

On the "steerability" of generative adversarial networks¹

February 26, 2020

¹ "On the "steerability" of generative adversarial networks" by Jahanian*, Chai*, and Isola

What this paper is about?

What this paper is about?

- ▶ It is believed that neural networks do not generalize “outside” of a dataset

What this paper is about?

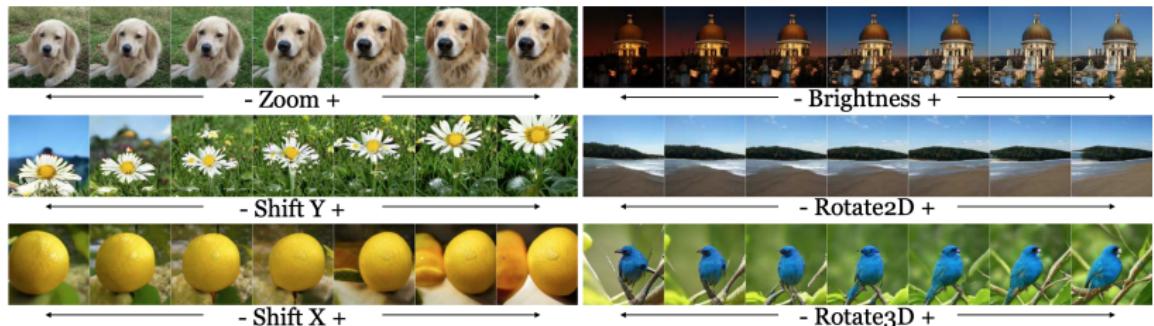
- ▶ It is believed that neural networks do not generalize “outside” of a dataset
- ▶ Do GANs suffer from the same issue?

What this paper is about?

- ▶ It is believed that neural networks do not generalize “outside” of a dataset
- ▶ Do GANs suffer from the same issue?
- ▶ Yes :(

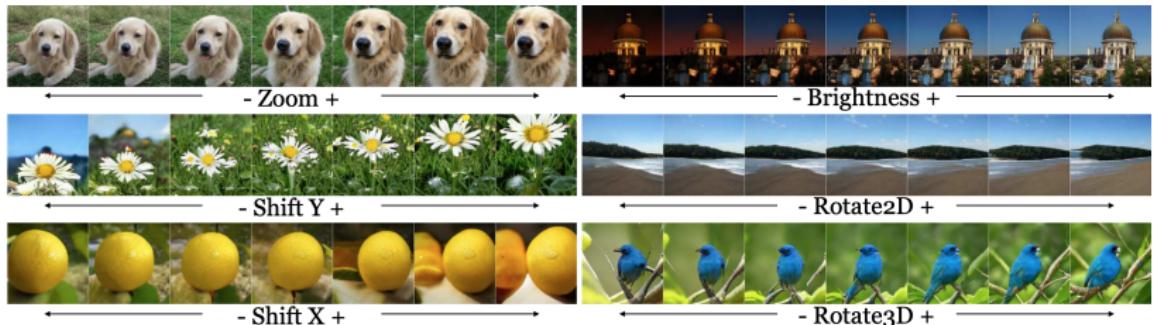
What this paper is about?

- ▶ It is believed that neural networks do not generalize “outside” of a dataset
- ▶ Do GANs suffer from the same issue?
- ▶ Yes :(
- ▶ But we can zoom/rotate/shift generated images by simple manipulations in latent space



What this paper is about?

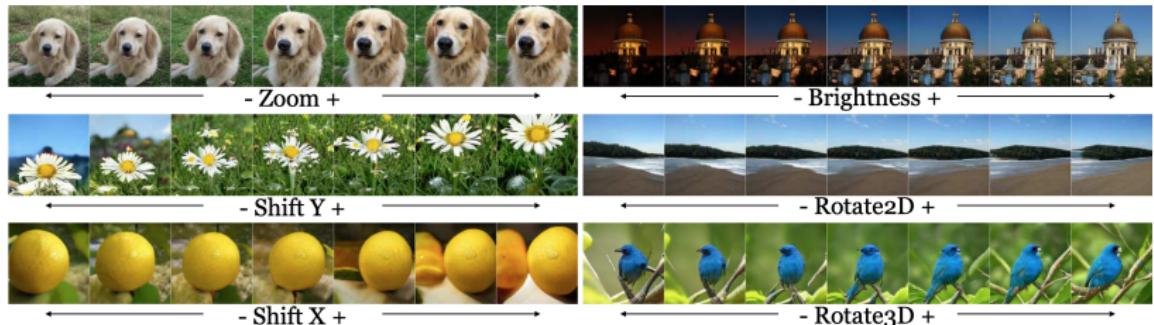
- ▶ It is believed that neural networks do not generalize “outside” of a dataset
- ▶ Do GANs suffer from the same issue?
- ▶ Yes :(
- ▶ But we can zoom/rotate/shift generated images by simple manipulations in latent space



- ▶ We can also add an additional loss term to improve *steerability*

What this paper is about?

- ▶ It is believed that neural networks do not generalize “outside” of a dataset
- ▶ Do GANs suffer from the same issue?
- ▶ Yes :(
- ▶ But we can zoom/rotate/shift generated images by simple manipulations in latent space

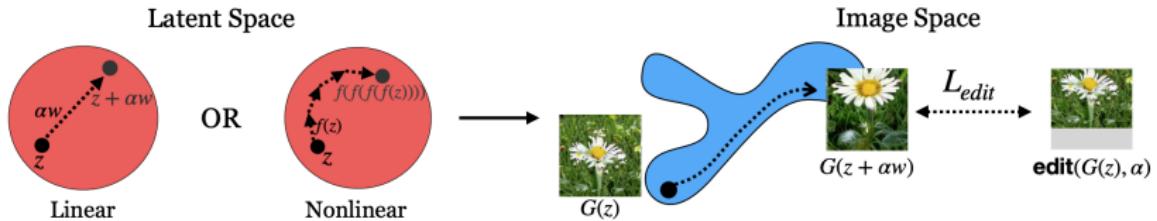


- ▶ We can also add an additional loss term to improve *steerability*
- ▶ This all shows us that GANs have some extrapolation abilities (but not a lot)

What is steerability?

What is steerability?

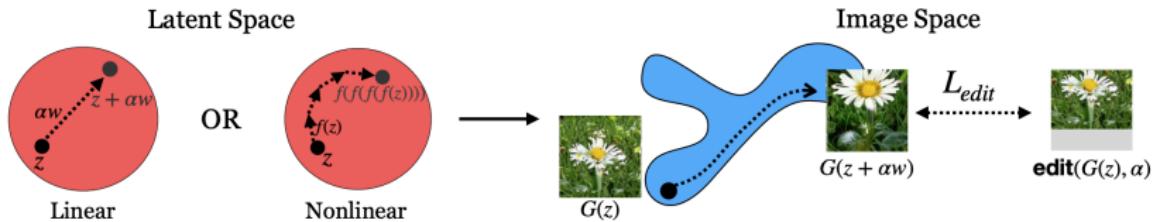
GAN model is *steerable* if we can change a specific property of an image by “walking” in a latent space.



On the image above:

What is steerability?

GAN model is *steerable* if we can change a specific property of an image by “walking” in a latent space.

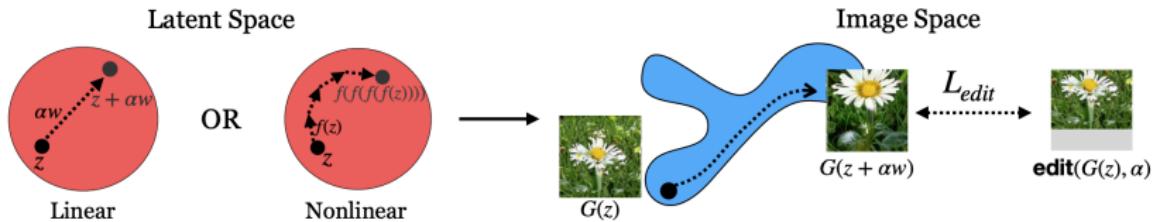


On the image above:

- ▶ We move via a linear or a nonlinear path in the latent space

What is steerability?

GAN model is *steerable* if we can change a specific property of an image by “walking” in a latent space.

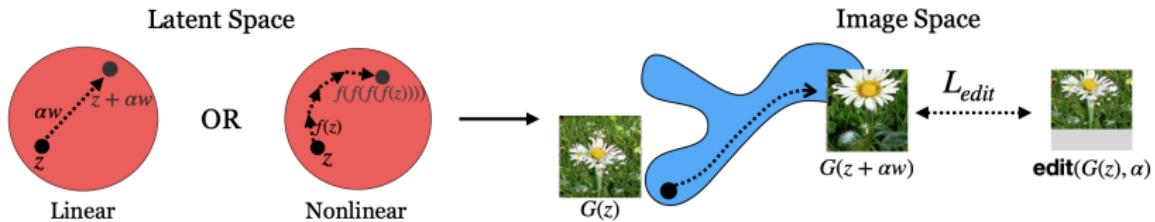


On the image above:

- ▶ We move via a linear or a nonlinear path in the latent space
- ▶ The daisy image shifts up in the image space

What is steerability?

GAN model is *steerable* if we can change a specific property of an image by “walking” in a latent space.



On the image above:

- ▶ We move via a linear or a nonlinear path in the latent space
- ▶ The daisy image shifts up in the image space
- ▶ (We can also construct a supervised signal by cropping and shifting the image, like in the rightmost image)

How to construct a path in a latent space?

How do we build a linear path in a latent space?

²Instead of using L2 loss one can use perceptual distance, but this does not change things much

How to construct a path in a latent space?

How do we build a linear path in a latent space?

1. Imagine that we want to find a path for zooming

²Instead of using L2 loss one can use perceptual distance, but this does not change things much

How to construct a path in a latent space?

How do we build a linear path in a latent space?

1. Imagine that we want to find a path for zooming
2. Generate a lot of synthetic images

²Instead of using L2 loss one can use perceptual distance, but this does not change things much

How to construct a path in a latent space?

How do we build a linear path in a latent space?

1. Imagine that we want to find a path for zooming
2. Generate a lot of synthetic images
3. Edit these synthetic images by zooming-in and zooming-out with different strengths

²Instead of using L2 loss one can use perceptual distance, but this does not change things much

How to construct a path in a latent space?

How do we build a linear path in a latent space?

1. Imagine that we want to find a path for zooming
2. Generate a lot of synthetic images
3. Edit these synthetic images by zooming-in and zooming-out with different strengths
4. Find optimal path direction w by optimizing²:

$$w^* = \arg \min_w \mathbb{E}_{z, \alpha} [\|G(z + \alpha w) - \text{edit}(G(z), \alpha)\|_2^2] \quad (1)$$

Here:

- ▶ $\text{edit}(G(z), \alpha)$ is the edited image
- ▶ α is a strength parameter

²Instead of using L2 loss one can use perceptual distance, but this does not change things much

How to construct a path in a latent space?

How do we build a non-linear path in a latent space?

How to construct a path in a latent space?

How do we build a non-linear path in a latent space?

1. We split a path into n pieces of size ϵ

How to construct a path in a latent space?

How do we build a non-linear path in a latent space?

1. We split a path into n pieces of size ϵ
2. We want to find a function f_θ such that applying it n times will zoom an image in $n \cdot \epsilon$ times

How to construct a path in a latent space?

How do we build a non-linear path in a latent space?

1. We split a path into n pieces of size ϵ
2. We want to find a function f_θ such that applying it n times will zoom an image in $n \cdot \epsilon$ times
3. We find it by optimizing:

$$\theta^* = \arg \min_{\theta} \mathbb{E}_{z,n} \left[\| G(\underbrace{f(f(\dots f(z)))}_{n \text{ times}}) - \text{edit}(G(z), n \cdot \epsilon) \|_2^2 \right]$$

Authors parametrize f as a neural network.

How to construct a path in a latent space?

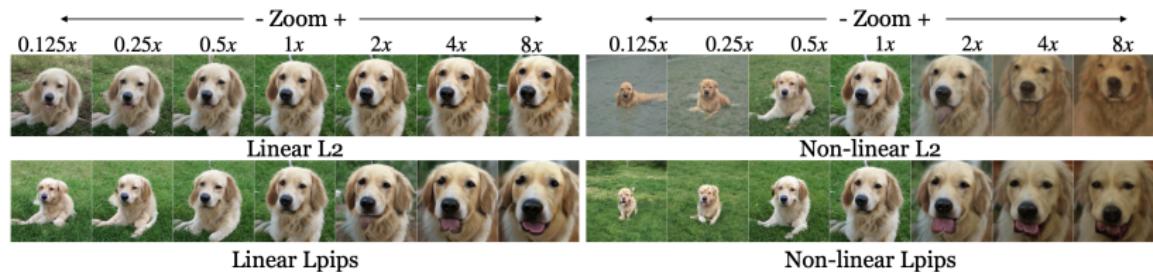
How do we build a non-linear path in a latent space?

1. We split a path into n pieces of size ϵ
2. We want to find a function f_θ such that applying it n times will zoom an image in $n \cdot \epsilon$ times
3. We find it by optimizing:

$$\theta^* = \arg \min_{\theta} \mathbb{E}_{z,n} \left[\| G(\underbrace{f(f(\dots f(z)))}_{n \text{ times}}) - \text{edit}(G(z), n \cdot \epsilon) \|_2^2 \right]$$

Authors parametrize f as a neural network.

But it does not work much better than a linear path:



Can we improve steerability?

³Authors do not elaborate if D is also trained with the modified loss

Can we improve steerability?

Yes, by adding an additional regularization to Generator's objective:

$$G^*, w^* = \arg \min_{G,w} (\mathcal{L}_{edit} + \mathcal{L}_{GAN}) \quad (2)$$

Where:

$$\mathcal{L}_{edit} = L2(G(z + \alpha w) - \text{edit}(G(z), \alpha)) \quad (3)$$

And GAN loss is ³:

$$\mathcal{L}_{GAN} = \max_D (\mathbb{E}_{z,\alpha}[D(G(z + \alpha w))] - \mathbb{E}_{x,\alpha}[D(\text{edit}(x, \alpha))]) \quad (4)$$

³Authors do not elaborate if D is also trained with the modified loss

Steering has limits

There 2 types of problems we face when steer further:

Steering has limits

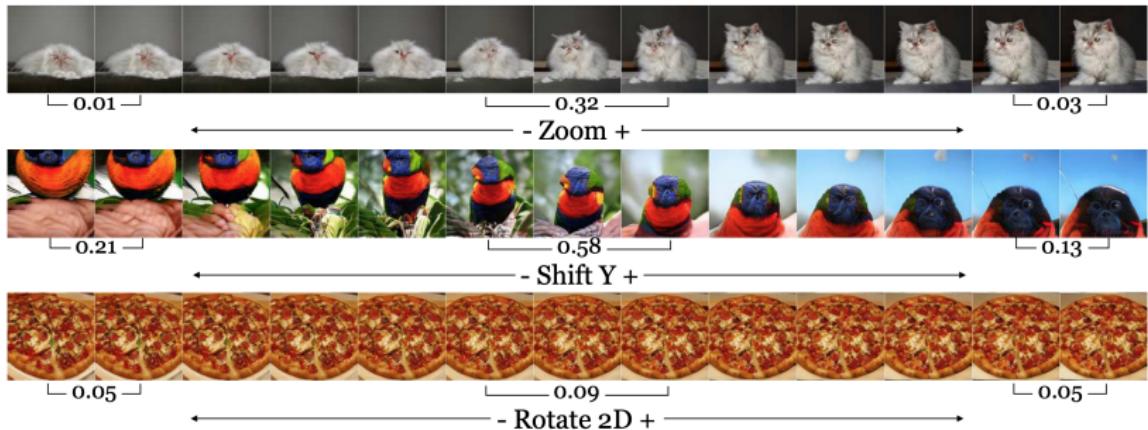
There 2 types of problems we face when steer further:

- ▶ Images become unrealistic

Steering has limits

There 2 types of problems we face when steer further:

- ▶ Images become unrealistic
- ▶ Images stop changing anymore



Numbers in the image are perceptual distances between images, showing that transformation converged.

Conclusion

Conclusion

- ▶ Authors test steerability for several existing GAN models (BigGAN, StyleGAN, etc)

Conclusion

- ▶ Authors test steerability for several existing GAN models (BigGAN, StyleGAN, etc)
- ▶ Authors proposed a way to quantify steerability (omitted in the presentation)

Conclusion

- ▶ Authors test steerability for several existing GAN models (BigGAN, StyleGAN, etc)
- ▶ Authors proposed a way to quantify steerability (omitted in the presentation)
- ▶ We need better editions for shifting/rotating/etc to find better paths

Conclusion

- ▶ Authors test steerability for several existing GAN models (BigGAN, StyleGAN, etc)
- ▶ Authors proposed a way to quantify steerability (omitted in the presentation)
- ▶ We need better editions for shifting/rotating/etc to find better paths
- ▶ Variability in data allows us to build more steerable models