of Contents

Introduction & Motivation

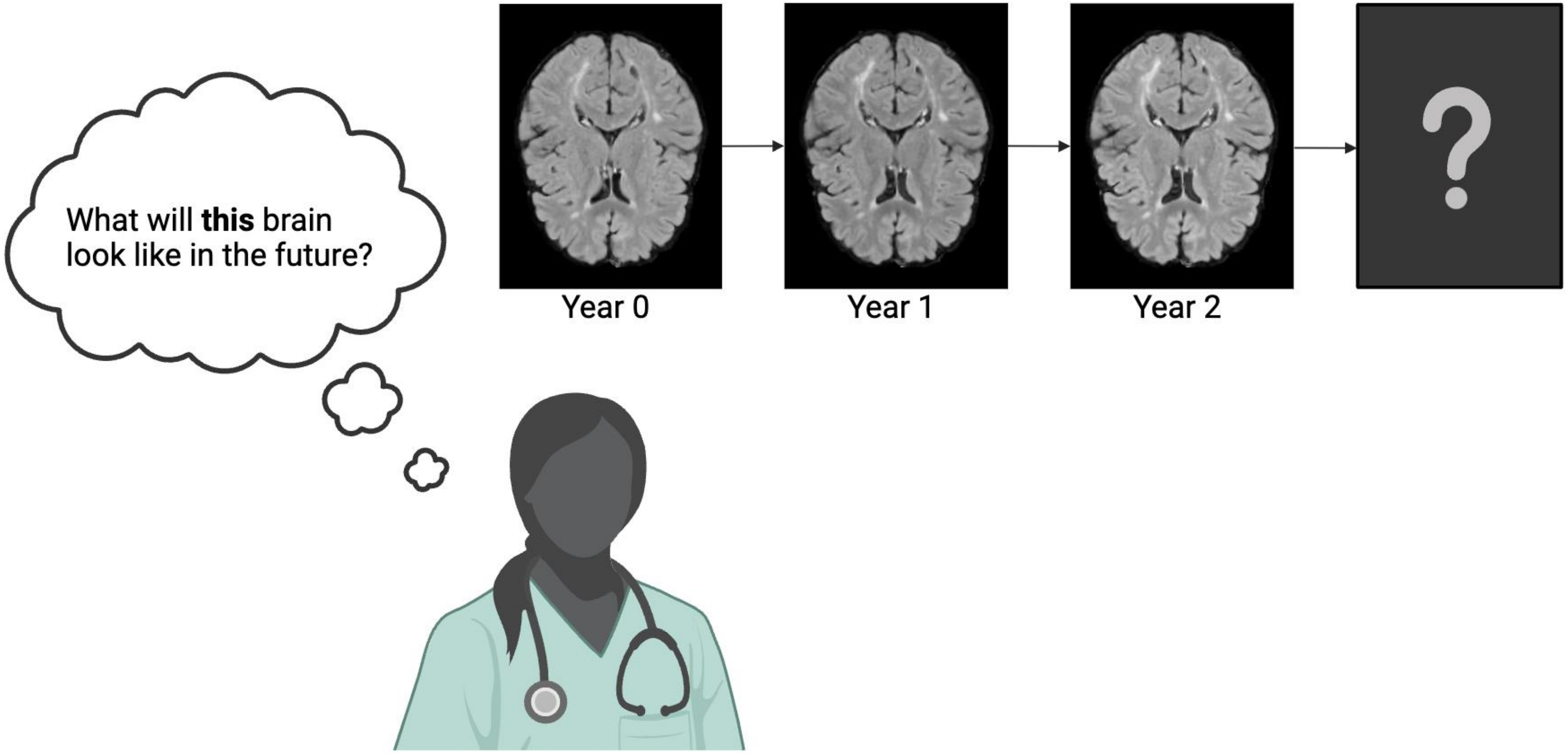
Generative Models

DALL-E2:
"photograph of an astronaut riding a horse"

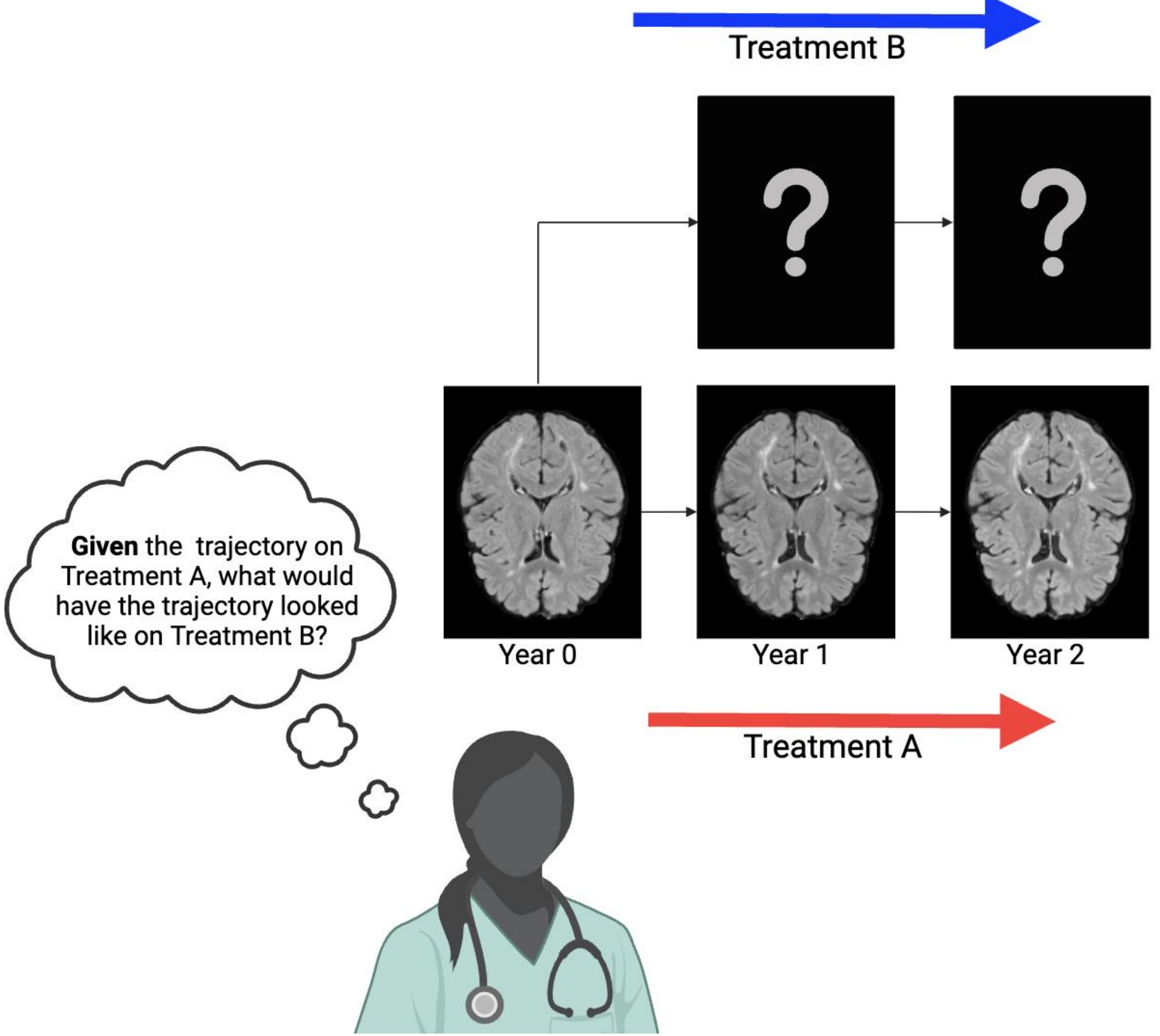
DALL-E2:
"an oil pastel drawing of an annoyed cat in a spaceship"

Why is there a need for generative models in (Multiple Sclerosis) Medical Imaging?

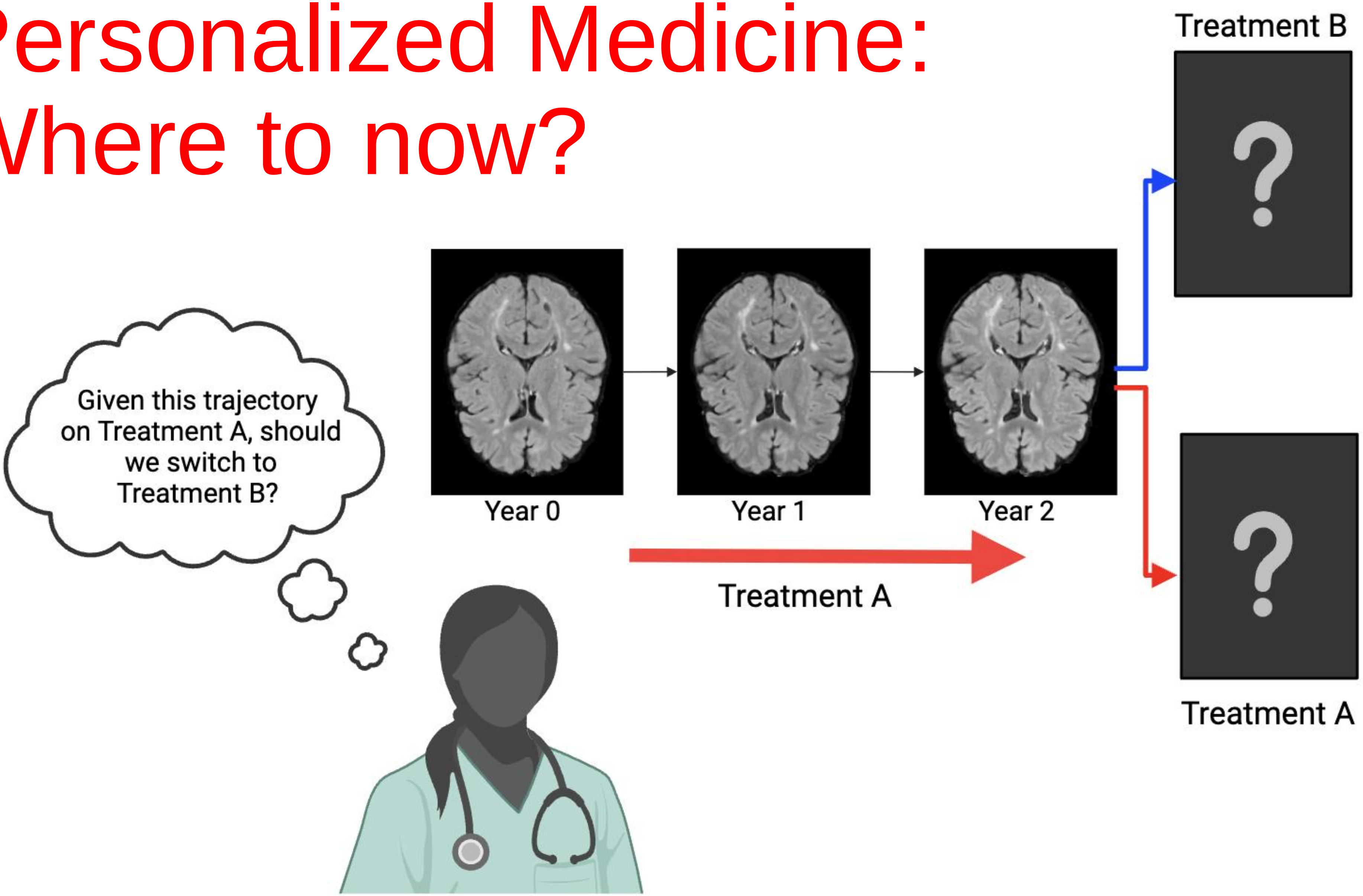
Disease Progression: Where in the Future?



Treatment Effect: Where in the Past?



Personalized Medicine: Where to now?



Constraints

Constraints in (Multiple Sclerosis) Imaging

- We want to generate **patient specific** images
- We want clinically relevant – not random - generation
- We want to distinguish between attributes that are often **highly correlated** (brain morphology and disease progression)

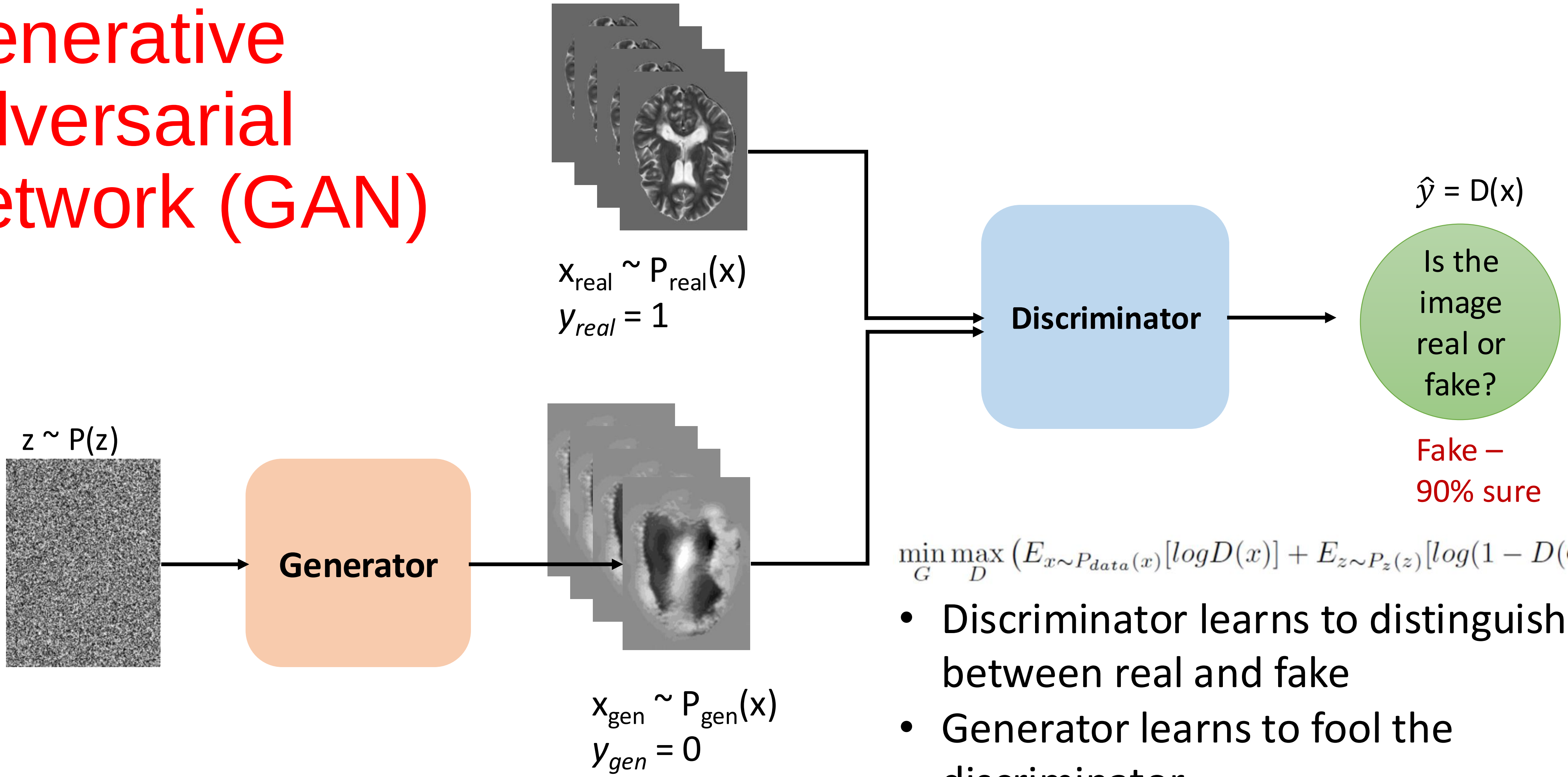
Objectives

Objectives

1. Generate images such that certain attributes (e.g., number of lesions) are changed
2. Preserve the attribute excluding details (e.g., brain identity, age)
3. Maintain realistic image generation

Background

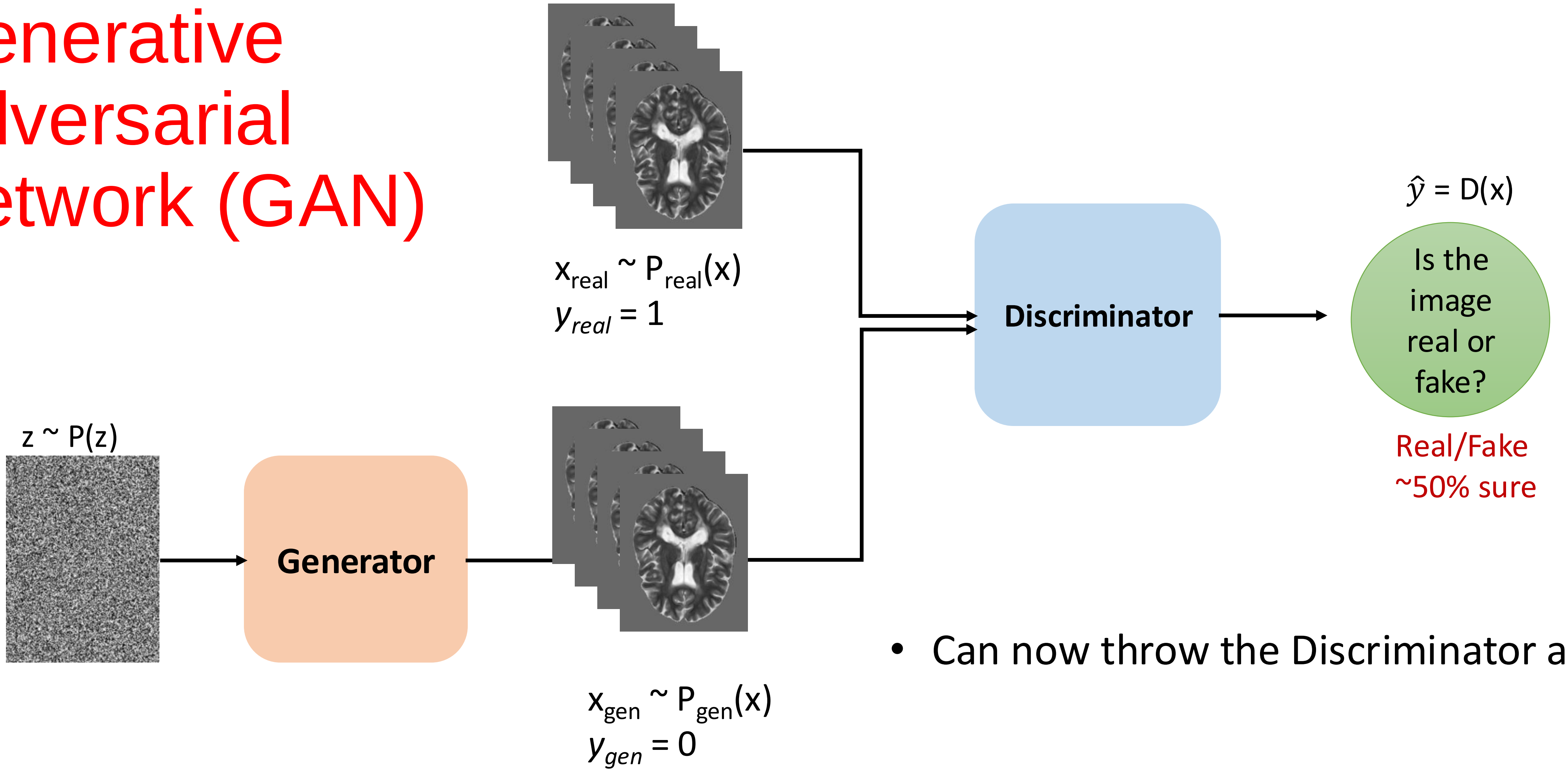
Generative Adversarial Network (GAN)



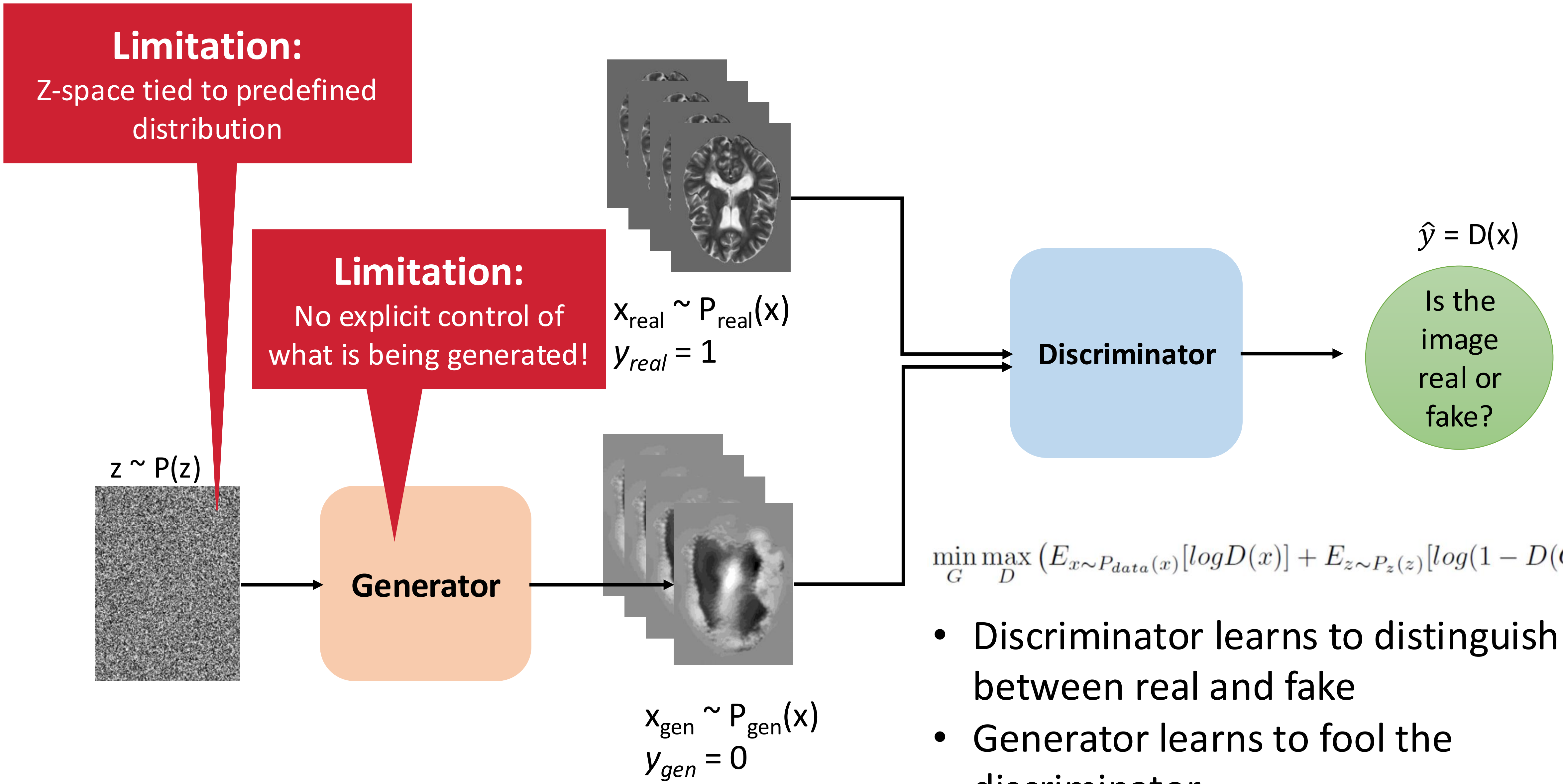
$$\min_G \max_D \left(E_{x \sim P_{data}(x)}[logD(x)] + E_{z \sim P_z(z)}[log(1 - D(G(z)))]\right)$$

- Discriminator learns to distinguish between real and fake
- Generator learns to fool the discriminator

Generative Adversarial Network (GAN)

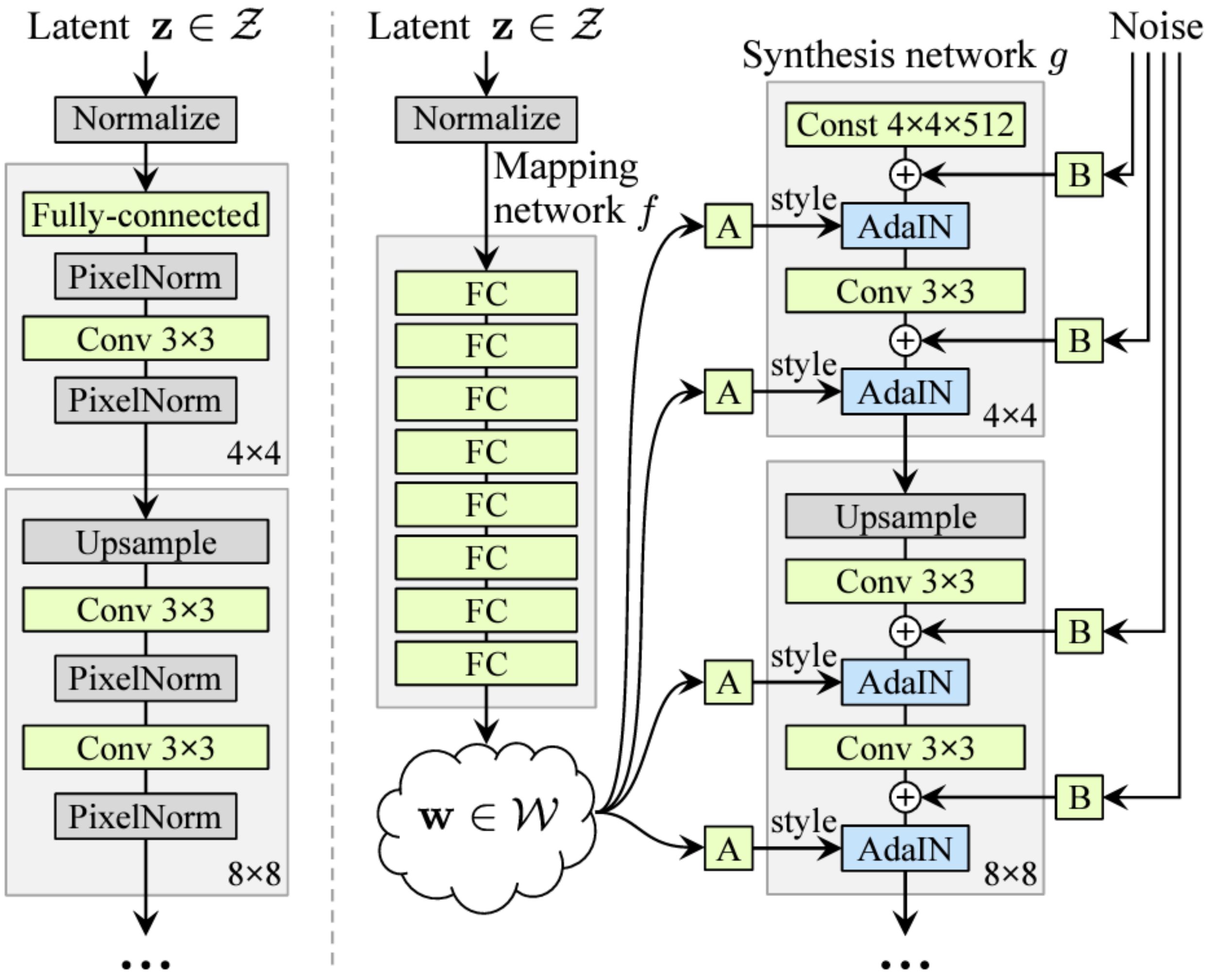


- Can now throw the Discriminator away



$$\min_G \max_D \left(E_{x\sim P_{data}(x)}[logD(x)] + E_{z\sim P_z(z)}[log(1-D(G(z)))]\right)$$

- Discriminator learns to distinguish between real and fake
- Generator learns to fool the discriminator



(a) Traditional (b) Style-based generator

StyleGAN

3 major improvements:

- **Mapping network f**
  W does not have to support sampling from fixed dist.
- **Adaptive Instance Normalization**
  Inject style of y into content of x

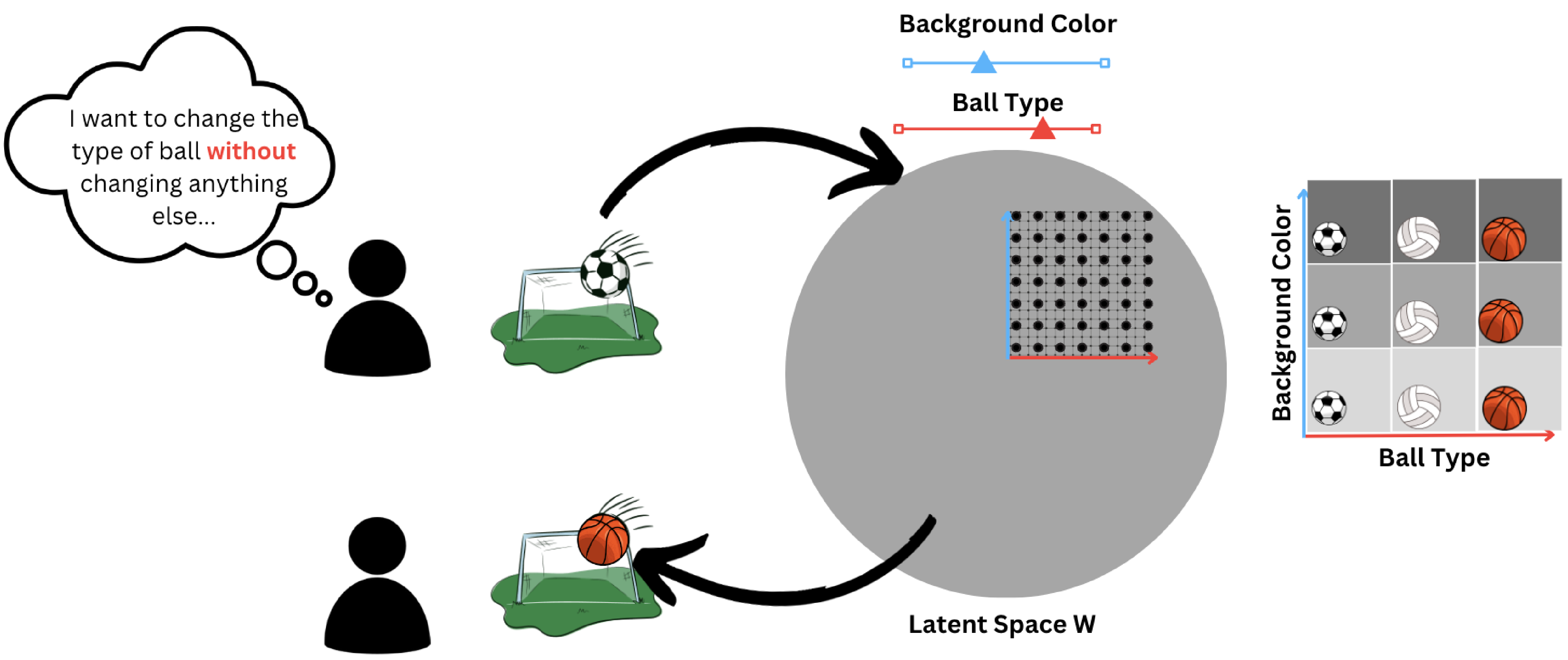
$$AdaIN(x, y) = \sigma(y)(\frac{x - \mu(x)}{\sigma(x)}) + \mu(y)$$

- **Noise**
  Induce stochastic variation in the image

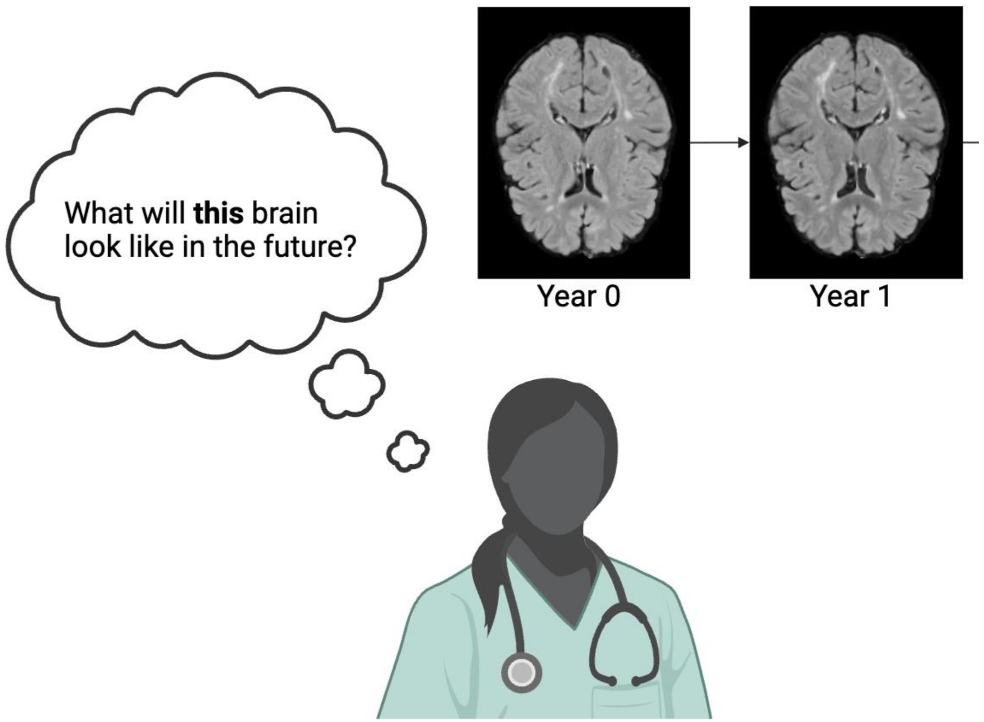
Overall idea:

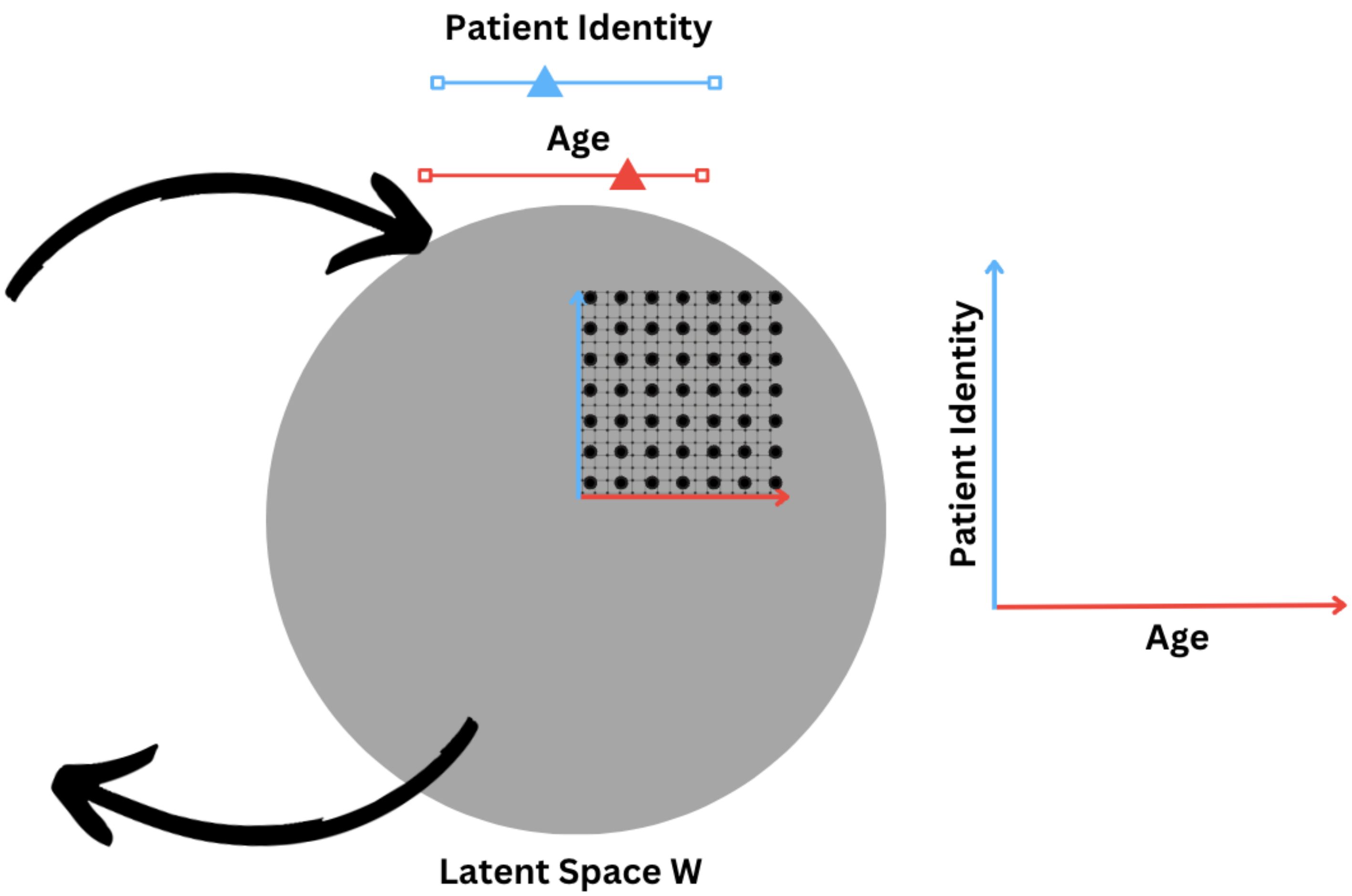
By separating style from content, StyleGAN provides a strong baseline for a **disentangled latent space** where we can theoretically **control** different attributes of an image

Disentanglement for Conditional Generation



Can we apply this to Medical Imaging?





Methodology

Methodology Pipeline

PHASE 1
Train a 2D StyleGAN

PHASE 2
Measure Disentanglement & Compare to GAN

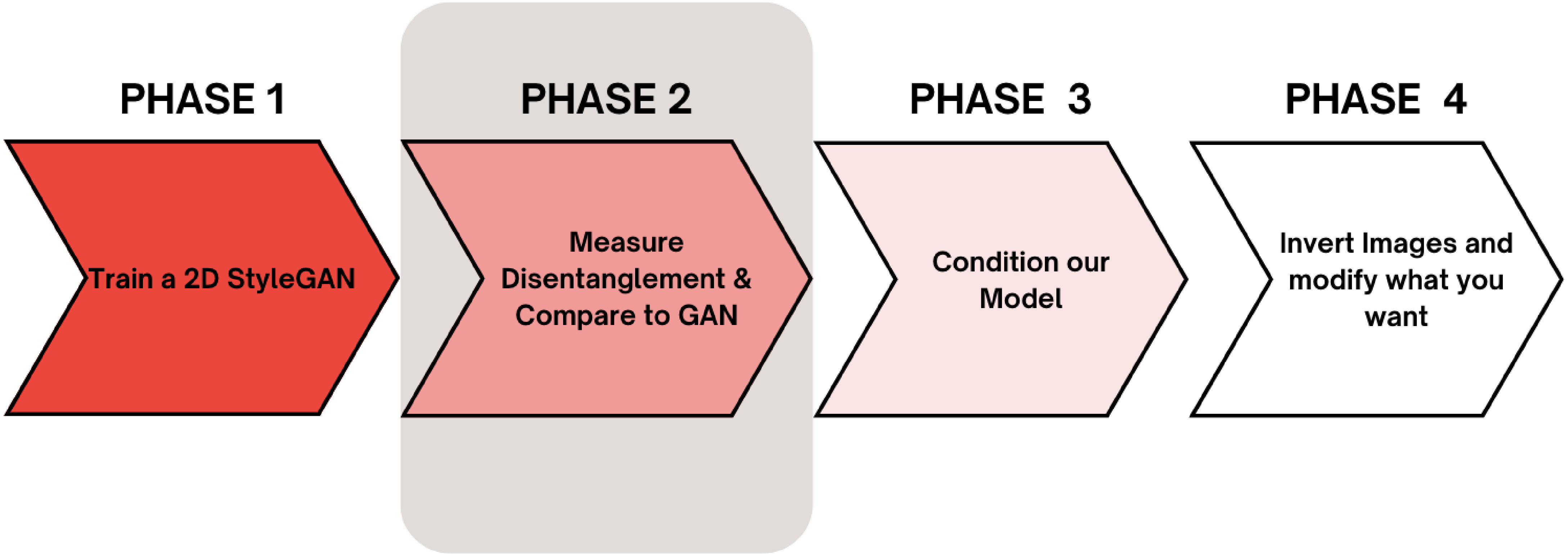
PHASE 3
Condition our Model

PHASE 4
Invert Images and modify what you want

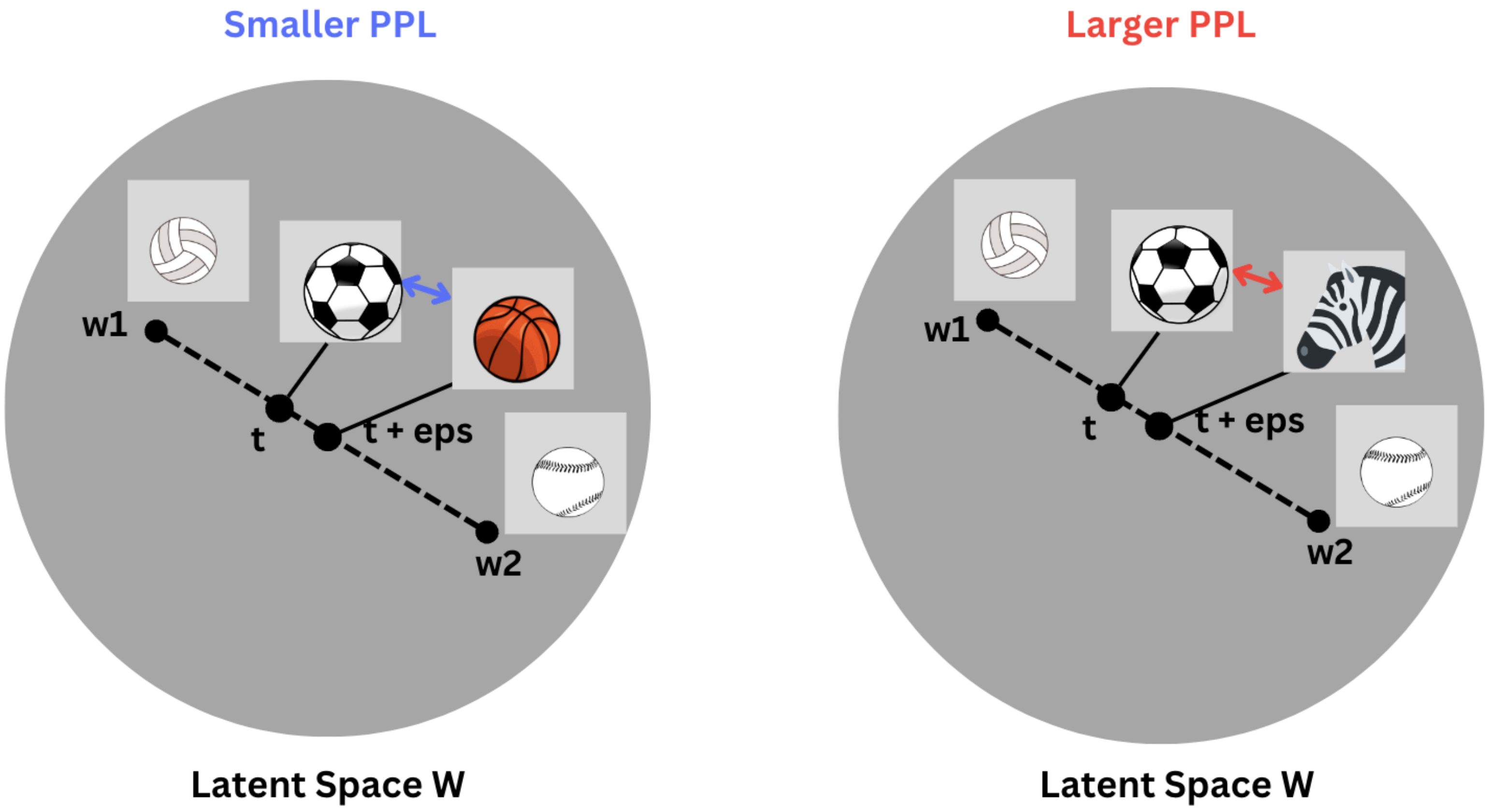
Dataset & Resources

- Multiple Sclerosis with ~5500 patients and access to 16 GB GPUs were provided by the PVG Lab
- Code for a 3D StyleGAN was written at the PVG Lab and thus only needed to be converted to 2D

Methodology Pipeline



Perceptual Path Length (PPL)



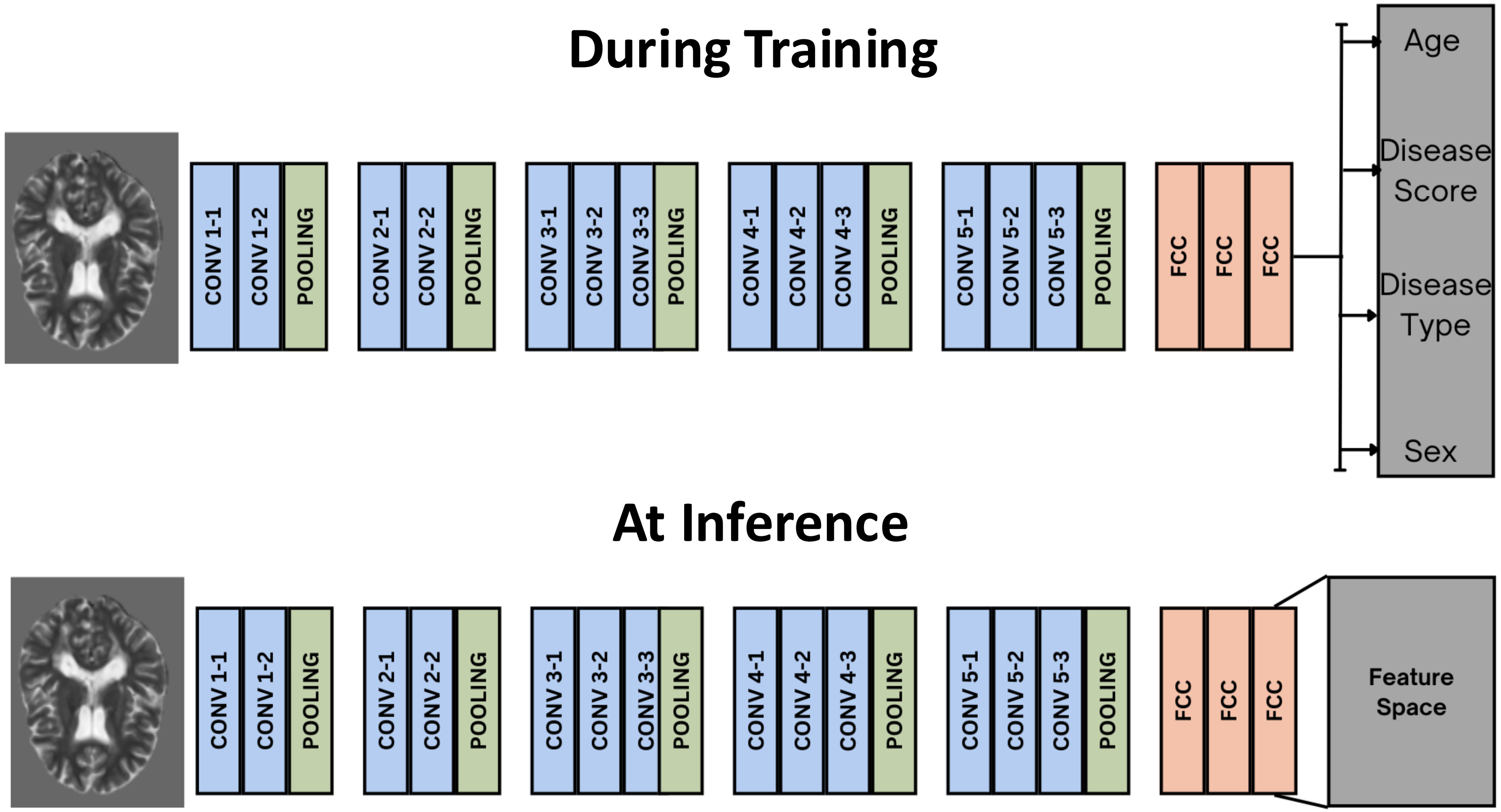
Perceptual Path Length:

$$l_{\mathcal{W}} = \mathbb{E}\Big[\frac{1}{\epsilon^2} d\big(g(\mathrm{lerp}(f(\mathbf{z}_1), f(\mathbf{z}_2); t)), g(\mathrm{lerp}(f(\mathbf{z}_1), f(\mathbf{z}_2); t+\epsilon))\big)\Big],$$

Idea:

Quantify this distance, d , such that **perceptually** similar images achieve a low PPL score

VGG-16 Model



Perceptual Path Length:

$$l_{\mathcal{W}} = \mathbb{E}\Big[\frac{1}{\epsilon^2} d\big(g(\mathrm{lerp}(f(\mathbf{z}_1), f(\mathbf{z}_2);\, t)),\\ g(\mathrm{lerp}(f(\mathbf{z}_1), f(\mathbf{z}_2);\, t+\epsilon))\big)\Big],$$

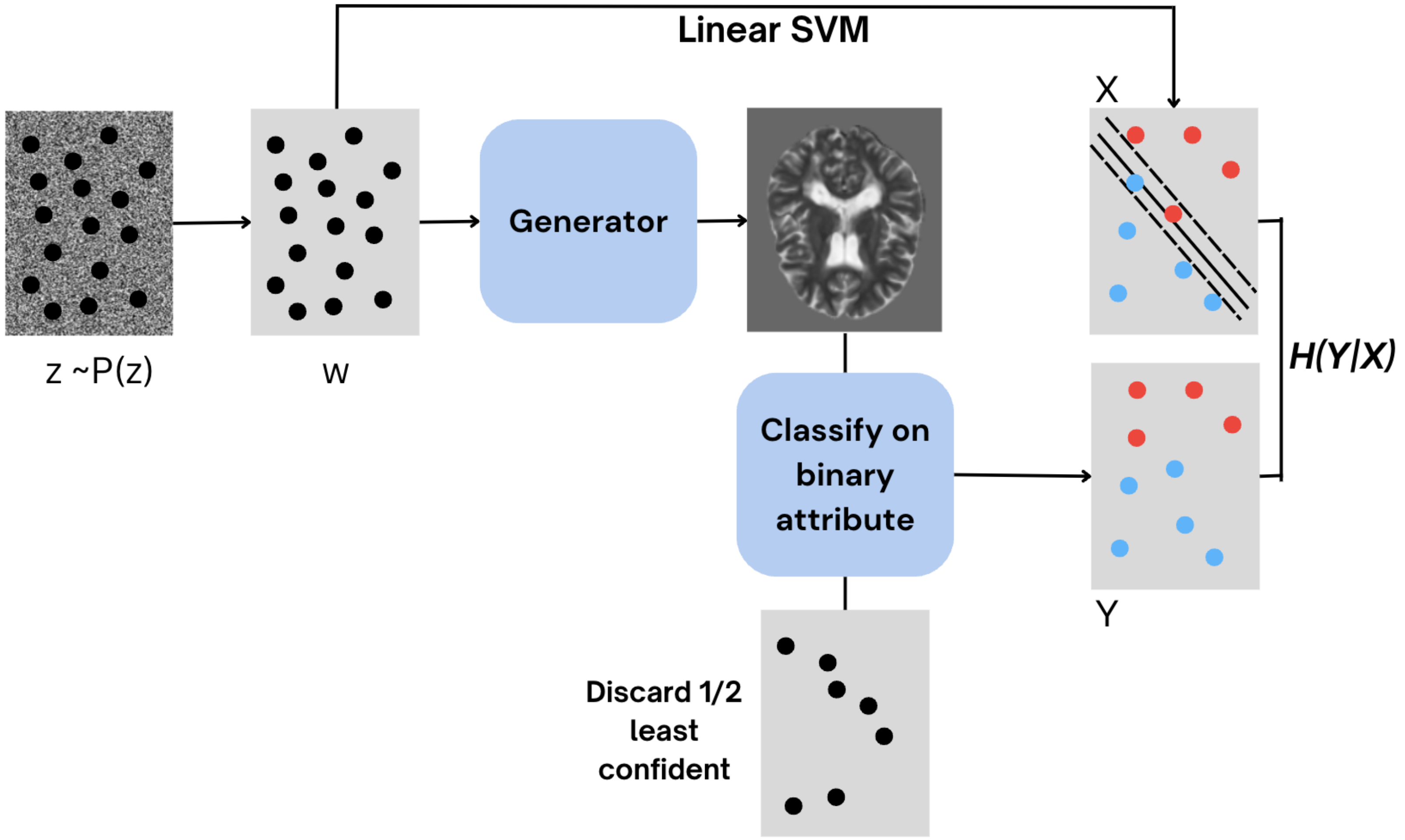
Idea:

Quantify this distance, d , such that **perceptually** similar images achieve a low PPL score

Solution:

1. Train a VGG-16 model
2. Use its final layer as a feature representation of the image
3. Find the distance between the feature vectors for PPL

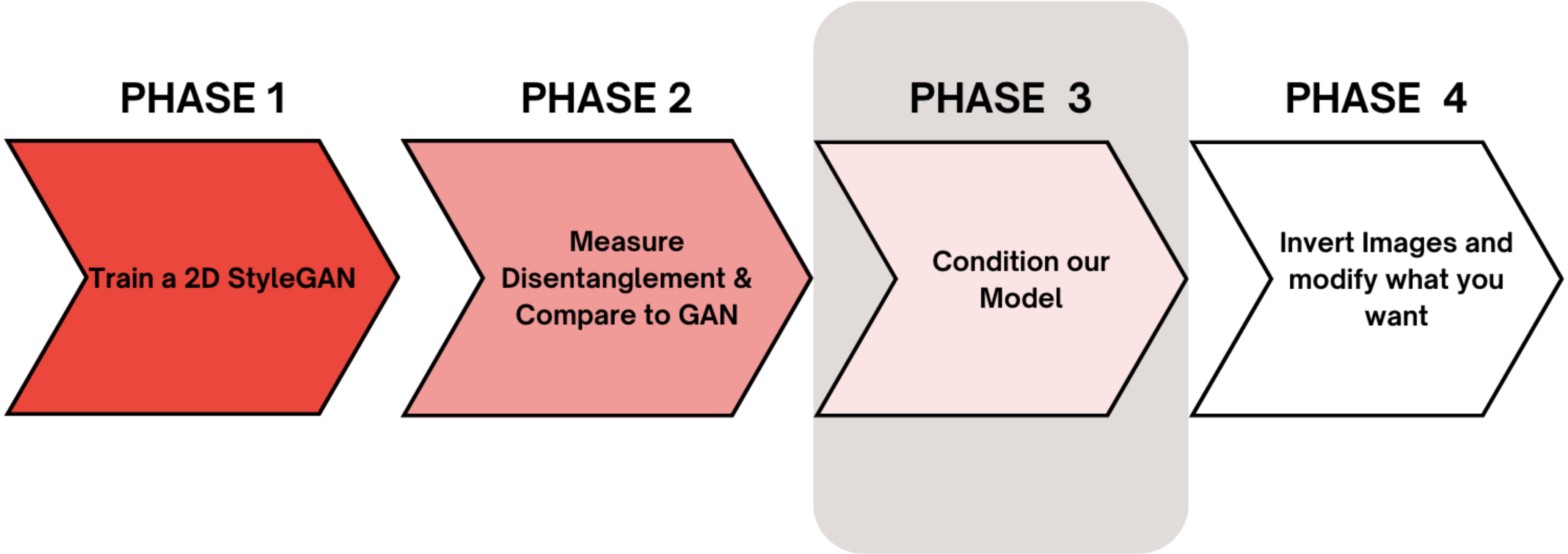
Linear Separability



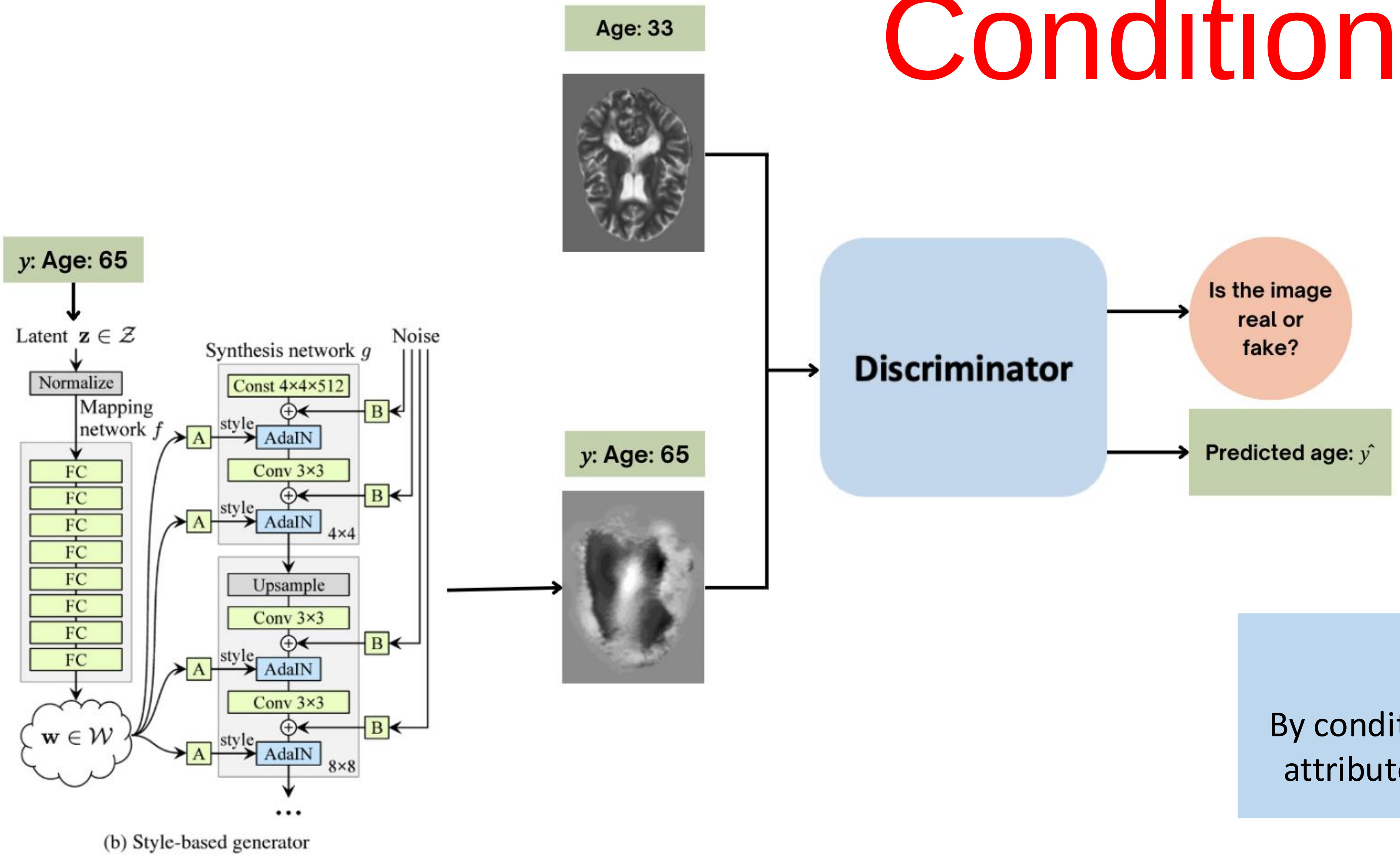
Idea:

1. Sample 200,000 points from z space and generate their images
2. Classify them using a pre-trained classifier based on binary attributes (Y)
3. Discard the 100,000 least confident images
4. Fit a linear SVM to categorize latent space points (X)
5. Compute H(Y|X)

Methodology Pipeline



Conditional model



Idea:

By conditioning, we can control for the attributes we want to add or remove

Results

Methodology Pipeline

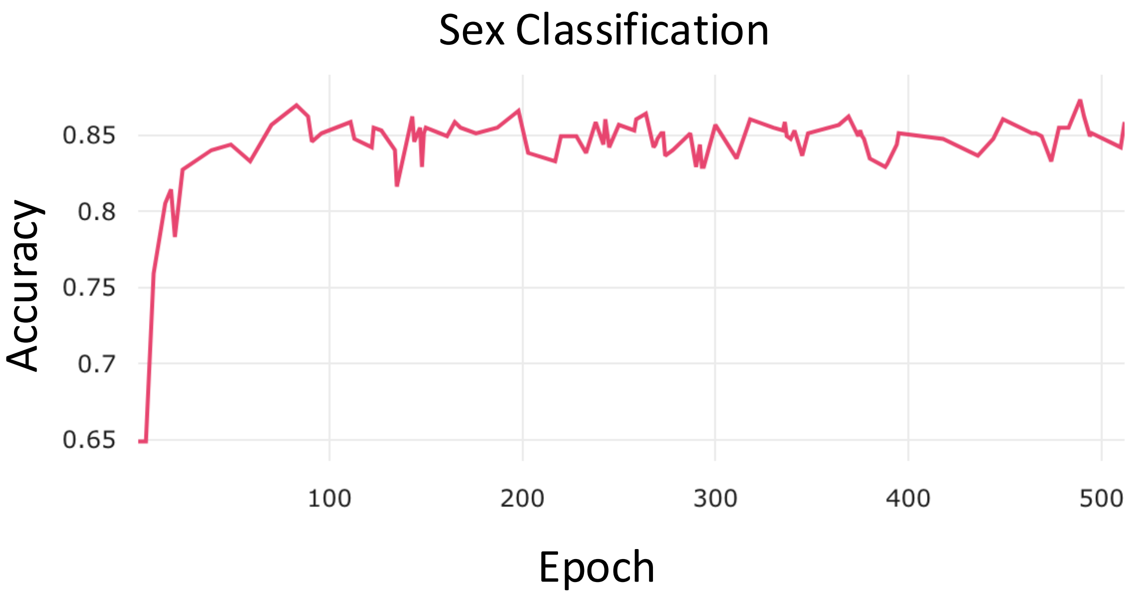
PHASE 1
Train a 2D StyleGAN

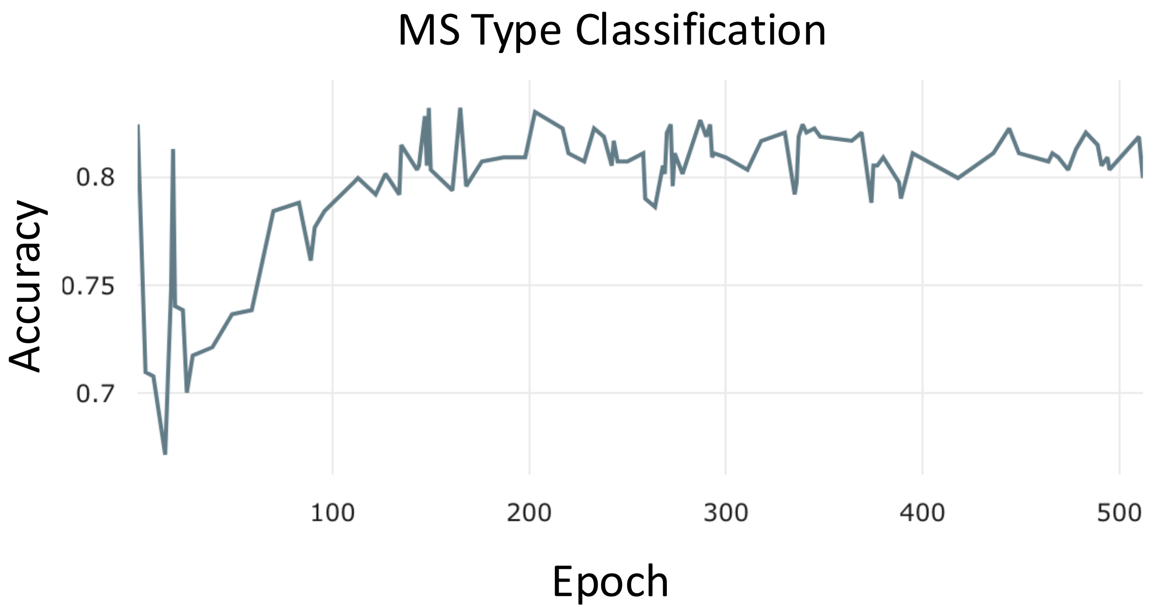
PHASE 2
Measure Disentanglement & Compare to GAN

PHASE 3
Condition our Model

PHASE 4
Invert Images and modify what you want

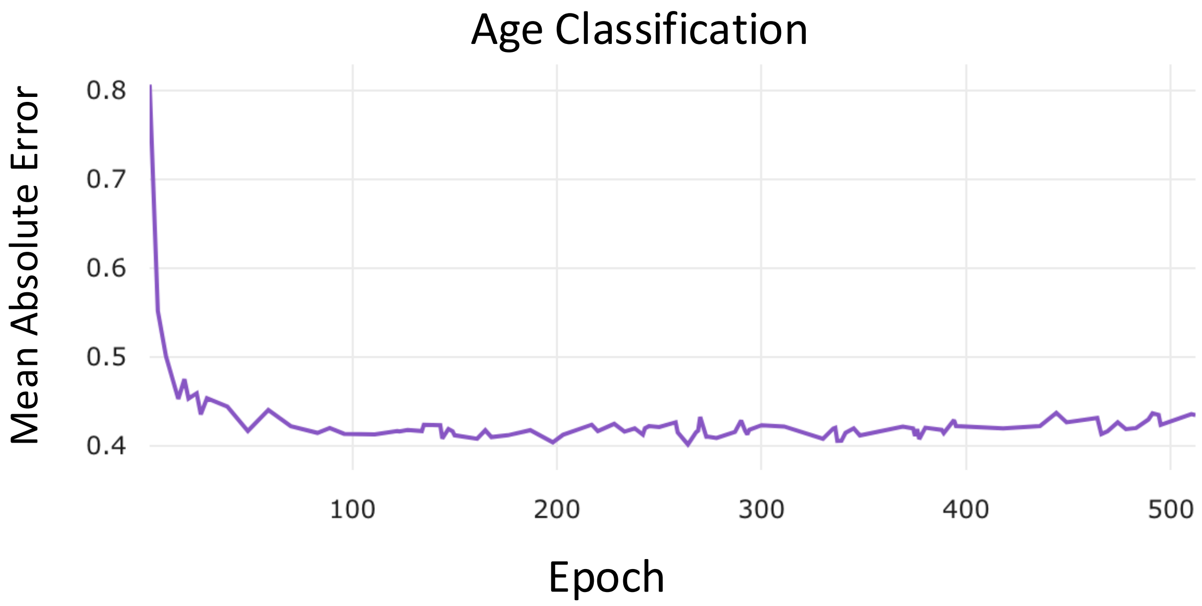
PPL Metric: Training of VGG-16 Model

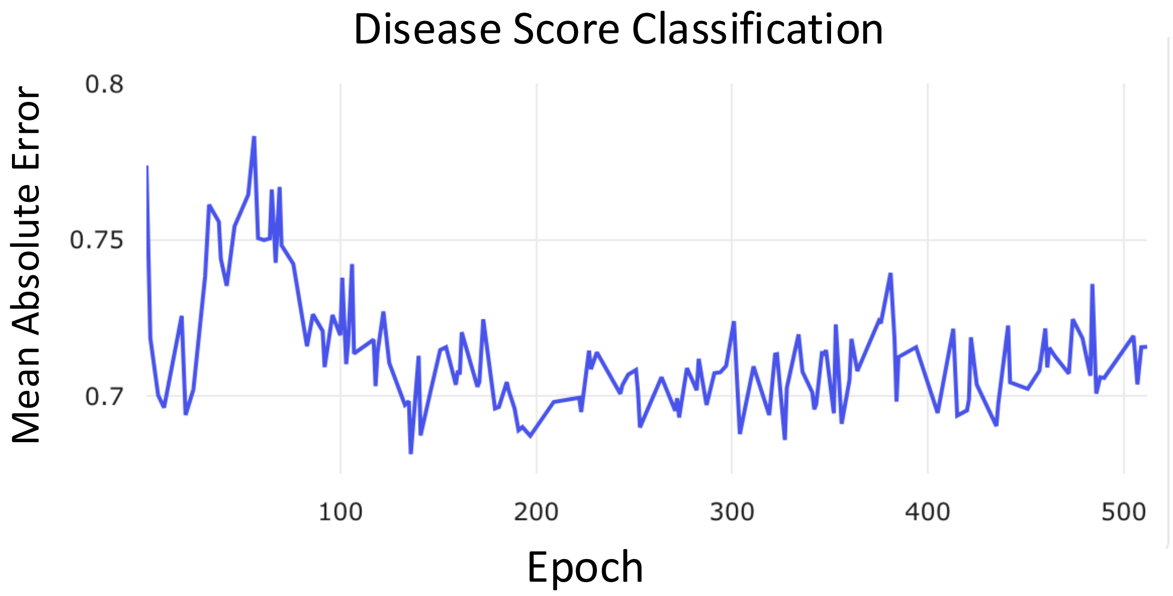




*Reminder: VGG-16 isn’t being used to get the most accurate classifier, but rather a feature space corresponding to human perception

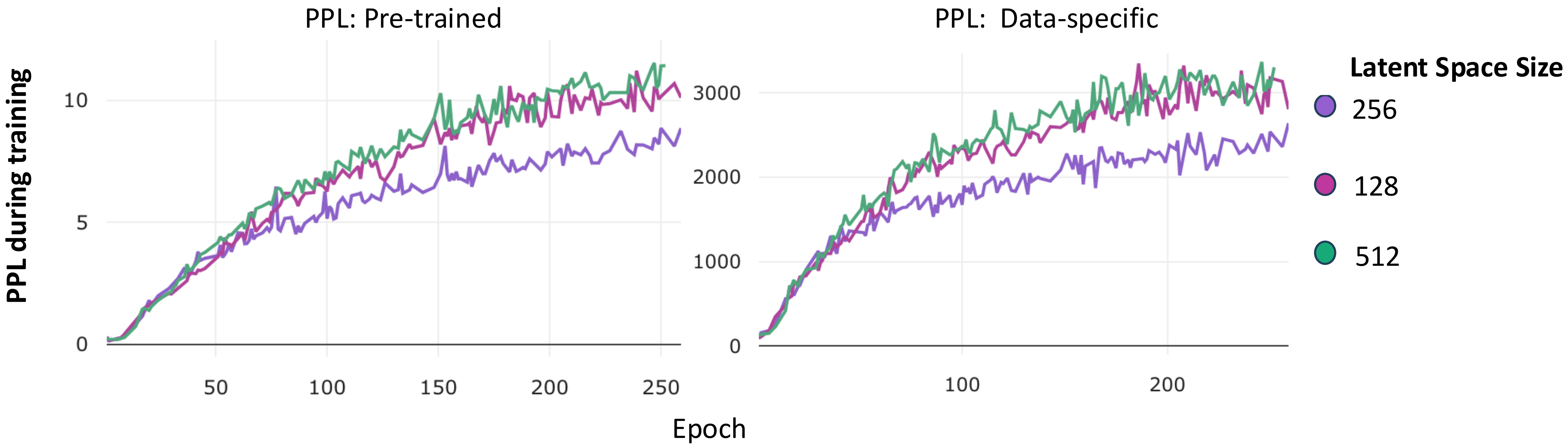
PPL Metric: Training of VGG-16 Model



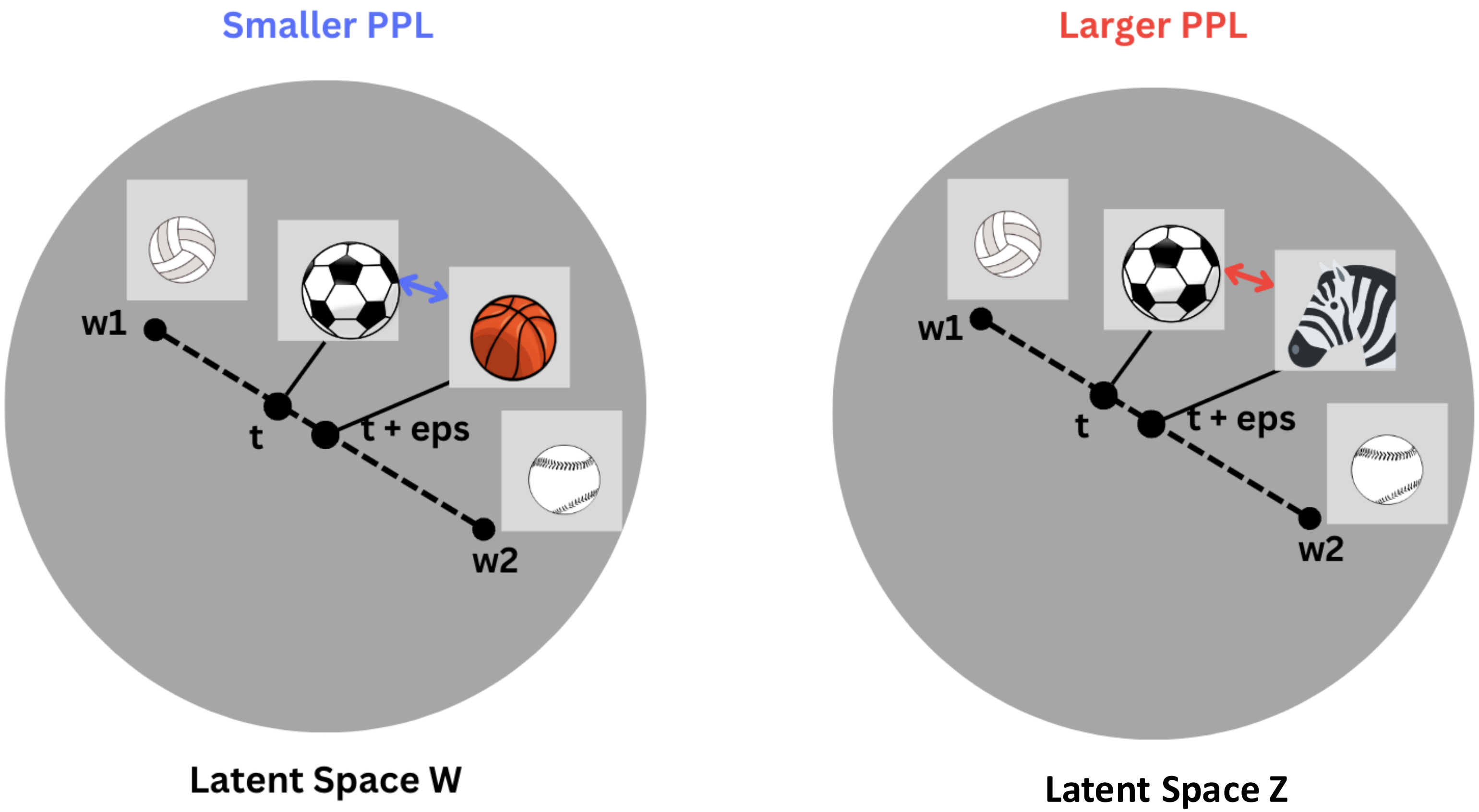


*Reminder: VGG-16 isn’t being used to get the most accurate classifier, but rather a feature space corresponding to human perception

PPL Metric: Comparison of our model with pre-trained VGG-16

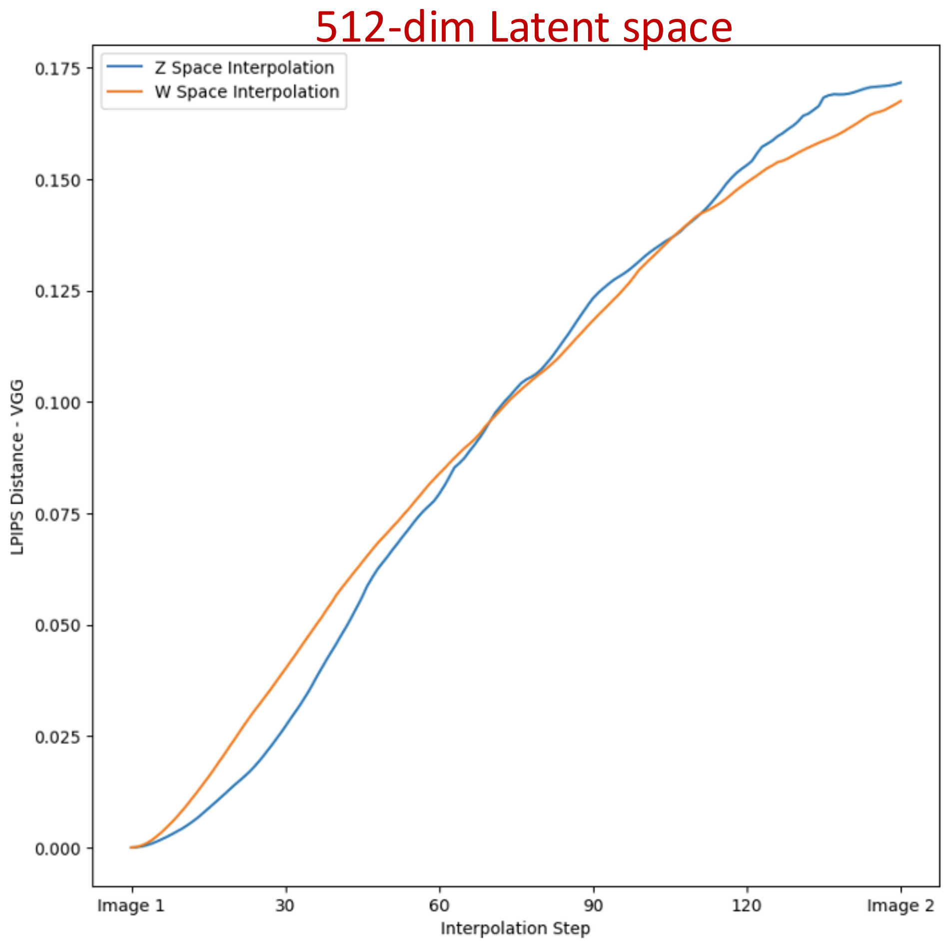


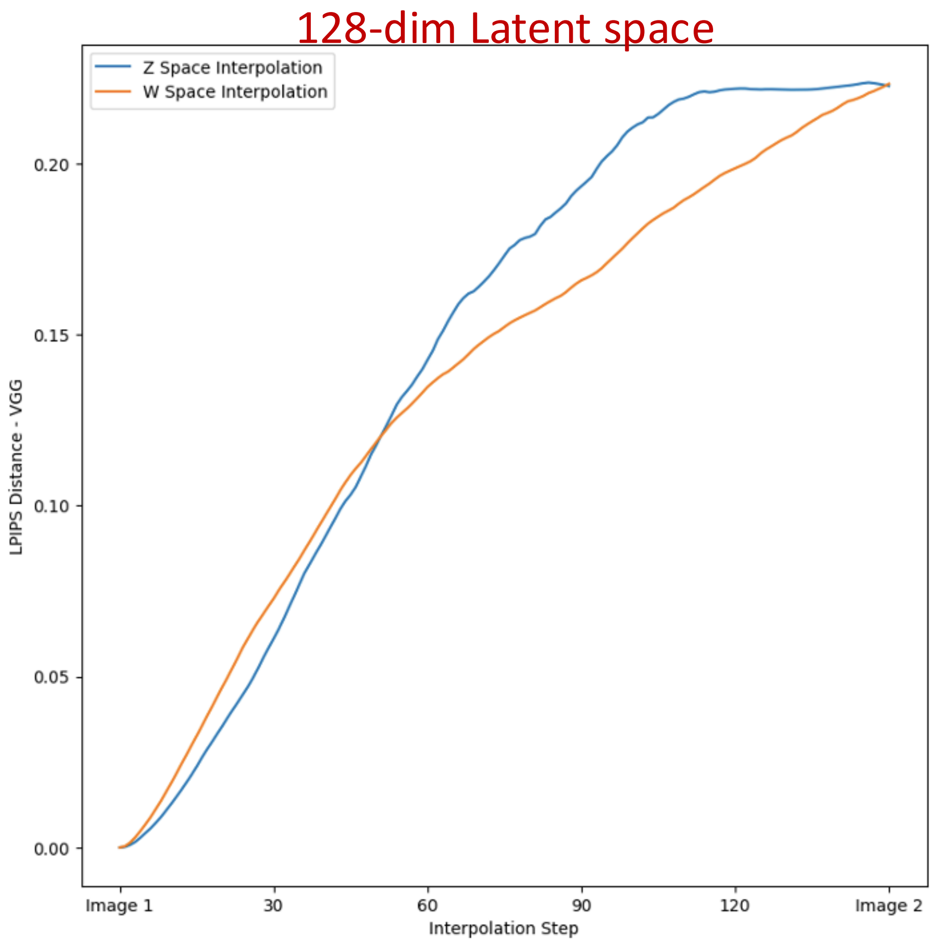
StyleGAN Interpolations



- Intuitively, we expect that if we **linearly** interpolate in Z-space (Gaussian!), changes in the image space will **not** be smooth
- If overparameterized (too big latent space), model can learn smooth transitions
- Ideally, smaller latent space would mean more precise conditioning

StyleGAN Interpolations

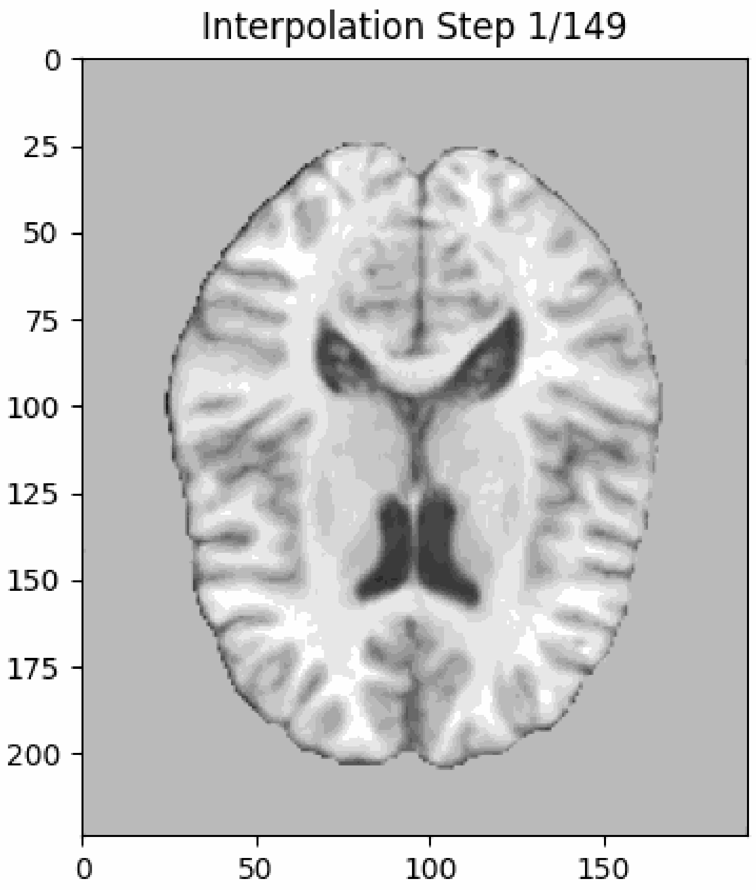




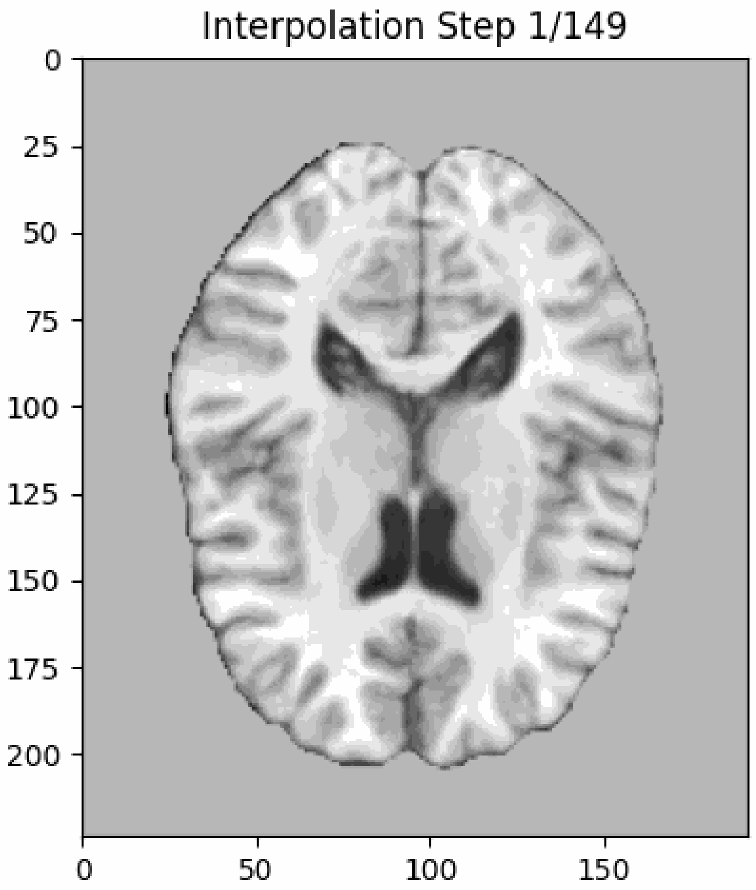
- Intuitively, we expect that if we **linearly** interpolate in Z-space (Gaussian!), changes in the image space will **not** be smooth
- If overparameterized, model can learn smooth transitions

StyleGAN Interpolations: 512-dim latent

W SPACE

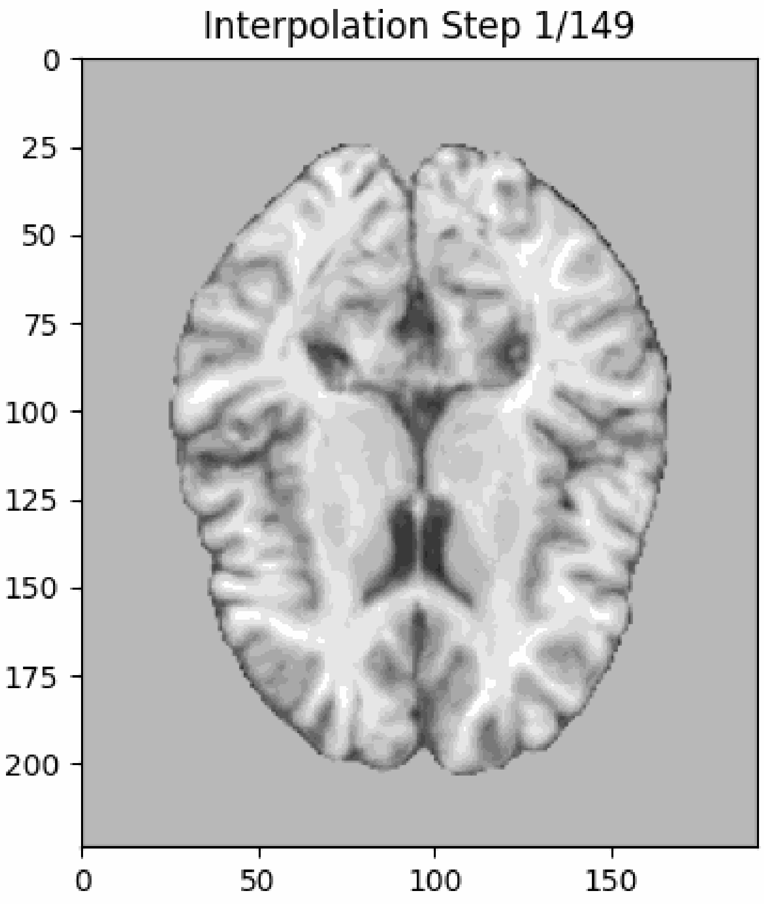


Z SPACE

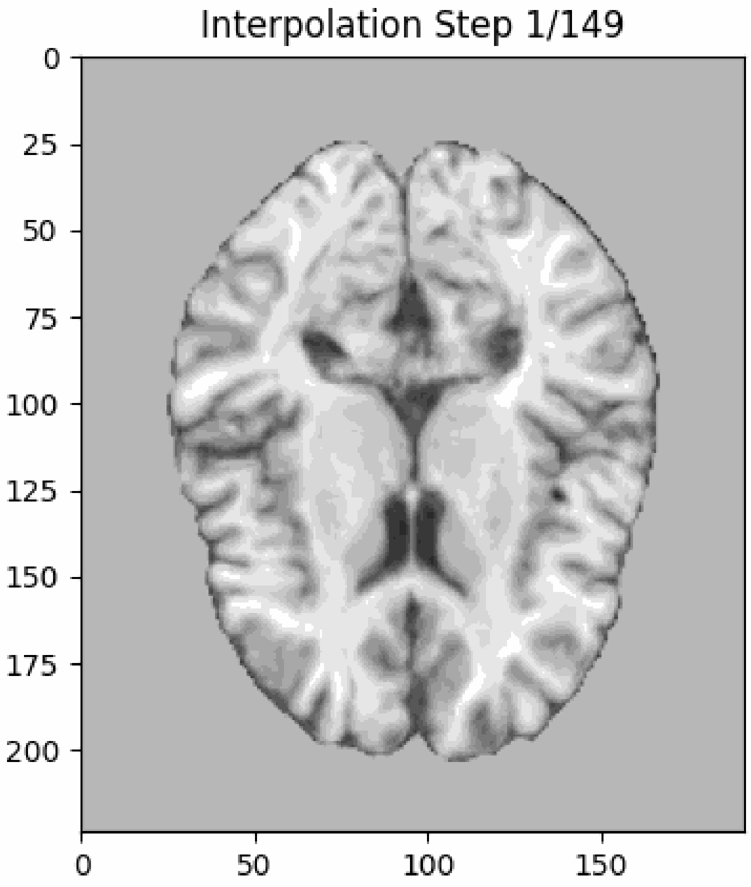


StyleGAN Interpolations: 128 dim latent

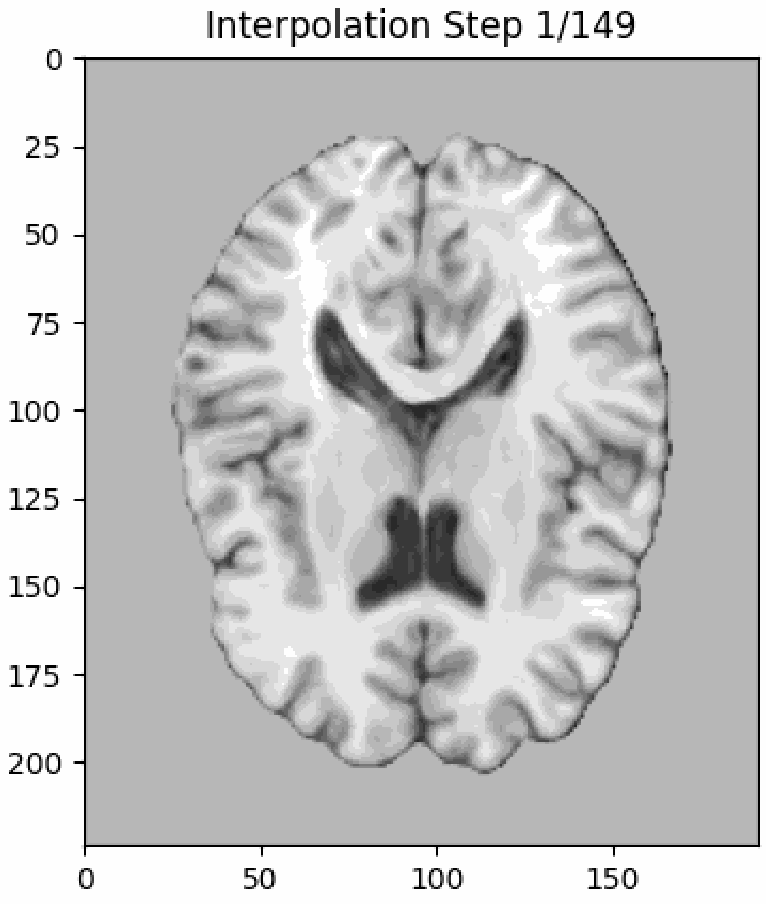
W SPACE

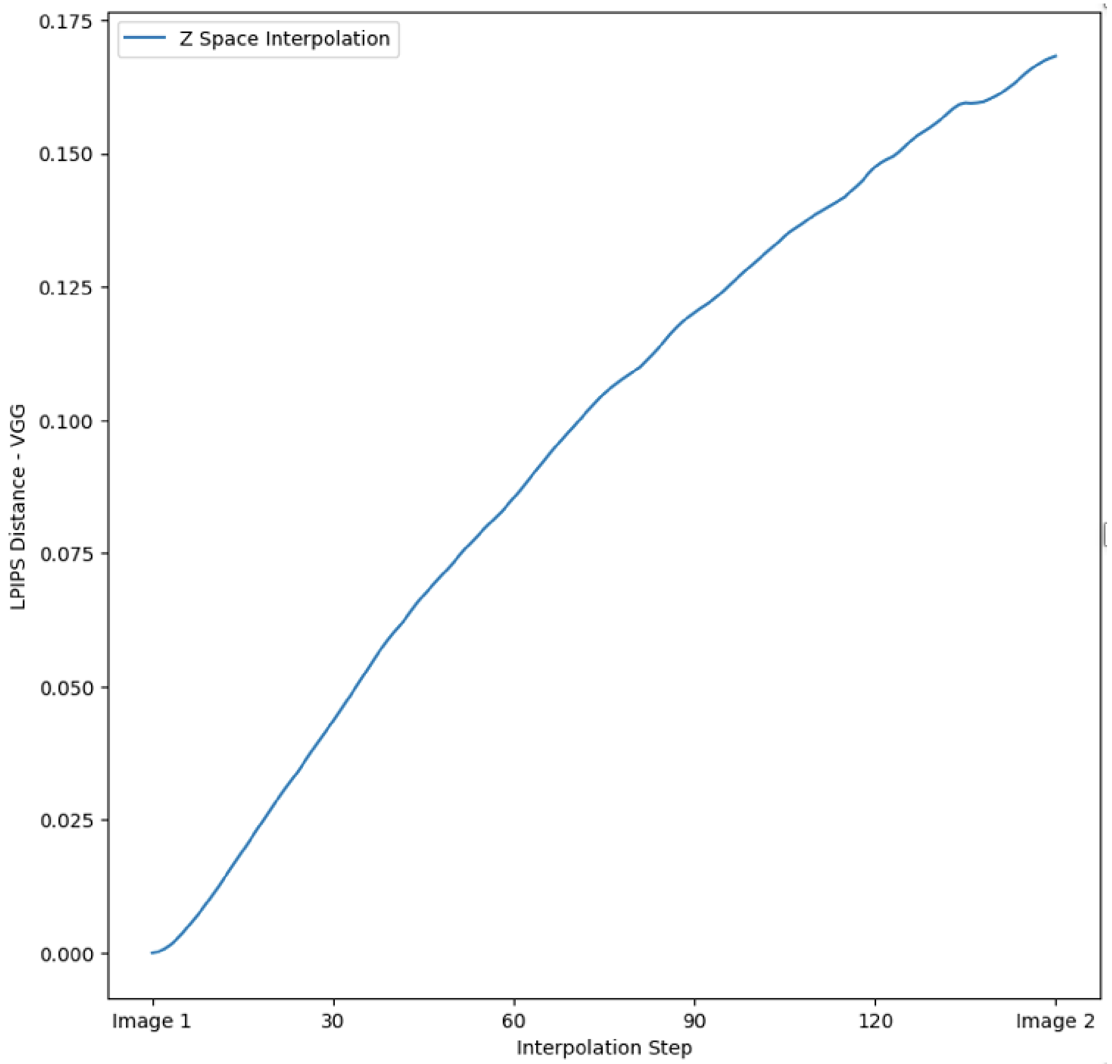


Z SPACE



GAN Interpolations: 256-dim latent





Future Work

Future Work

- Try smaller latent space sizes…currently overparameterized
- Add linear separability metric (amongst others)
- Evaluate metrics for StyleGAN vs GAN architecture
- Add conditioning and see if disentanglement improves
- Perform inversions…