On the “steerability” of generative adversarial networks

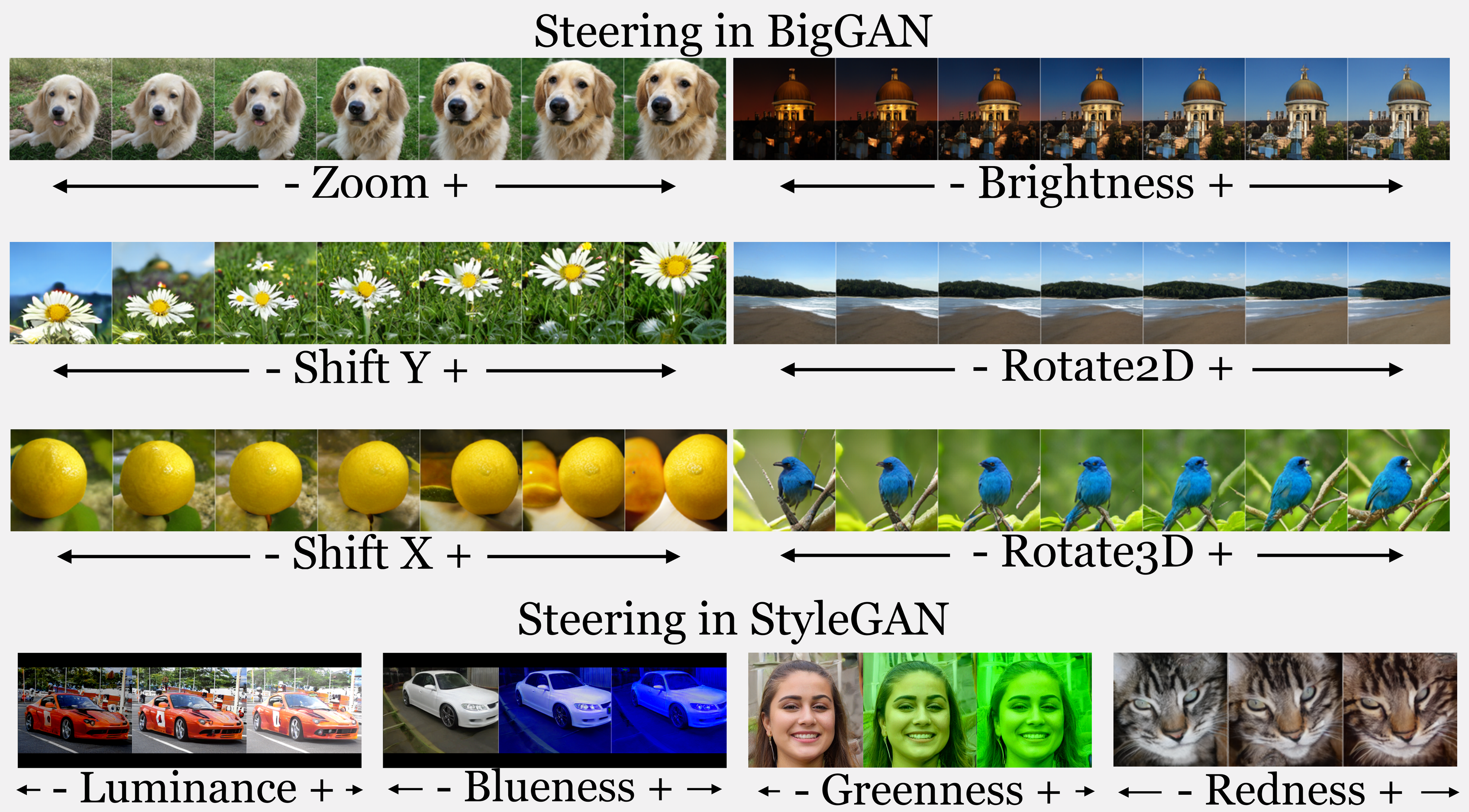
Massachusetts Institute of Technology

Ali Jahanian*, Lucy Chai*, Phillip Isola

ali-design@csail.mit.edu, lrchai@mit.edu, phillipi@mit.edu

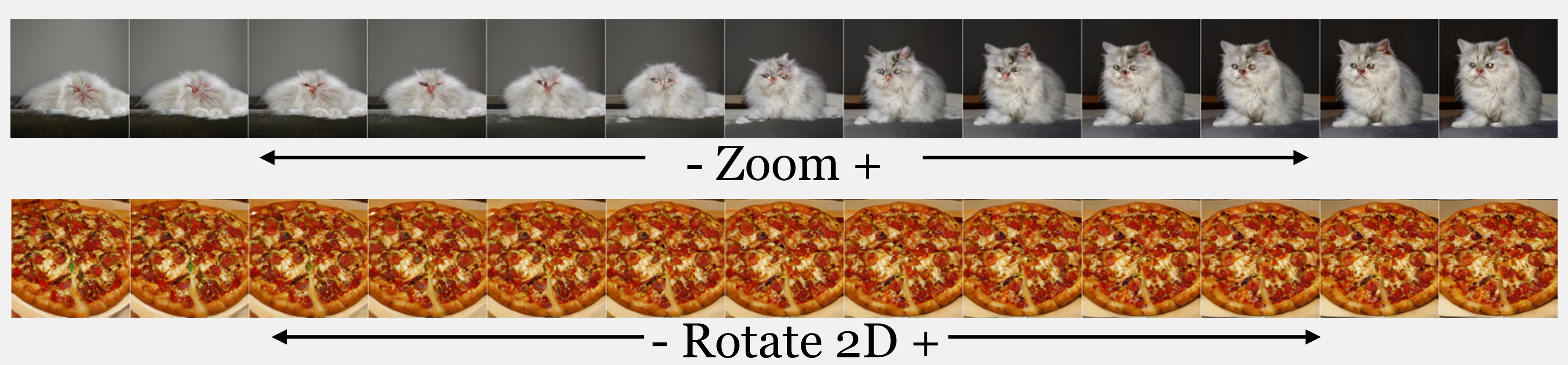
https://github.com/ali-design/gan_steerability

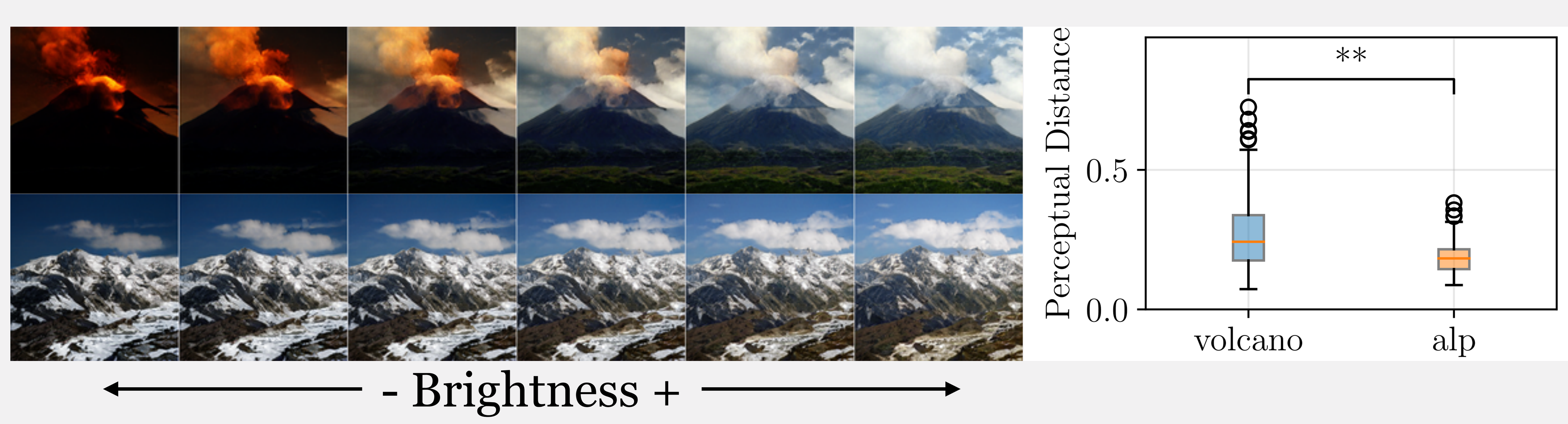
Can you “steer” a GAN in latent space?



- latent space trajectories in GANs correspond to simple image transformations
- following a trajectory changes the generated image distribution
- dataset biases limit the extent of transformations

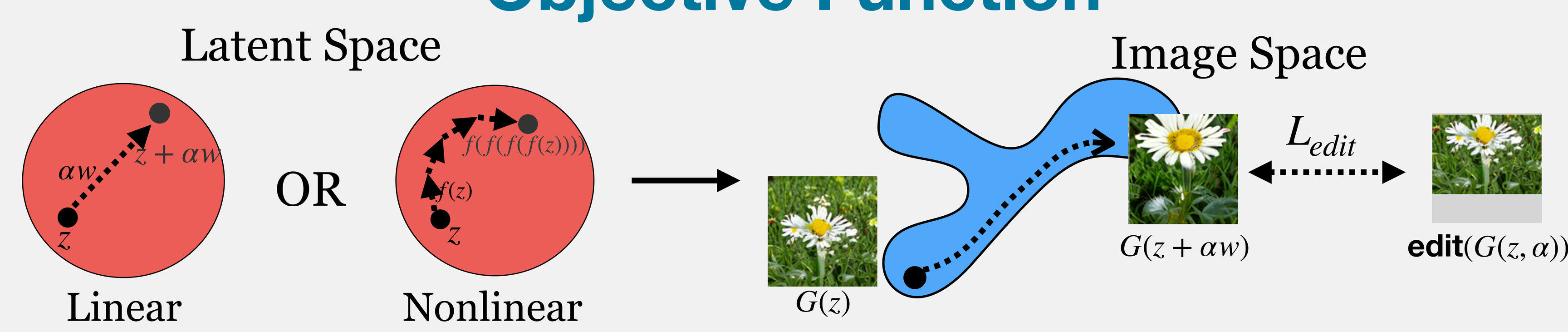
Limits of Latent Space Transformations





- ability to achieve arbitrary transformations is finite
- same trajectory has a different effect on different image classes; volcanoes explode when darkened but mountains do not
- effect of transformation is consistent with class semantics

Objective Function



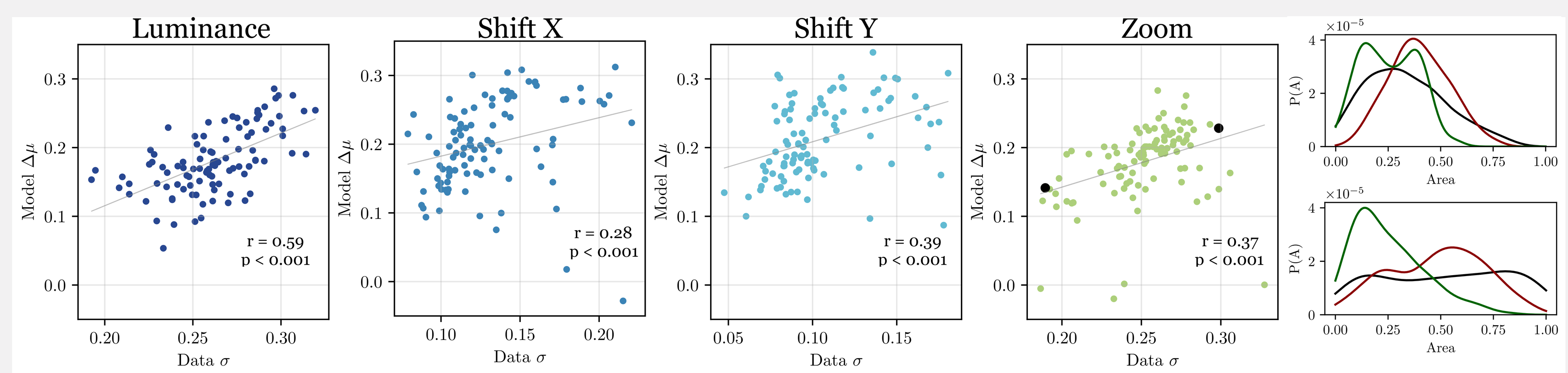
- latent space trajectories learned in a self-supervised manner
- L2 loss objective between applied latent walk and target transformation

$$w^* = \arg\min_{w} \mathbb{E}_{z,\alpha}\left[\mathcal{L}(G(z+\alpha w), \mathrm{edit}(G(z),\alpha))\right]$$

- alternatively instead of a linear walk, we can use a nonlinear one:

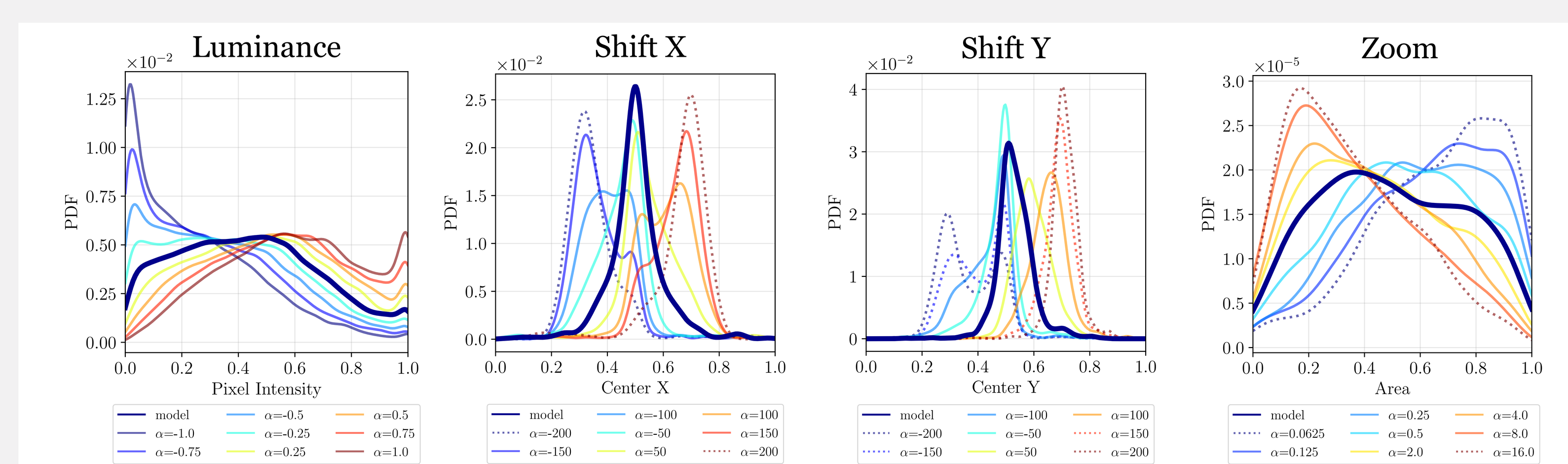
$$\mathcal{L} = \mathbb{E}_{z,n}\left[\,||(f^n(z) - \mathrm{edit}(G(z), n\epsilon))||\,\right]$$

Transformations Governed by Data Biases



- steerability of image classes is governed by dataset diversity
- we achieve greater degree of transformation when dataset spread is large

Shifting Attributes of Generated Distributions

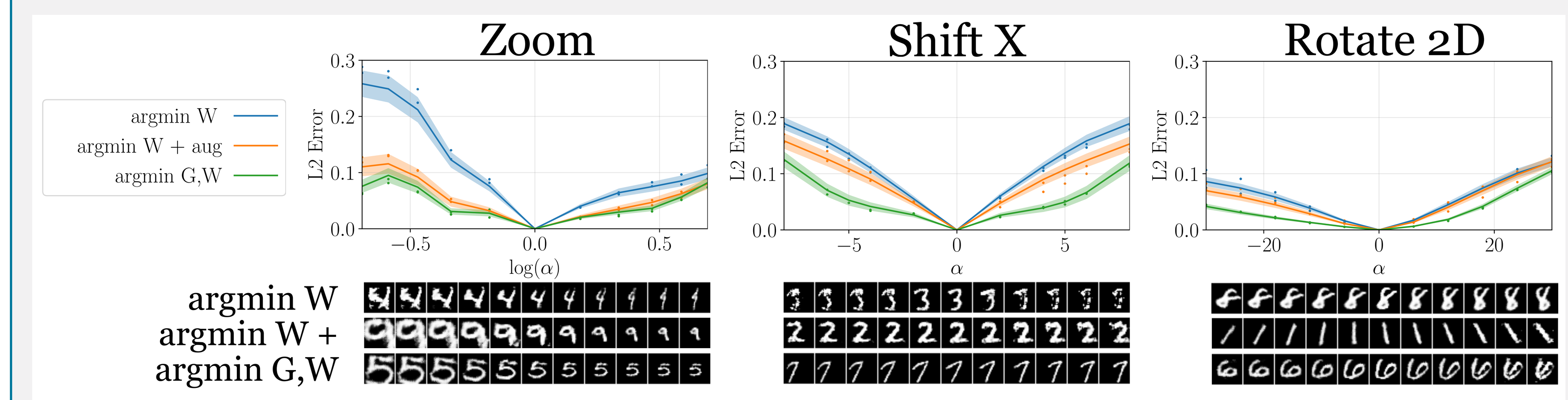


- latent space walks shift the generated image distribution
- generated images become unrealistic beyond a range of transformation

Steerable GANs

- instead of freezing the generator parameters when learning a latent space walk, we can jointly train the walk and the GAN

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



argmin W: Vanilla GANs w/o data augmentation
argmin W + aug: Vanilla GANs with data augmentation
argmin G,W: Jointly optimizing the generator and the walk parameters with data augmentation

- data augmentation and joint training of the generator and walk increases transformation range

Our Main Findings

1) A simple self-supervised walk in the latent space of GANs achieves camera motion and color transformations in the output image space.
2) The linear walk is as effective as more complex non-linear walks, without being explicitly trained for.
3) The extent of each transformation is limited, and it is a function of the dataset variability.
4) The transformations are a general-purpose and work with different architectures, e.g. BigGAN, StyleGAN, and DCGAN, and illustrate different disentanglement properties in their respective latent spaces.
5) Data augmentation and jointly training the walk trajectory and the generator weights improves steerability, resulting in larger transformation effects.

ACKNOWLEDGEMENTS

We would like to thank Quang H Le, Lore Goetschalckx, Alex Andonian, David Bau, and Jonas Wulff for helpful discussions. This work was supported by a Google Faculty Research Award to P.I., and a U.S. National Science Foundation Graduate Research Fellowship under Grant No. 1122374 to L.C.