Generative models as data visualization

Phillip Isola, MIT

Image and Video Synthesis Workshop, ICCV 2019

Work with Lore Goetschalckx, Alex Andonian, Ali Jahanian, Lucy Chai, Aude Oliva

What’s the use of a generative model?

- Prediction
- Representation
- Visualization

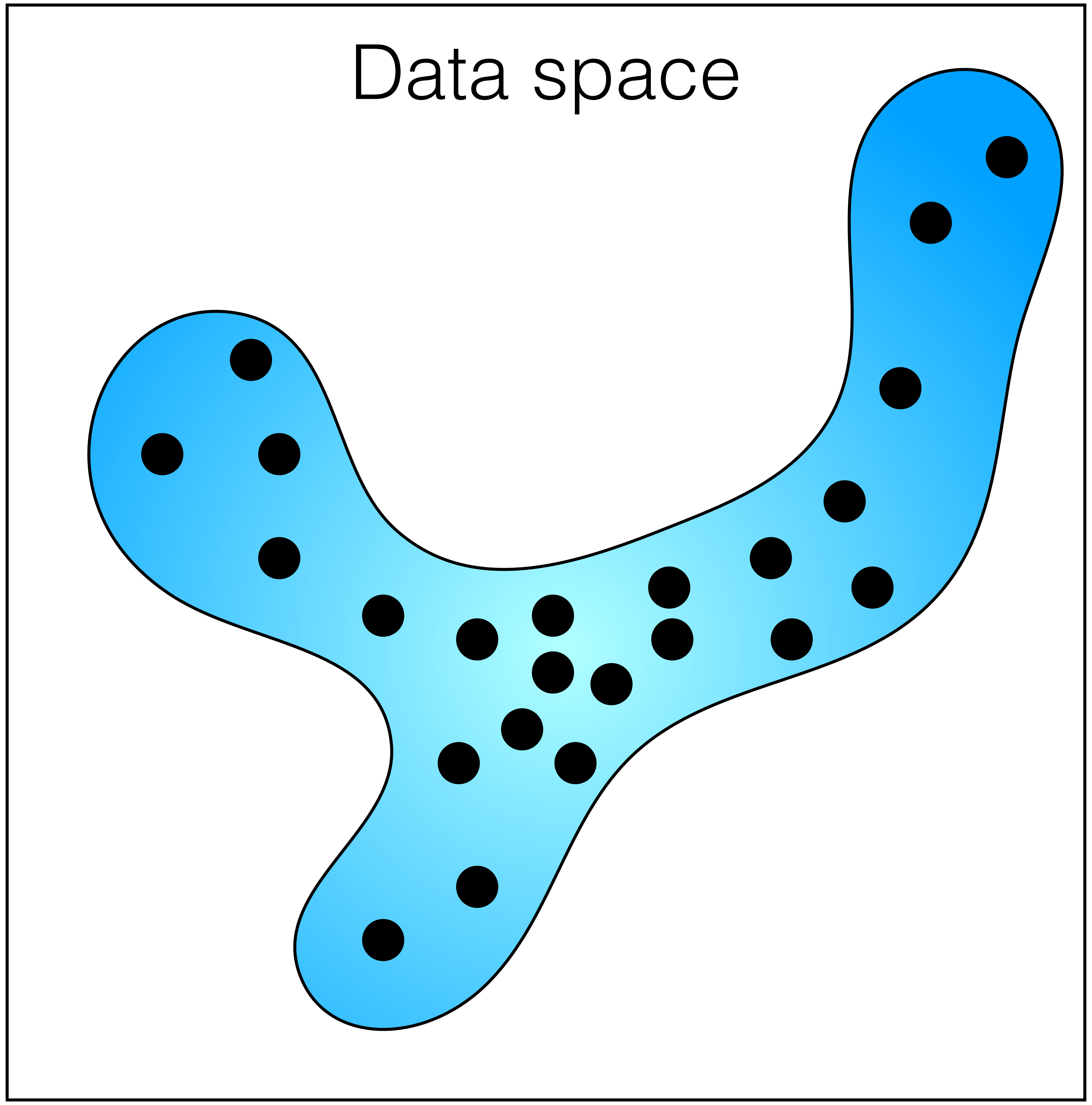
What's the use of a generative model?

- Prediction
- Representation
- **Visualization**

Generative models as data visualization

Data samples

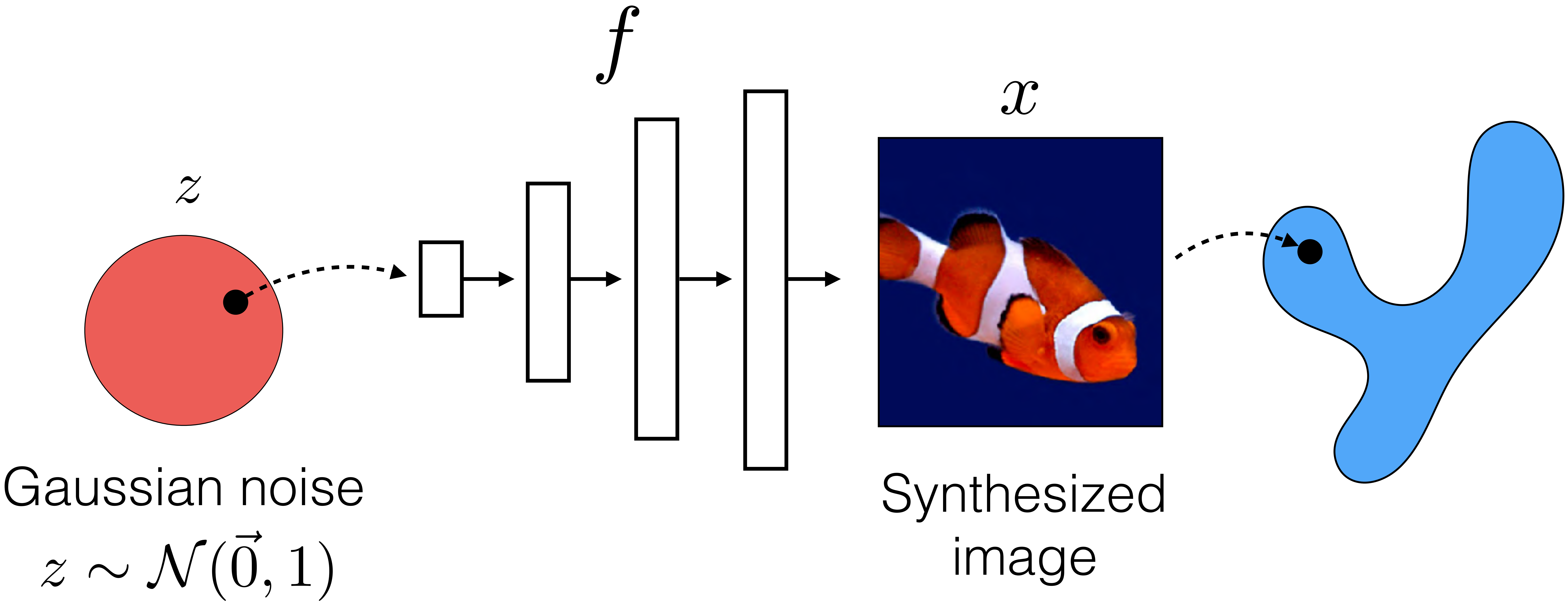
[FFHQ dataset, Karras et al. 2018]



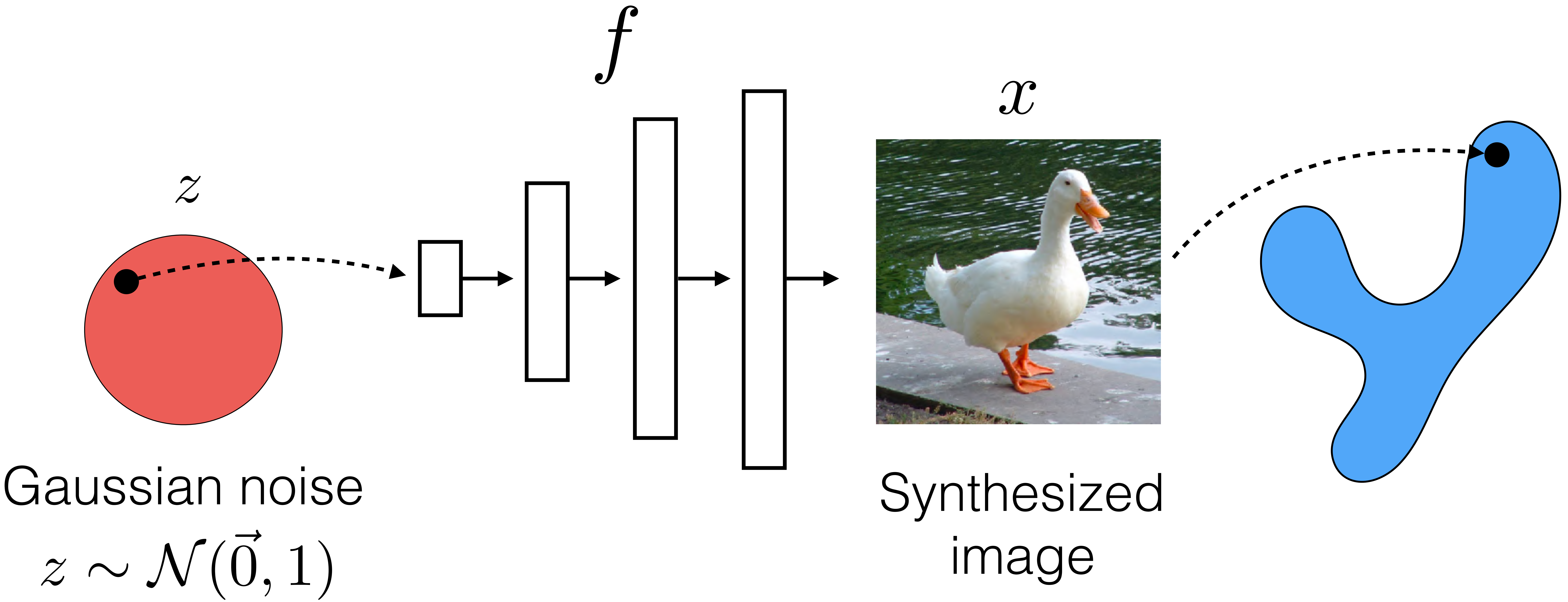
Model interpolations

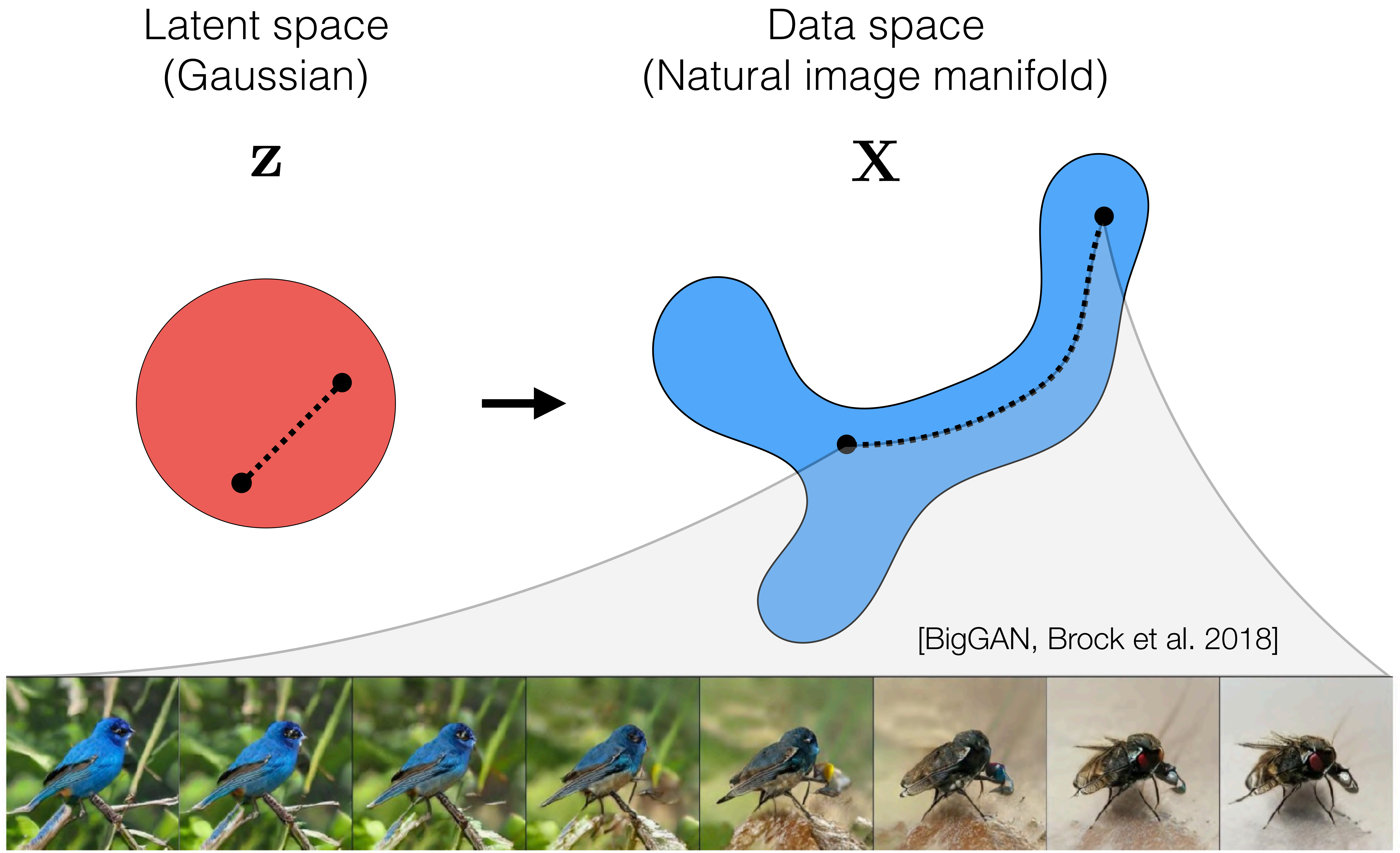
[StyleGAN, Karras et al. 2018]

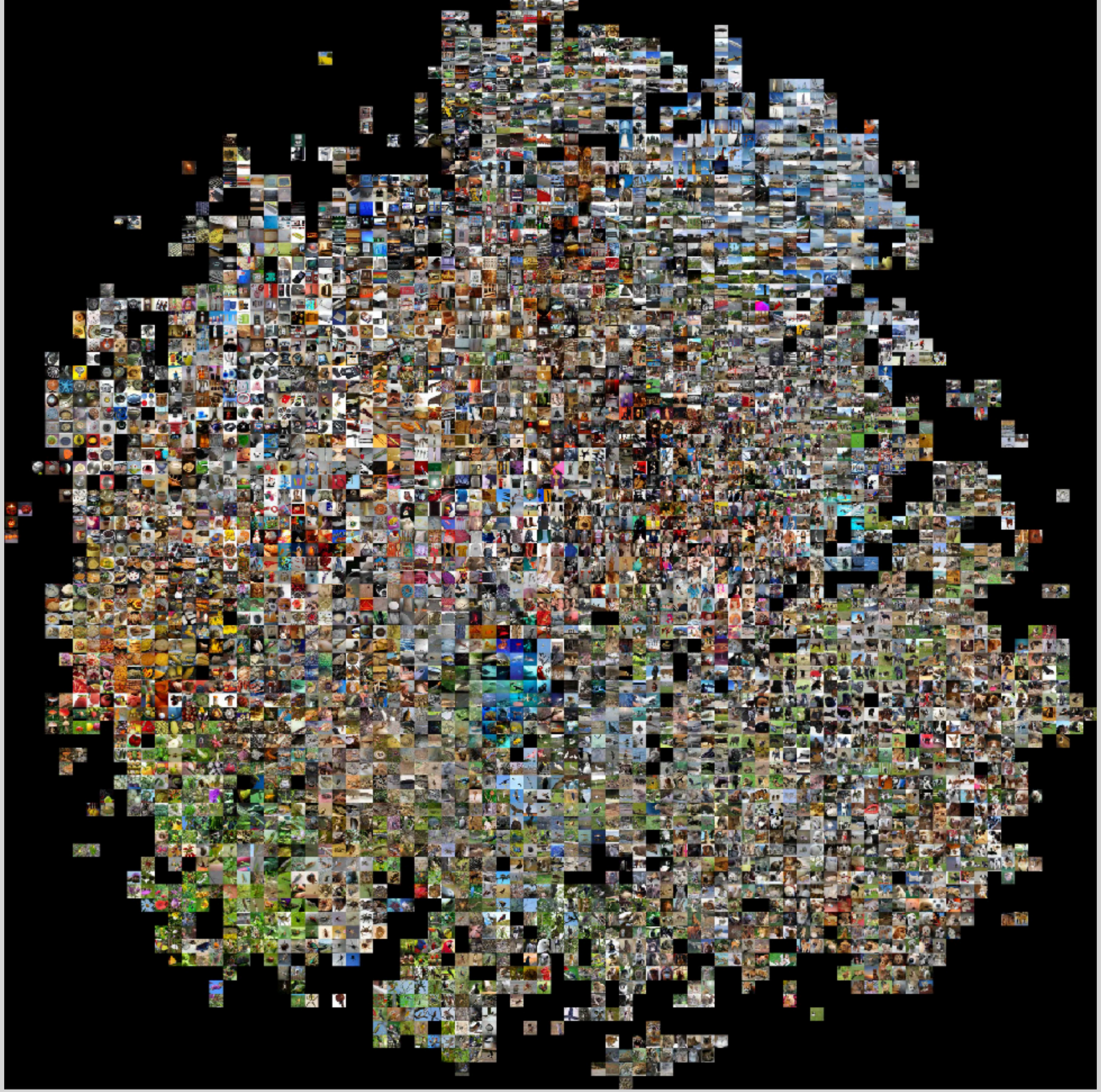
Deep generative models are distribution transformers



Deep generative models are distribution transformers

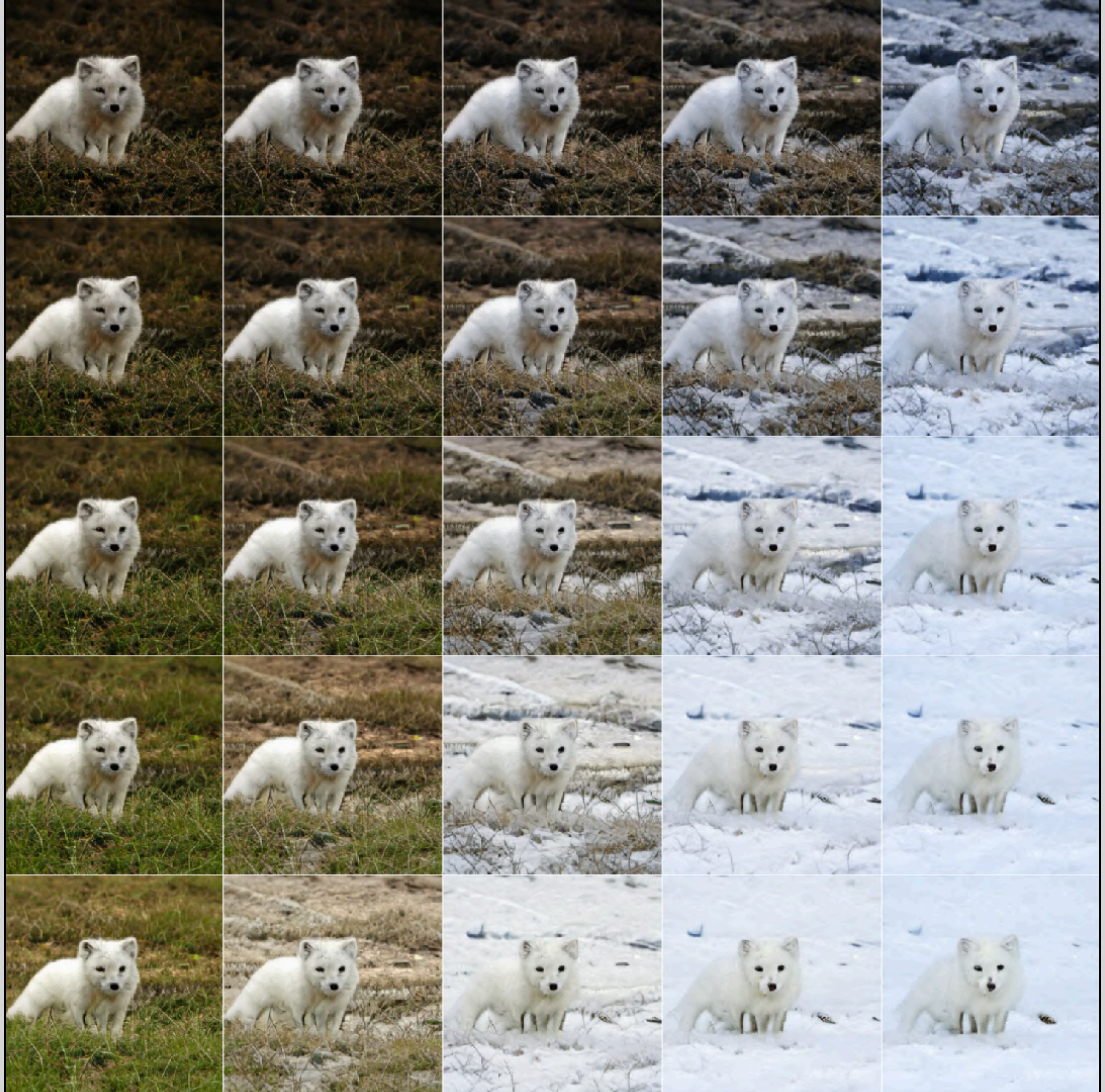



Latent space
(Gaussian)
$\mathbf{z}$
Data space
(Natural image manifold)
$\mathbf{x}$
[BigGAN, Brock et al. 2018]



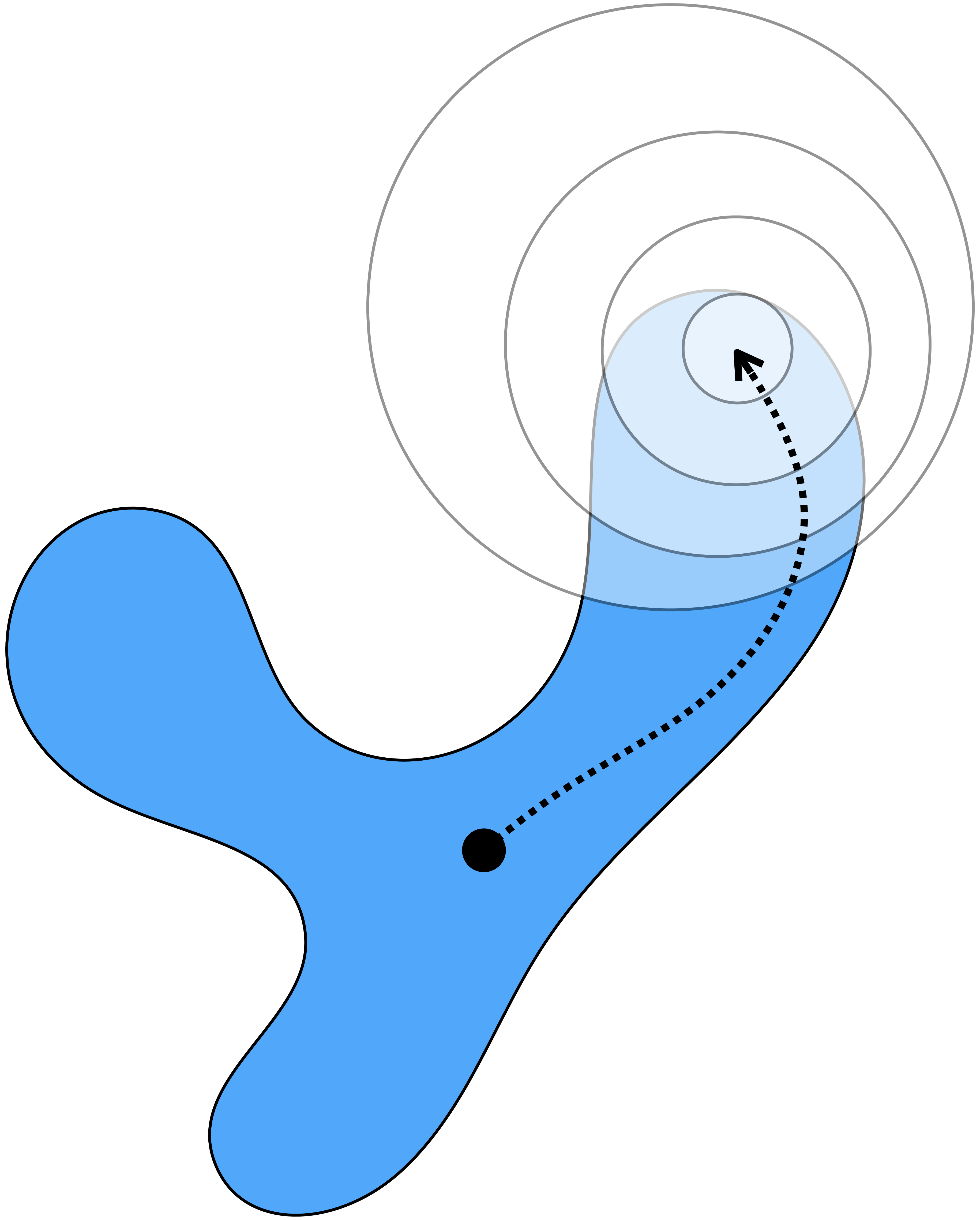
[https://cs.stanford.edu/people/karpathy/cnnembed/]

[https://cs.stanford.edu/people/karpathy/cnnembed/]



GANs let us explore the manifold of natural images

- What does it look like to make an image more white?



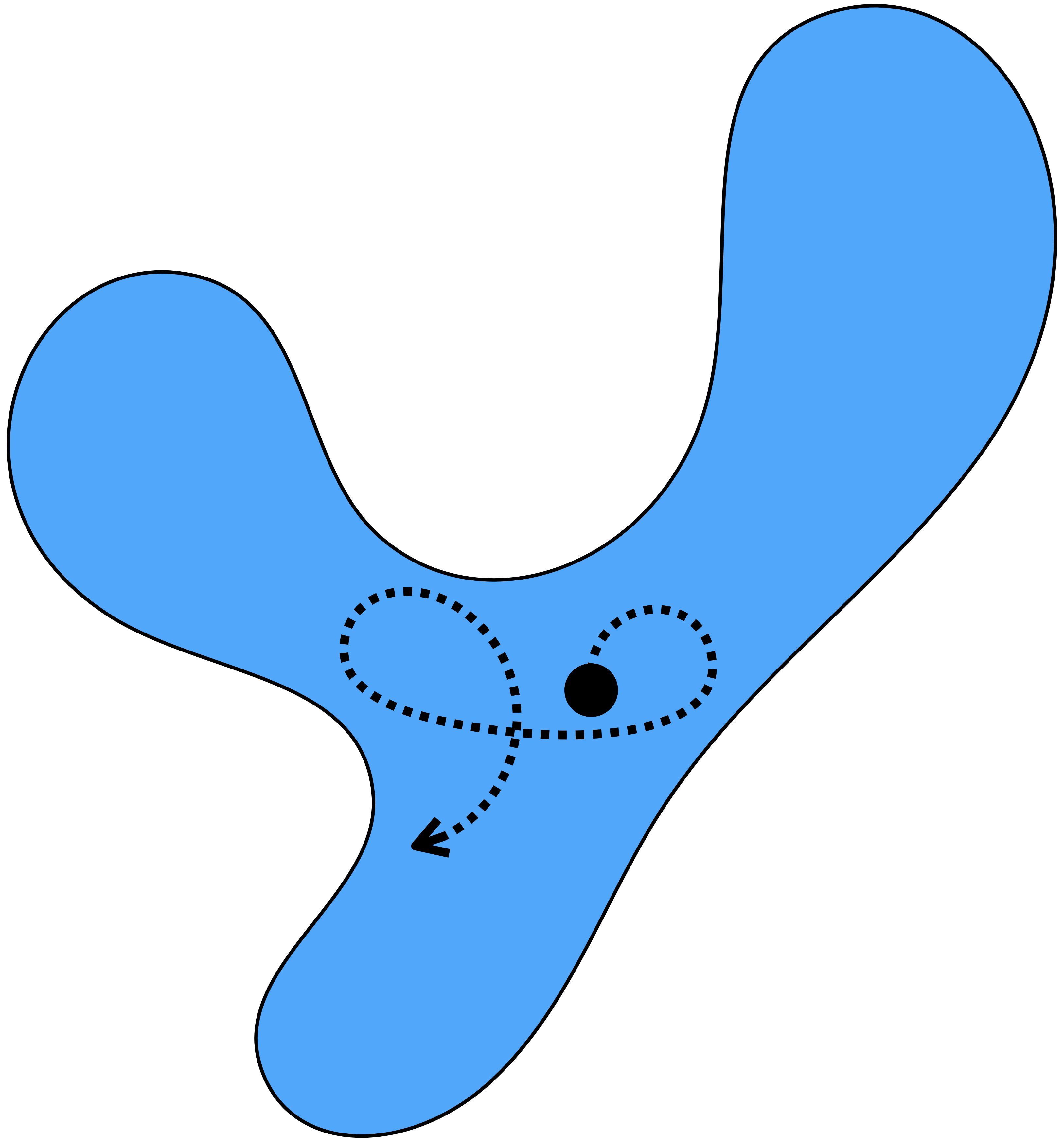
GANs let us explore the manifold of natural images

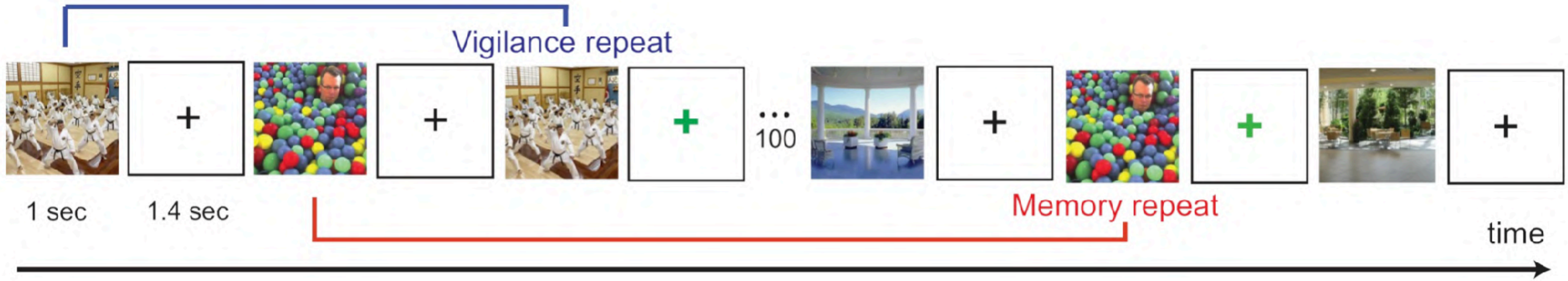
- What does it look like to make an image more memorable?

Lore Goetschalckx

Alex Andonian

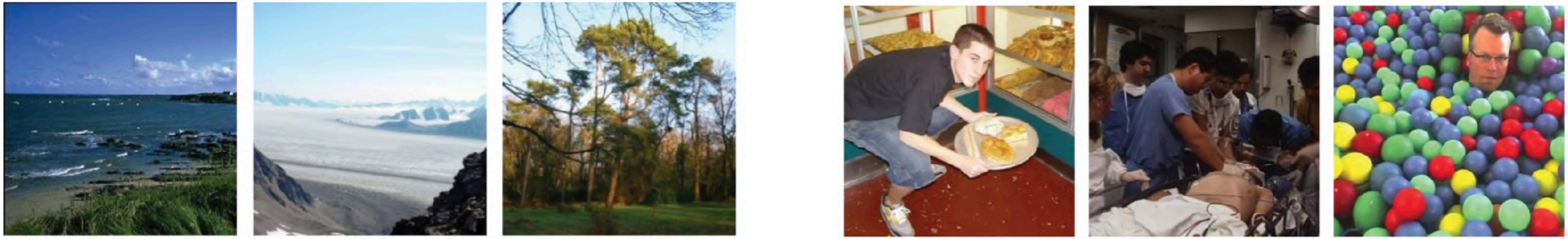
“GANalyze: Toward Visual Definitions of Cognitive Image Properties”





Forgettable

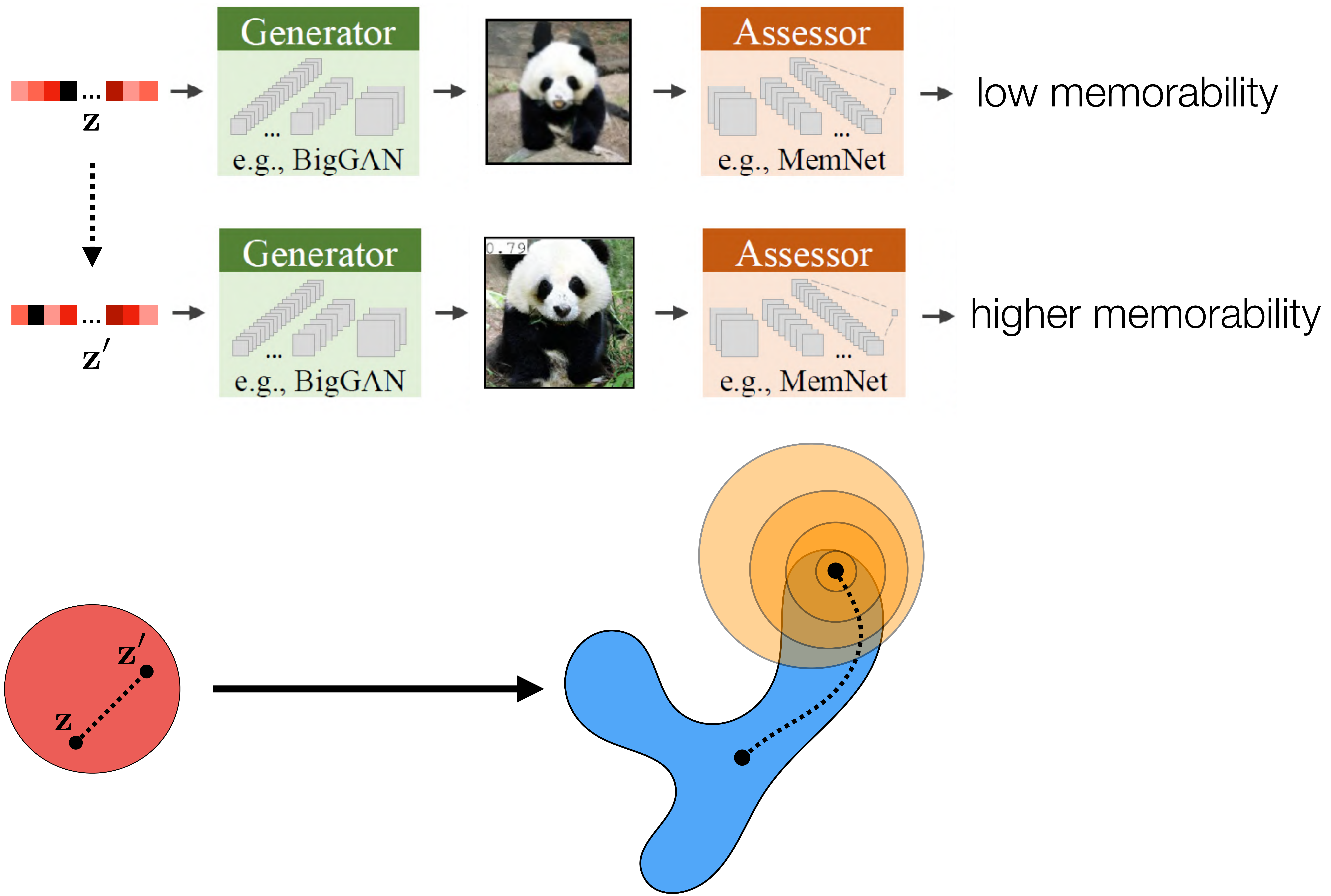
Memorable

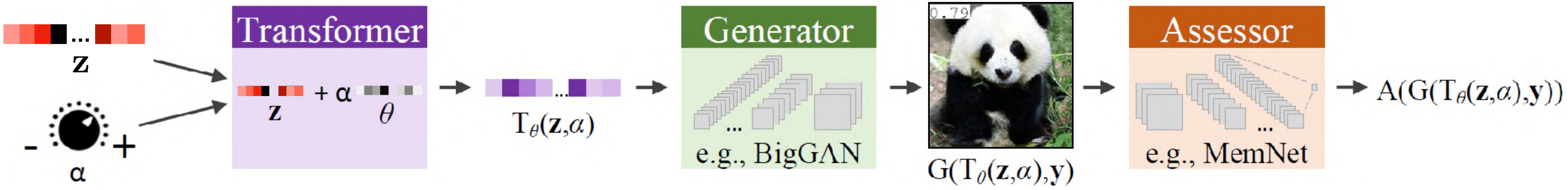


[Isola et al. 2011, Khosla et al. 2015]

Nonparametric visualization

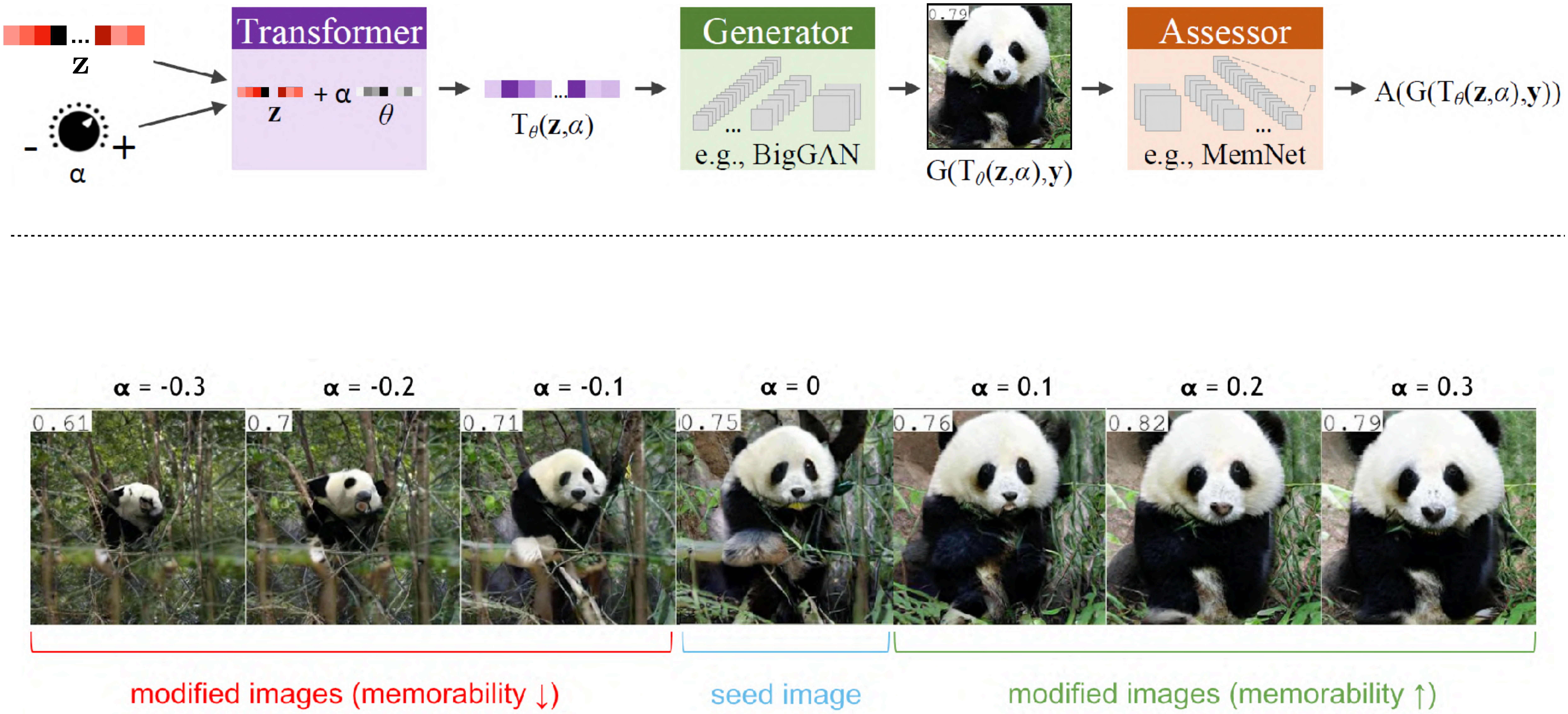
Memorability increases left-to-right, then top-to-bottom


Generator
e.g., BigGAN
Assessor
e.g., MemNet
low memorability
z
Generator
e.g., BigGAN
Assessor
e.g., MemNet
higher memorability
z′
z′
z



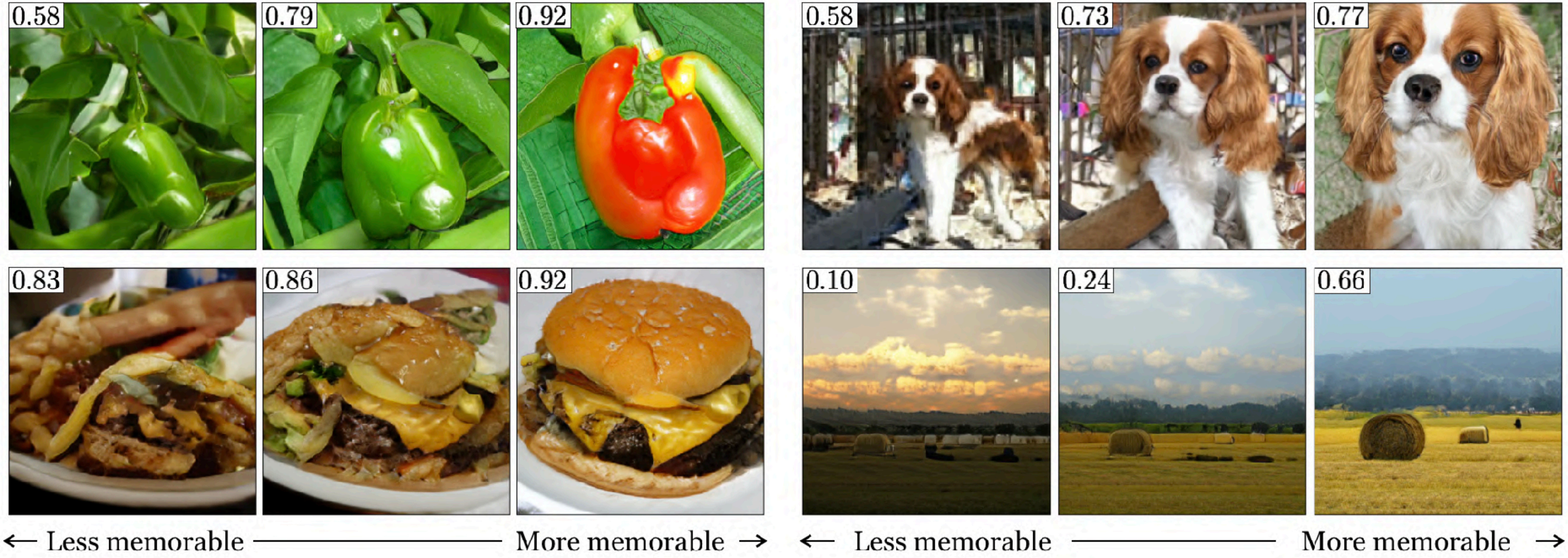
$$\mathcal{L}(\theta) = \mathbb{E}_{\mathbf{z},\mathbf{y},\alpha}\Big[\Big(\overbrace{A(G(T_\theta(\mathbf{z},\alpha),\mathbf{y}))}^{\text{Assessor score of the modified image}} - \big(\underbrace{A(G(\mathbf{z},\mathbf{y}))}_{\text{Assessor score of the seed image}} + \underbrace{\alpha}_{\text{Incremented by } \boldsymbol{\alpha}}\big)\Big)^2\Big]$$

$$T(\mathbf{z},\alpha) = \mathbf{z} + \alpha\theta$$



z
α
Transformer
z
+ α
θ
Tθ(z,α)
Generator
e.g., BigGAN
0.79
G(Tθ(z,α),y)
Assessor
e.g., MemNet
A(G(Tθ(z,α),y))
α = -0.3
α = -0.2
α = -0.1
α = 0
α = 0.1
α = 0.2
α = 0.3
0.61
0.7
0.71
0.75
0.76
0.82
0.79
modified images (memorability ↓)
seed image
modified images (memorability ↑)

What does it look like to make an image more memorable?



Zoom


0.57
0.58
0.65
0.81
0.89
0.94
0.96
0.27
0.45
0.55
0.56
0.64
0.77
0.8
Less memorable
More memorable

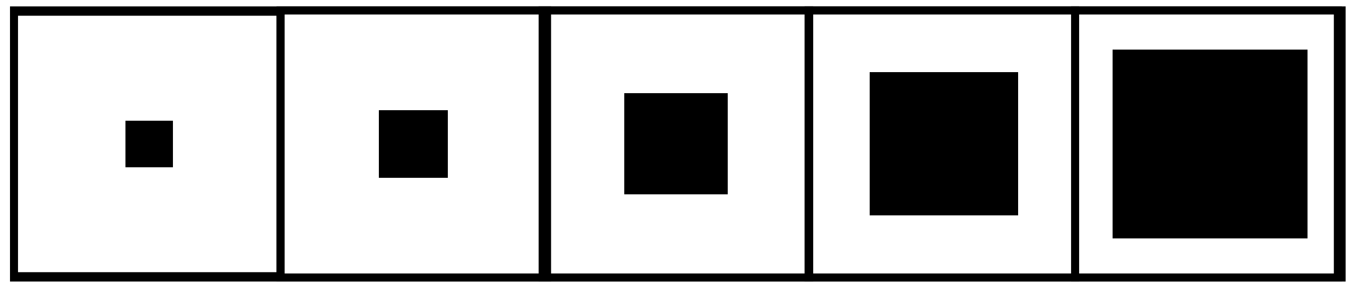
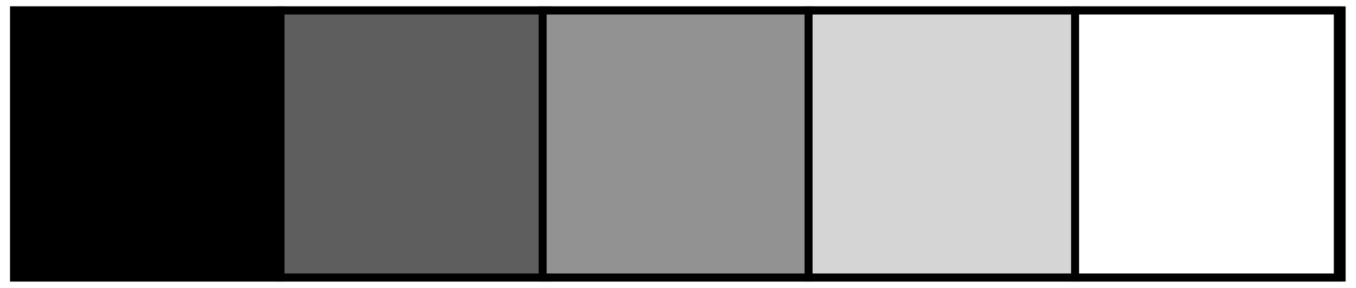
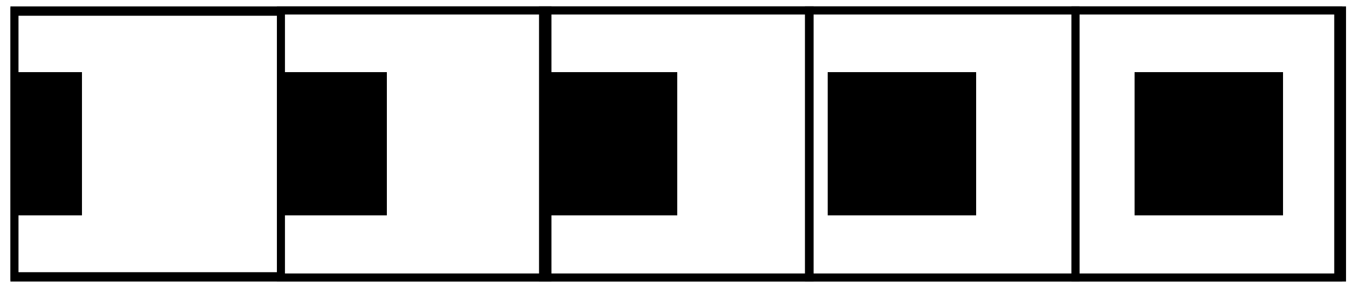
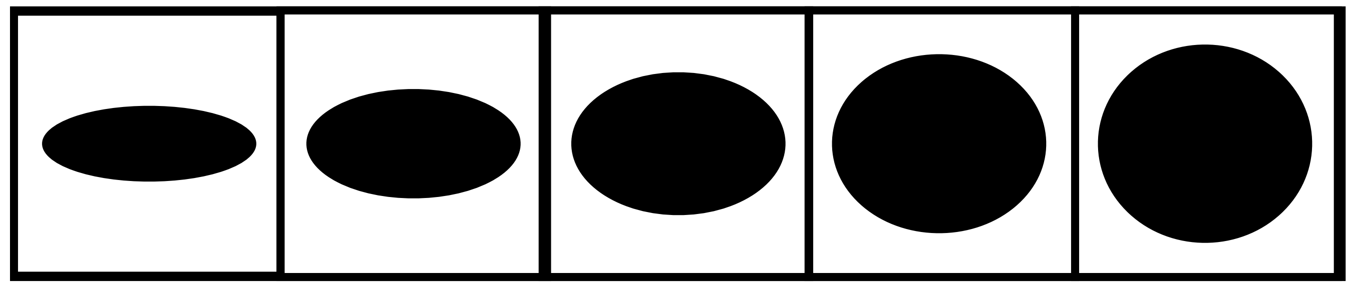
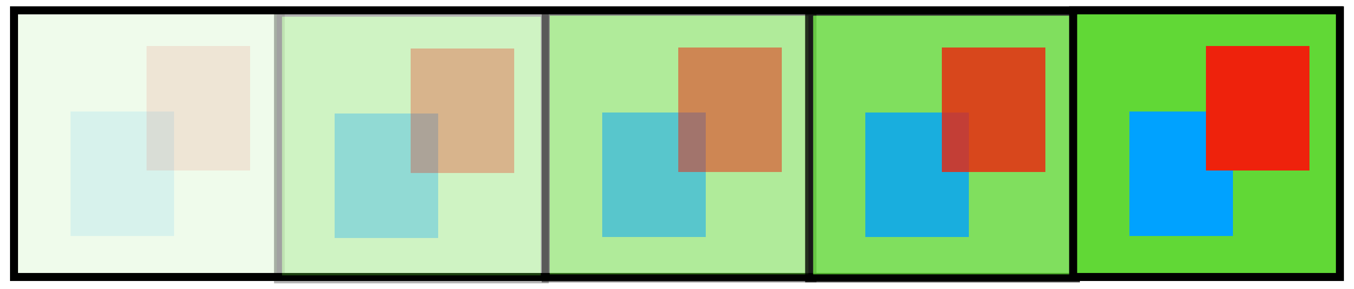
Circularity

← Less memorable ——— More memorable →

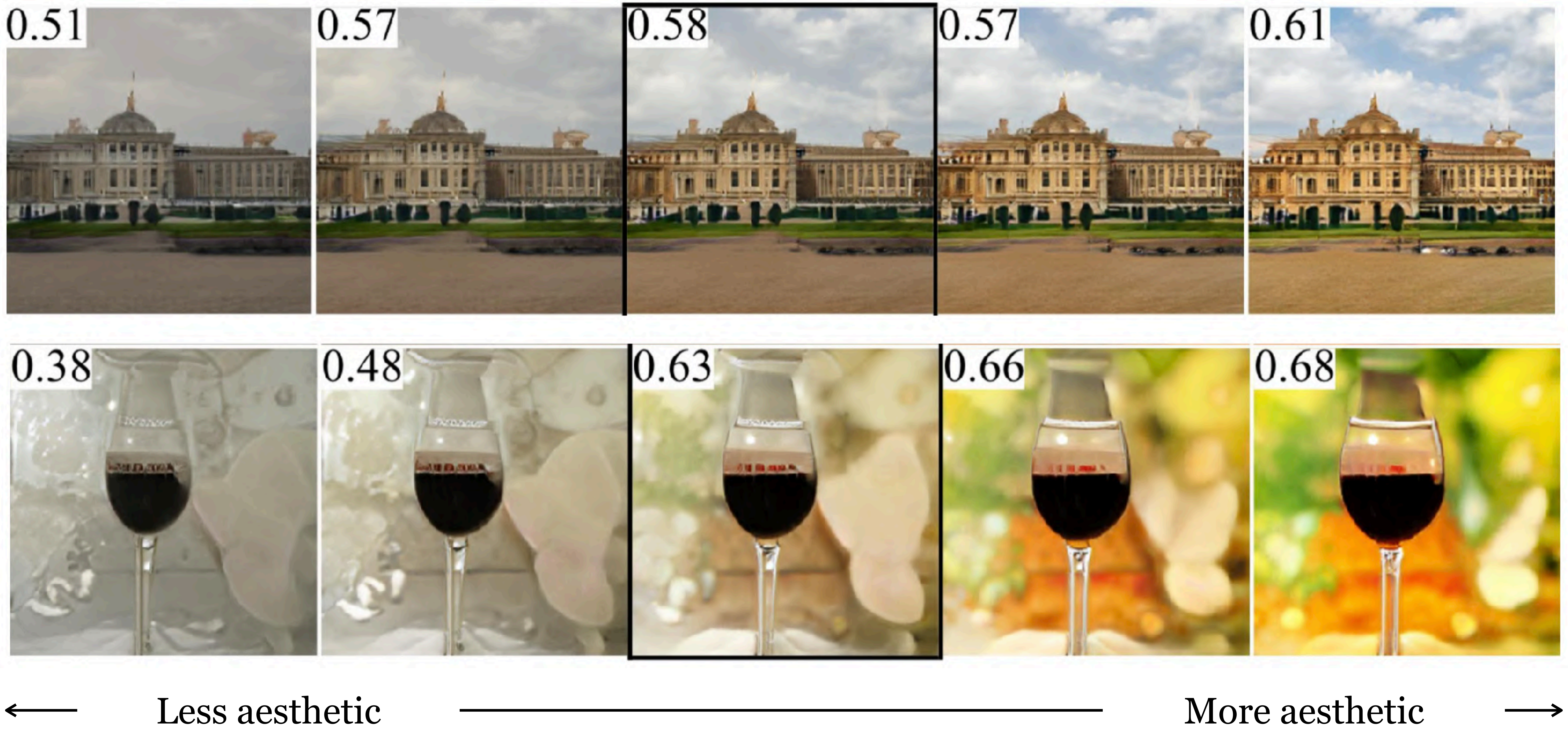
Simplicity


0.54
0.57
0.59
0.62
0.68
0.76
0.85
0.84
0.82
0.83
0.92
0.91
0.89
0.91
Less memorable
More memorable

Expressive face ("cuteness"?)

Factor		Log Odds
Object size		0.32
Brightness		0.28
Centeredness		0.24
Object shape		0.19
Colorfulness		0.17

What does it look like to make an image more *aesthetic*?

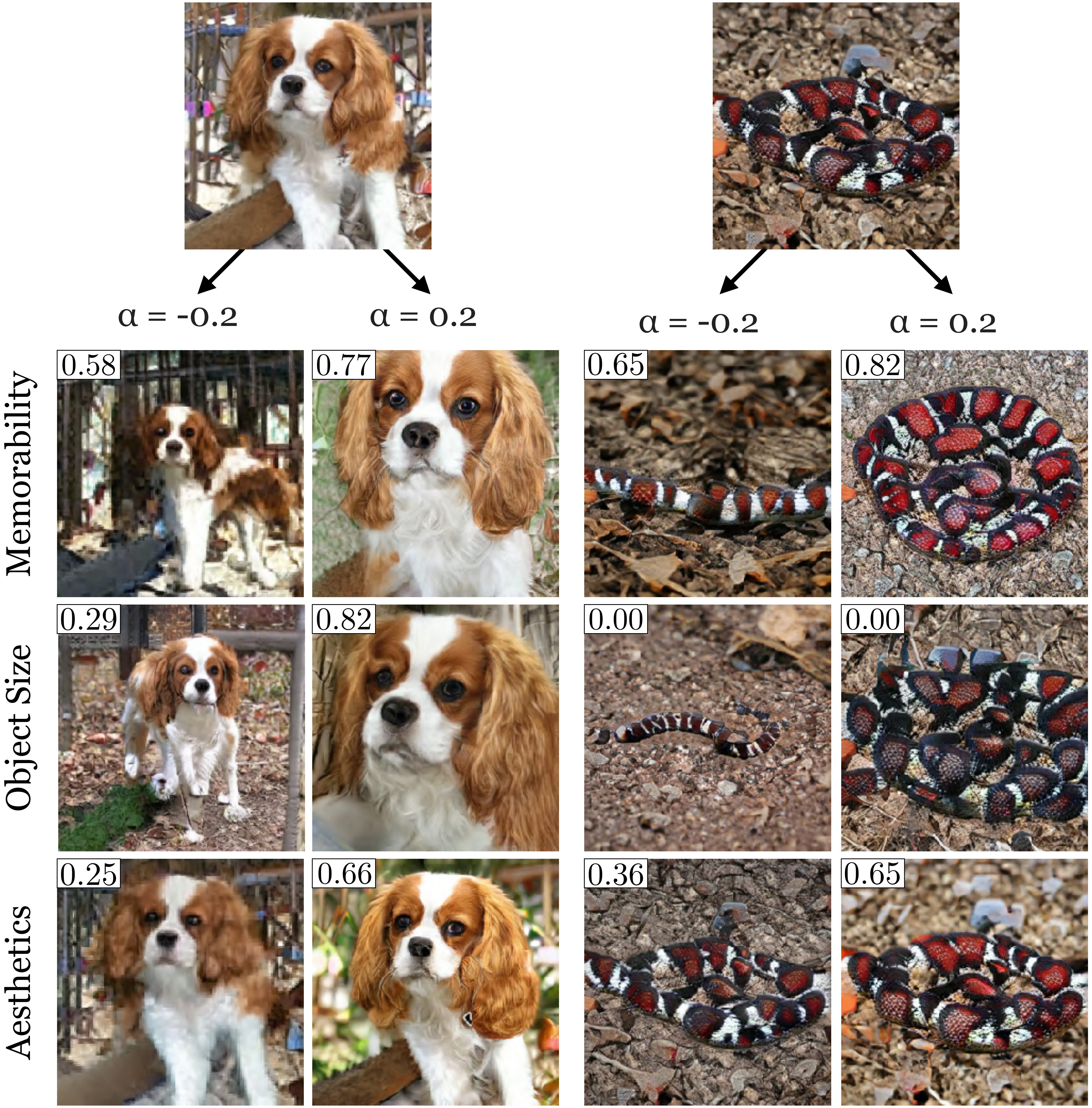


Intsagram popularity

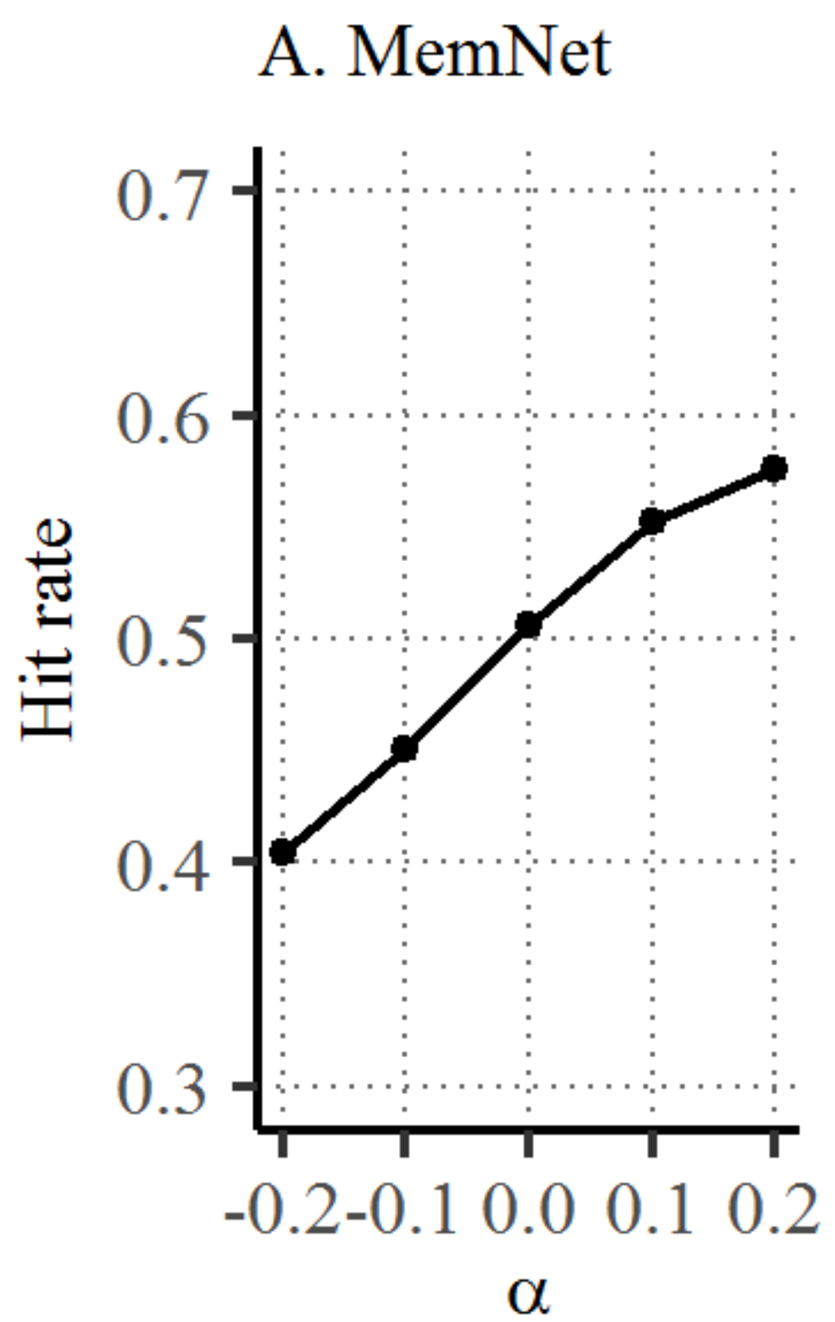
← Unpopular ——— Popular →

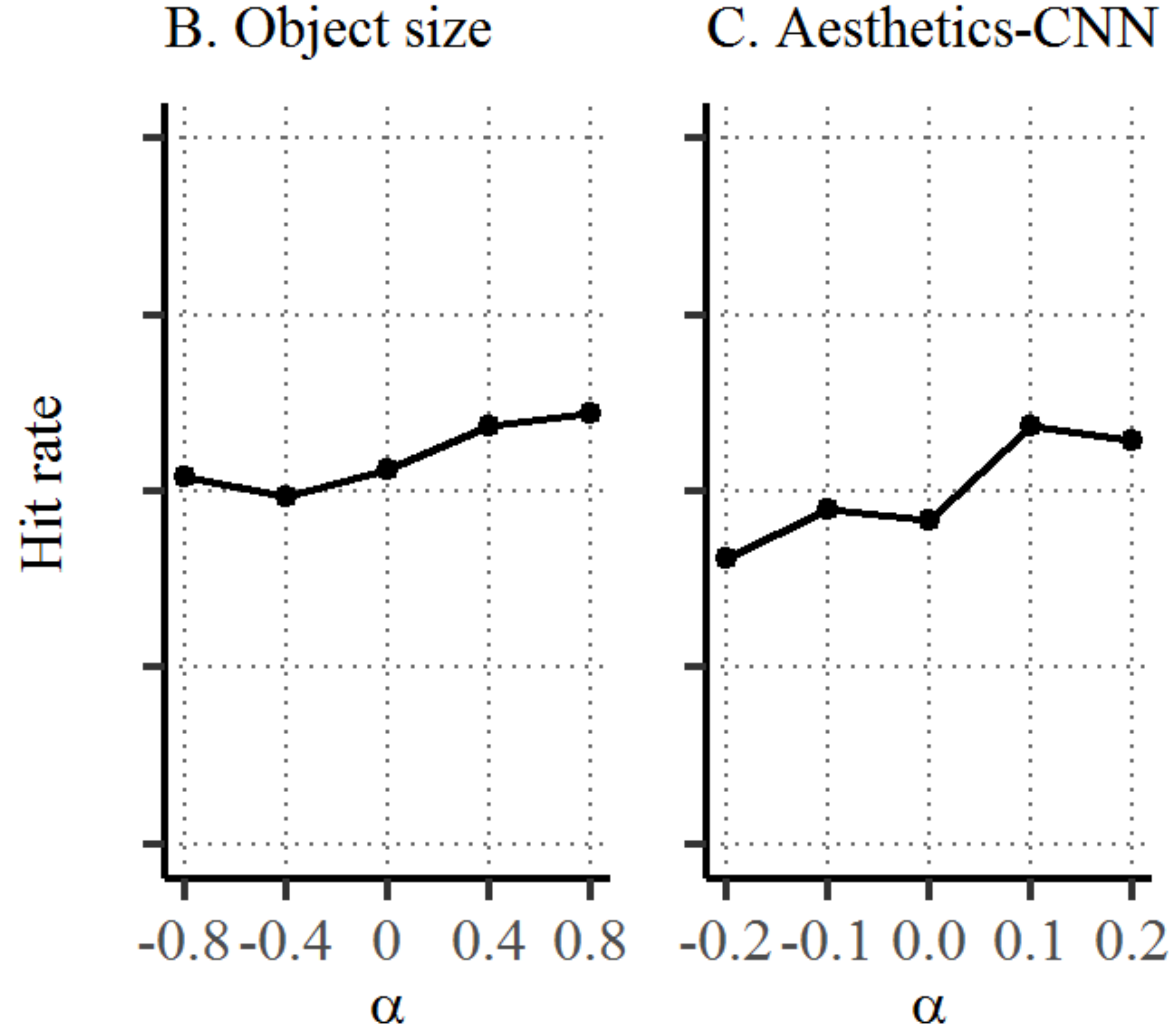
Emotional valence

← Negative ——— Positive →


α = -0.2
α = 0.2
α = -0.2
α = 0.2
Memorability
Object Size
Aesthetics
0.58
0.77
0.65
0.82
0.29
0.82
0.00
0.00
0.25
0.66
0.36
0.65

Causal?


A. MemNet
Hit rate
0.7
0.6
0.5
0.4
0.3
-0.2
-0.1
0.0
0.1
0.2
α


B. Object size
C. Aesthetics-CNN
Hit rate
-0.8
-0.4
0
0.4
0.8
α
-0.2
-0.1
0.0
0.1
0.2
α

“Visual definitions”

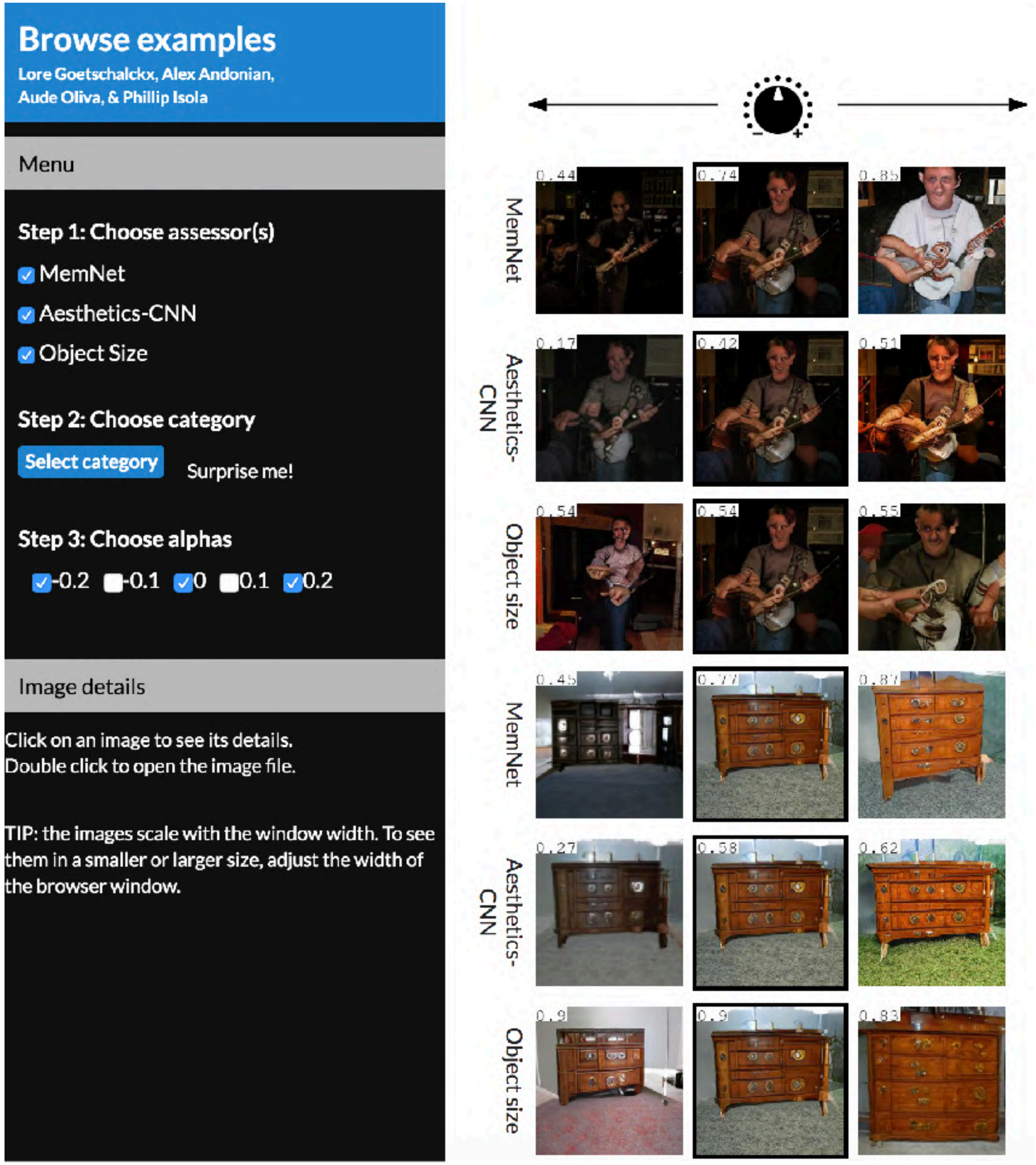
- Visually map out the full extent and continuum of a concept

+ Visualize abstract concepts (memorability, beauty, emotional valence, etc)

+ Scientific understanding

+ Graphics applications

[c.f. “Activation Atlases”, Carter et al. 2019]



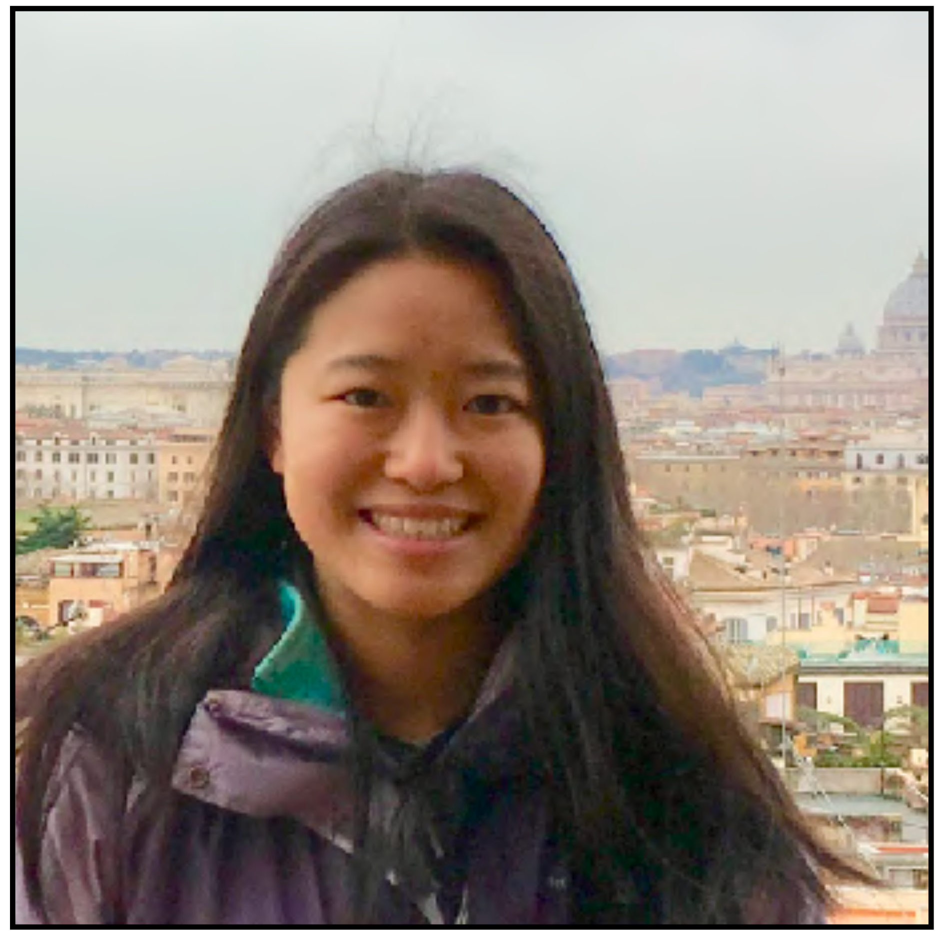
[http://memgame.csail.mit.edu/explorer/browse_examples.html]

“On the ‘steerability’ of Generative Adversarial Networks”

[Jahanian*, Chai*, Isola, arXiv 2019]

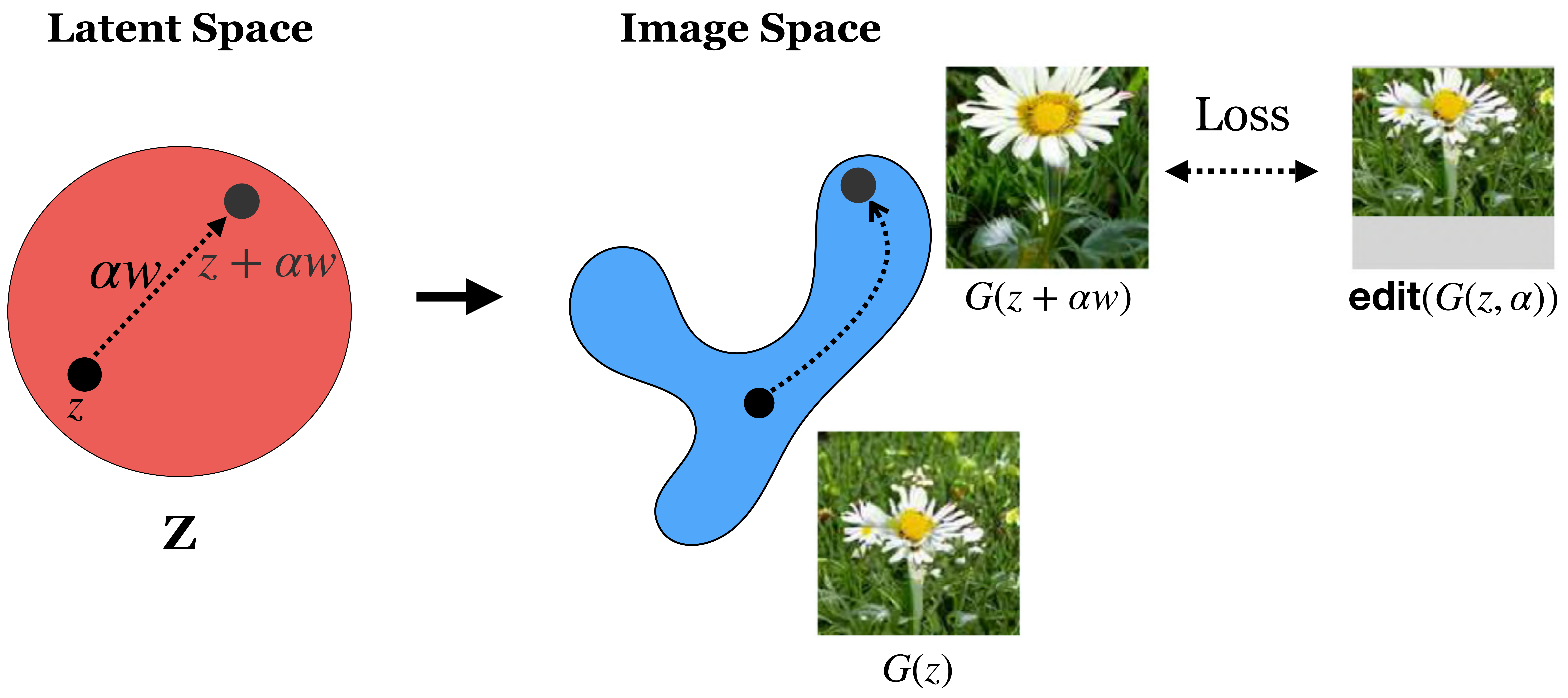
Let’s take a step back and look at basic color and camera transformations

How close is a GAN to a classic graphics engine?

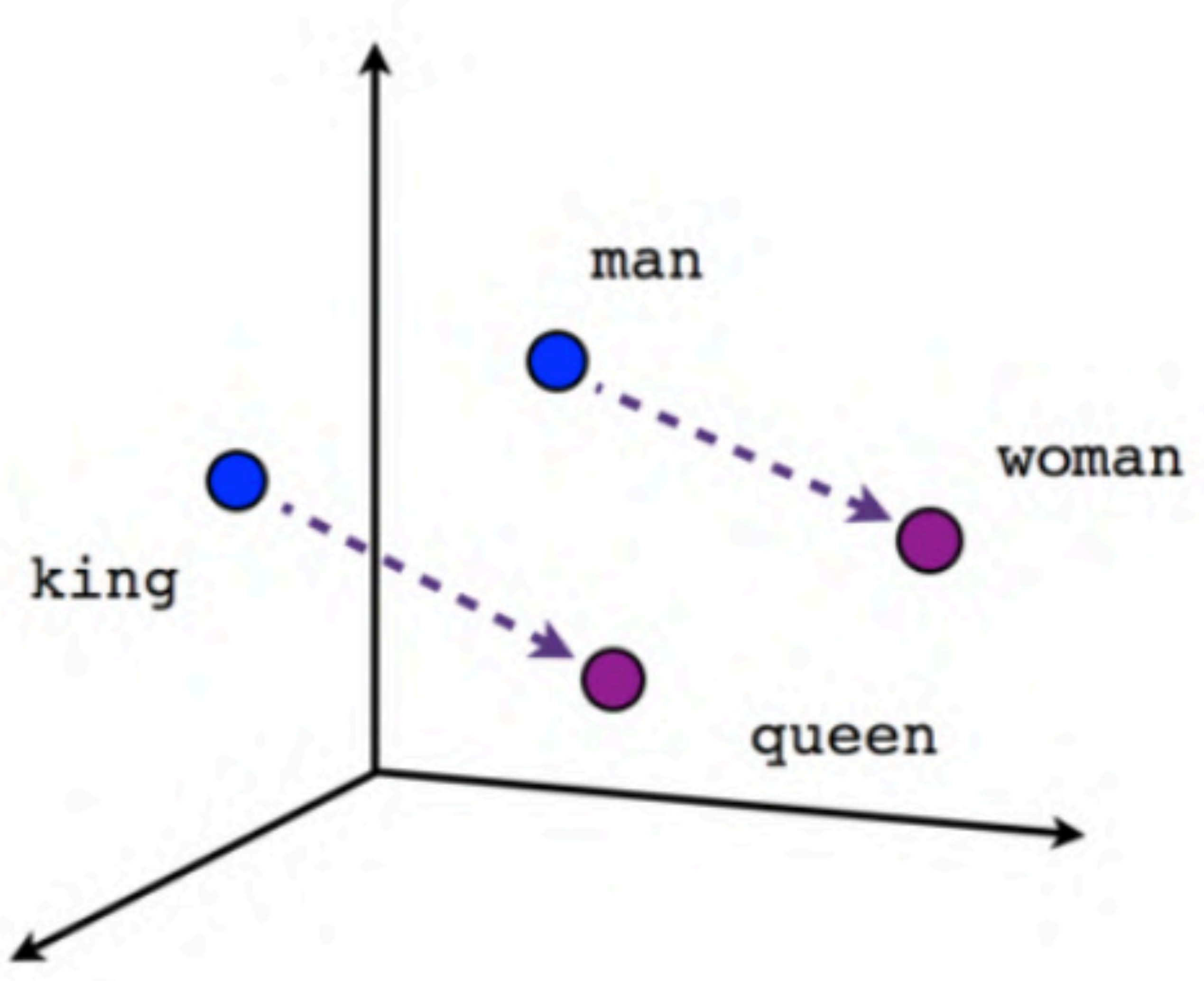


Lucy Chai

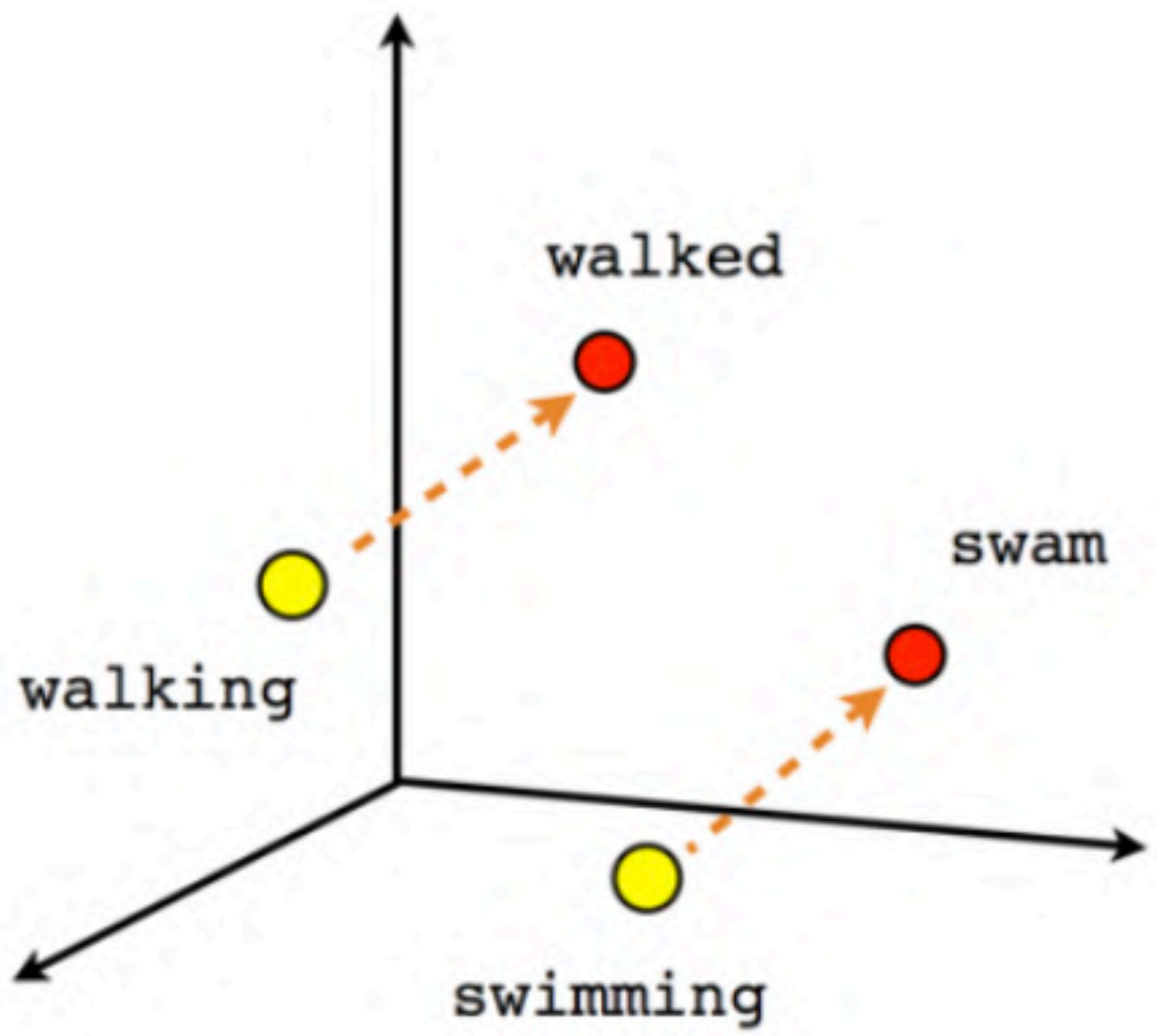
Ali Jahanian



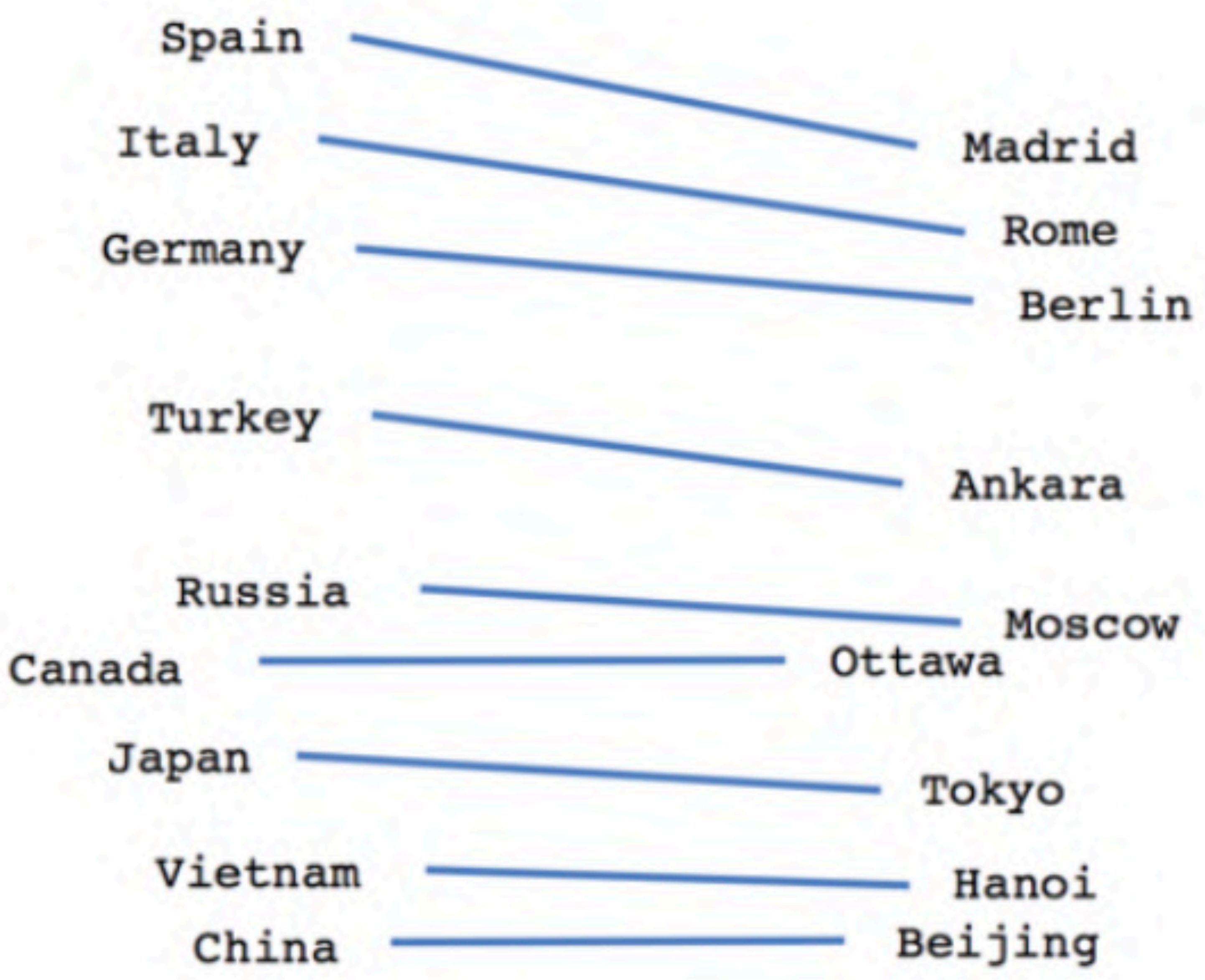
$$w^* = \arg\min_w \mathbb{E}_{z,\alpha}[\mathcal{L}(G(z+\alpha w), \mathtt{edit}(G(z), \alpha))]$$



Male-Female

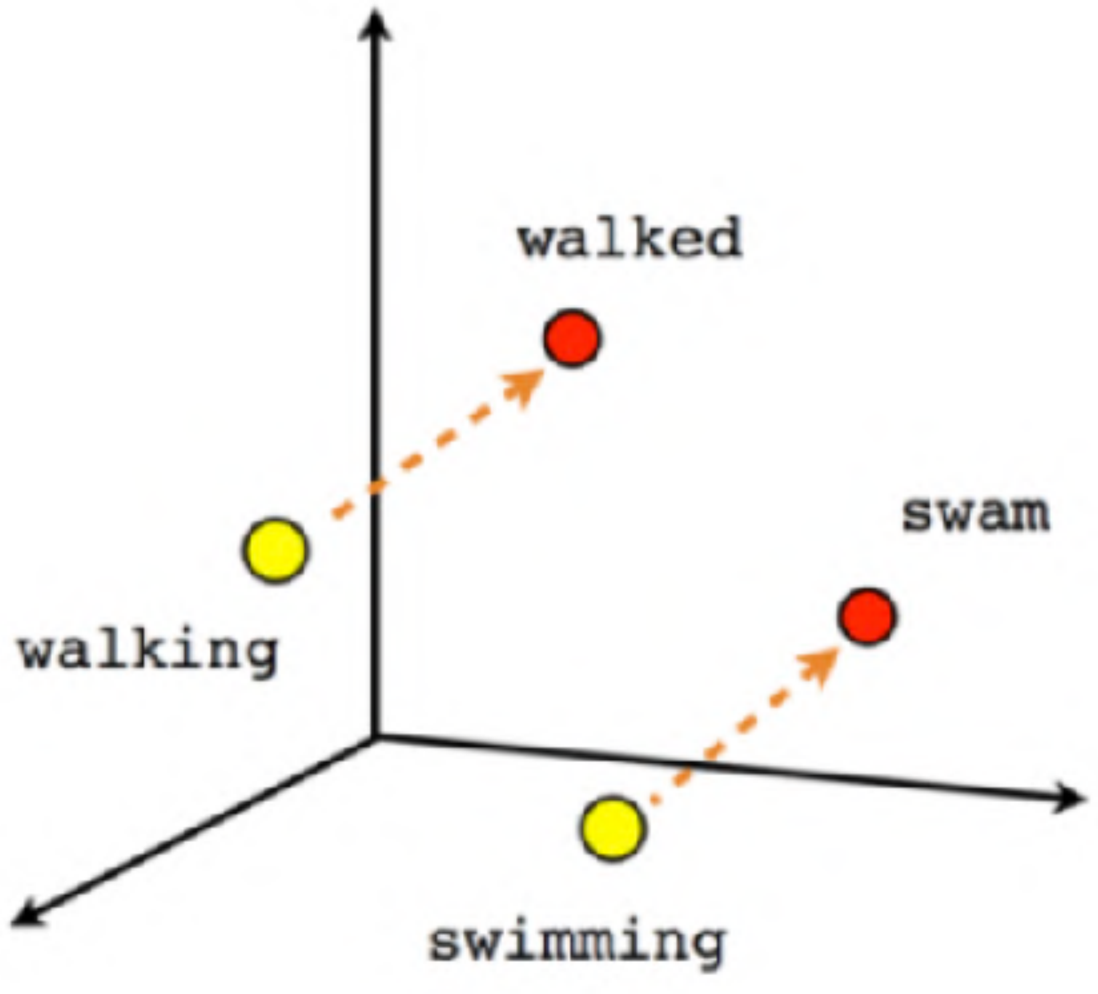


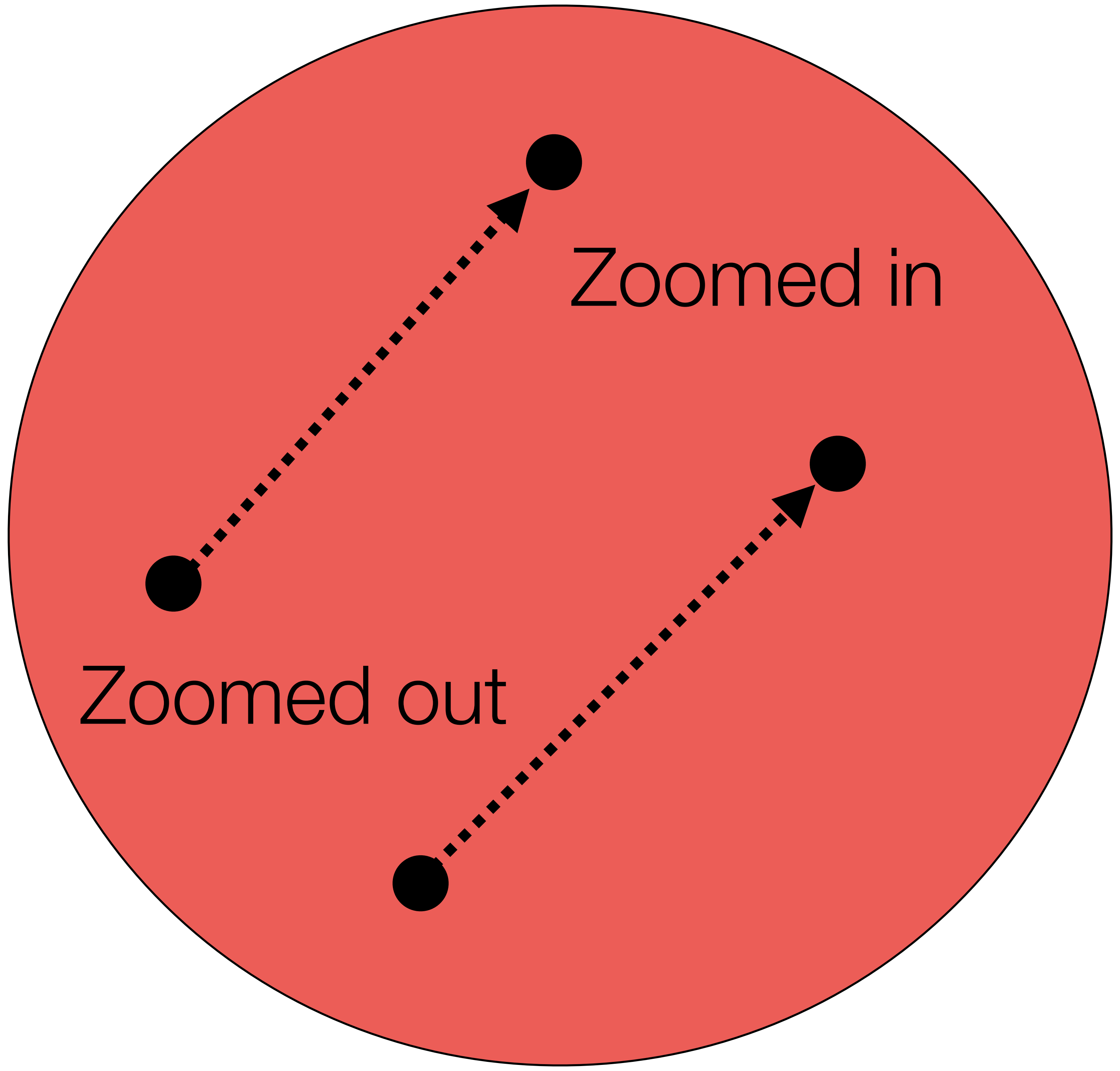
Verb tense



Country-Capital

Examples from https://www.tensorflow.org/tutorials/representation/word2vec





Zoom in vector

Shift vector

Color transformation vectors

Winter to spring

input output

Turning on the lights

input output

Day to night

input output

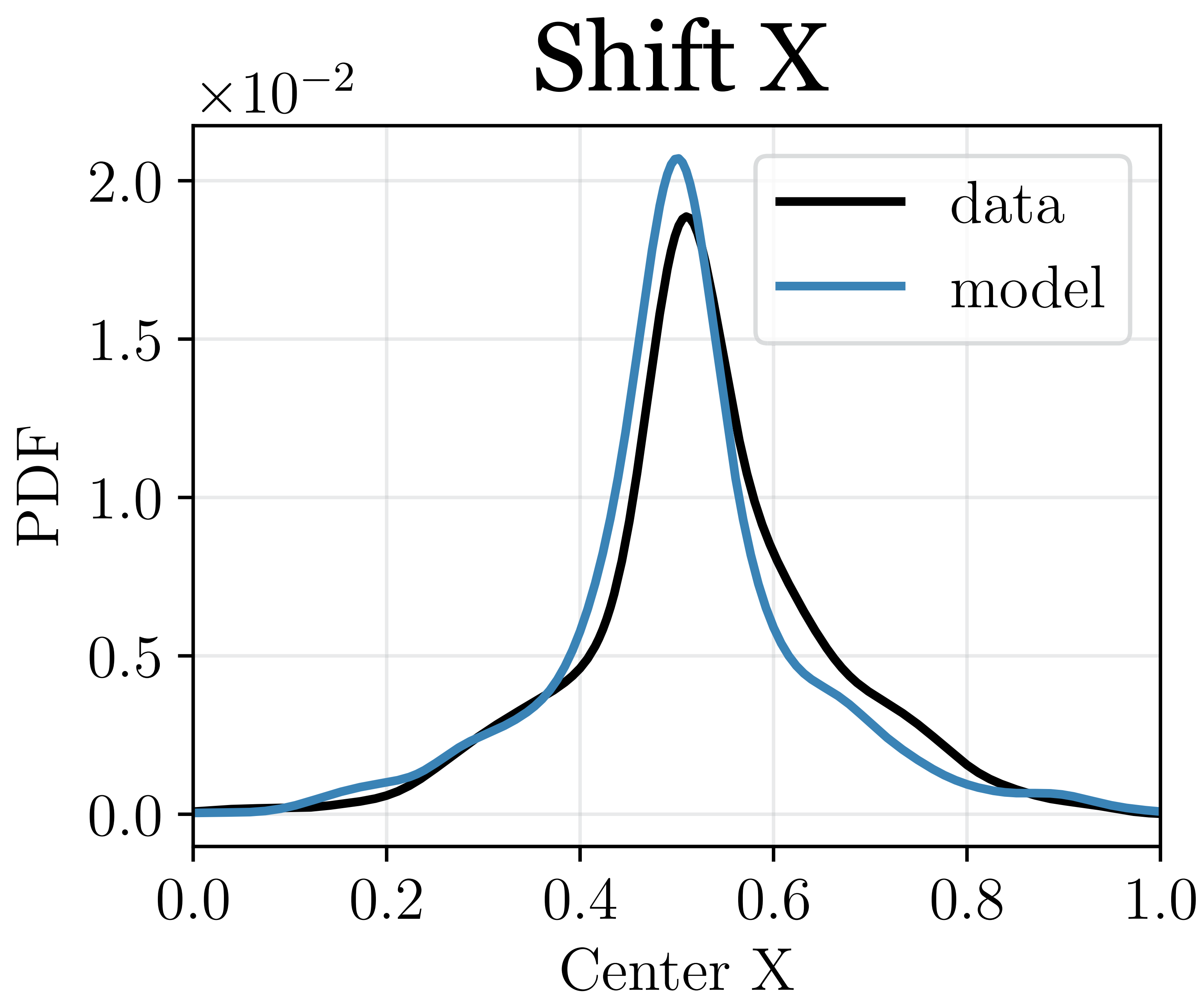
Volcano!

input output

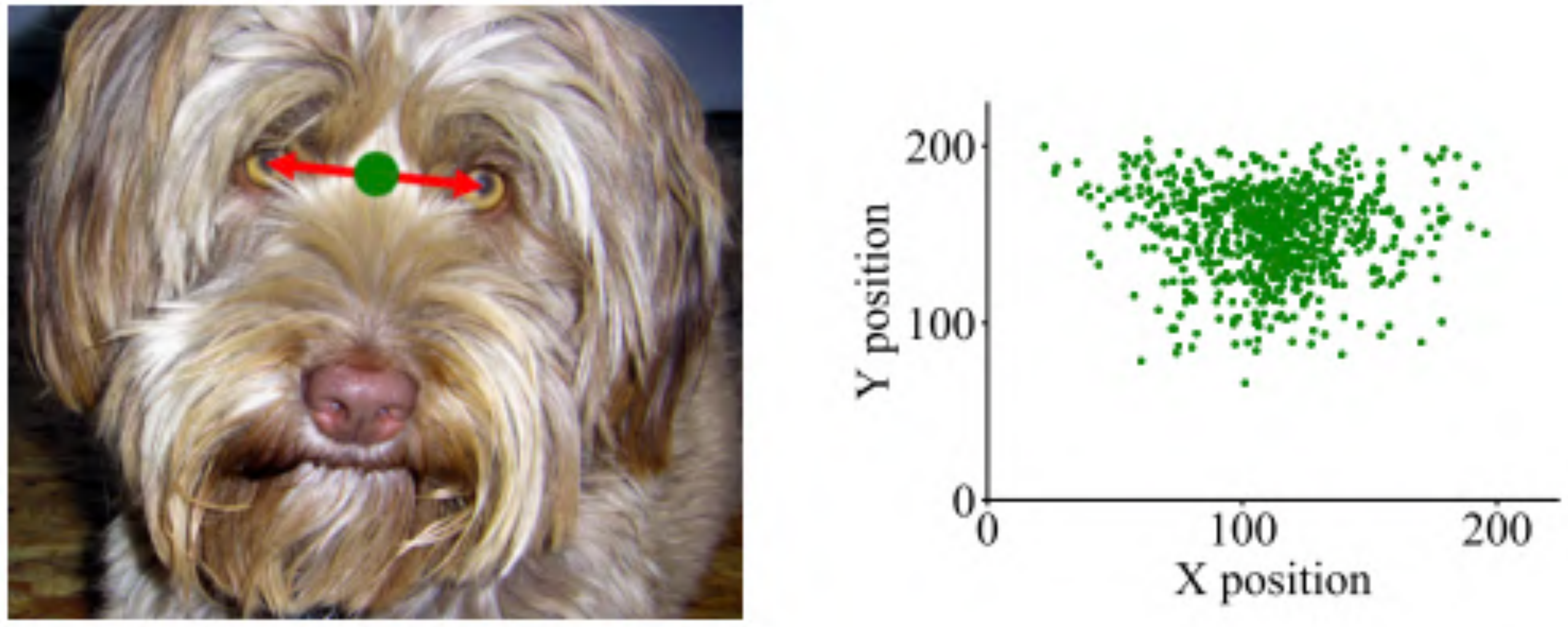
There is a latent space “vector” for each of these transformations —
the “spring vector”, the “volcano exploding vector”


position
alpha (strength of transformation)
0.01
0.32
0.03
- Zoom +
0.21
0.58
0.13
- Shift Y +
0.05
0.09
0.05
- Rotate 2D +

Is this due to bias in the dataset or the model?

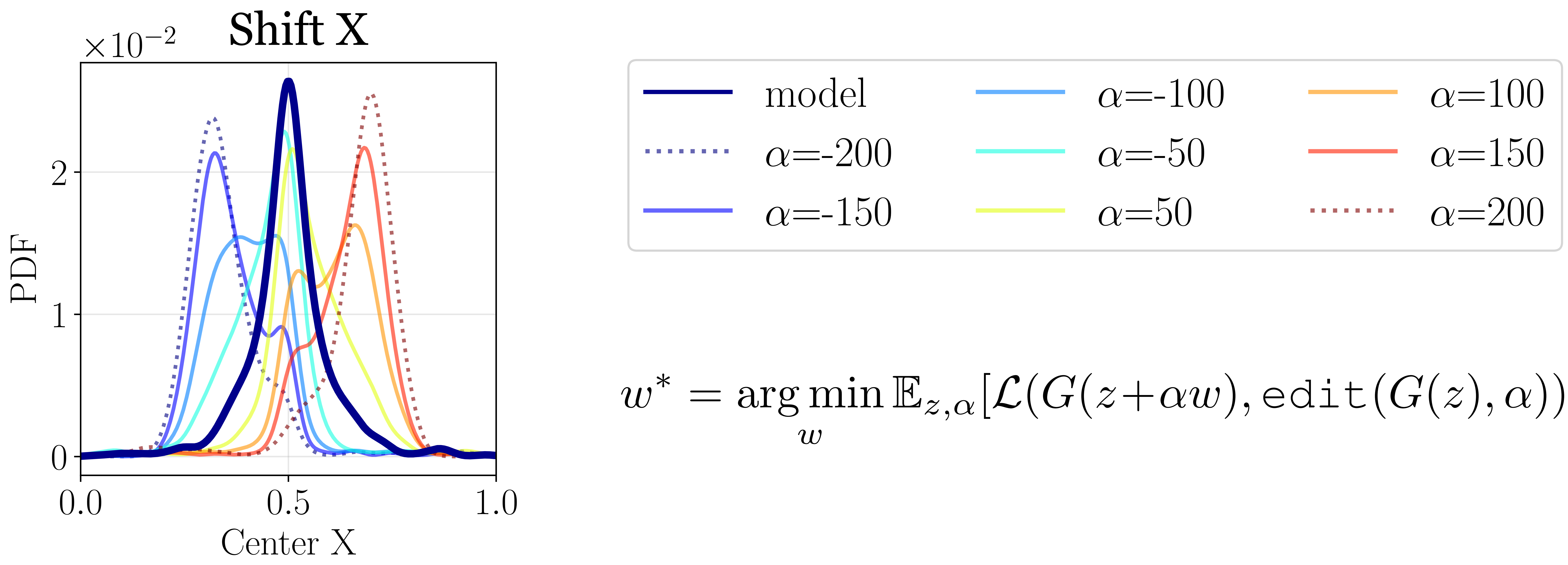


"Why do deep convolutional networks generalize so poorly to small image transformations?"



[Azulay & Weiss, 2019]

Are the transformations shifting the data distribution?

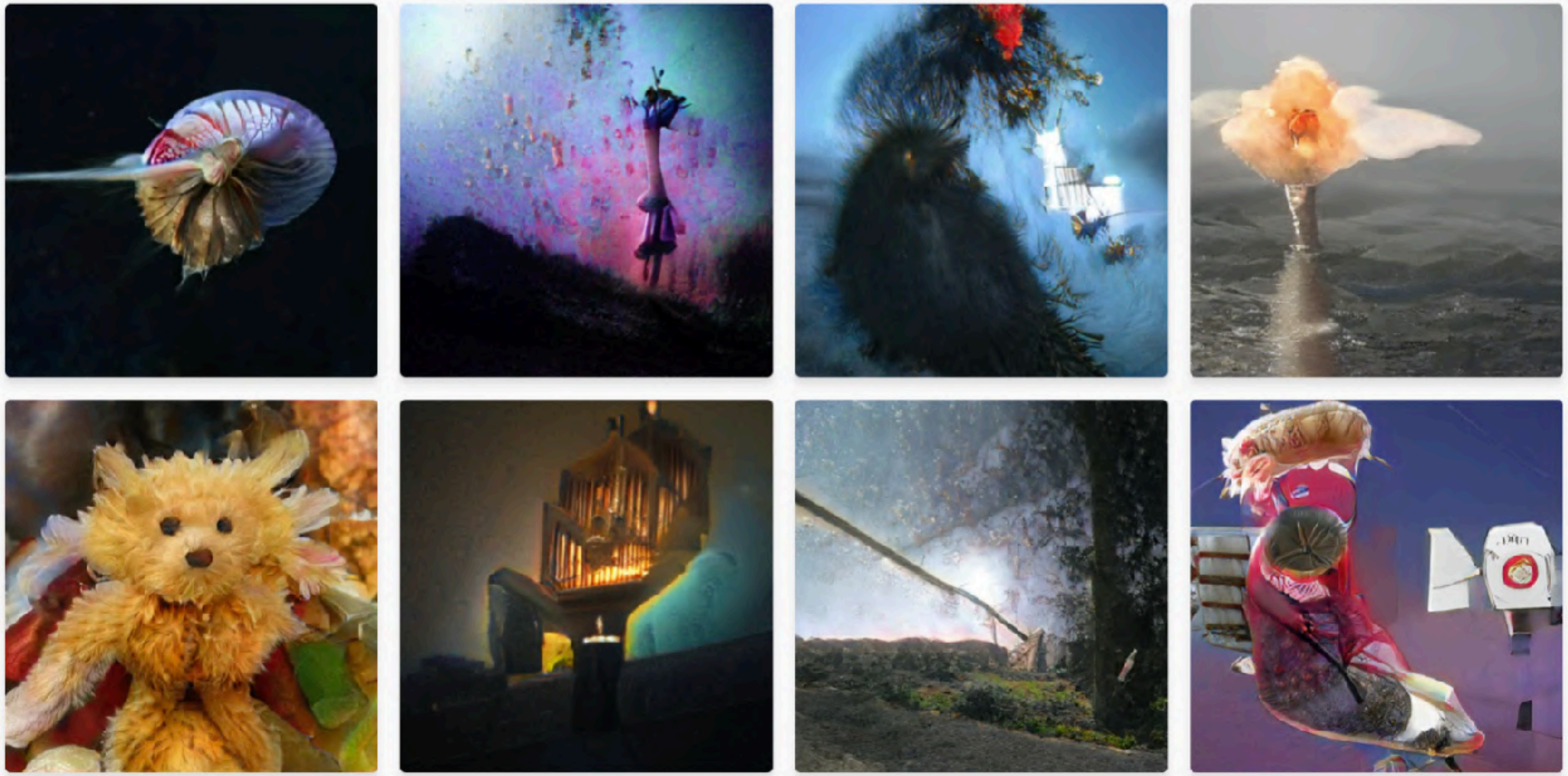


$$w^* = \arg\min_{w} \mathbb{E}_{z,\alpha}[\mathcal{L}(G(z+\alpha w), \texttt{edit}(G(z), \alpha))]$$

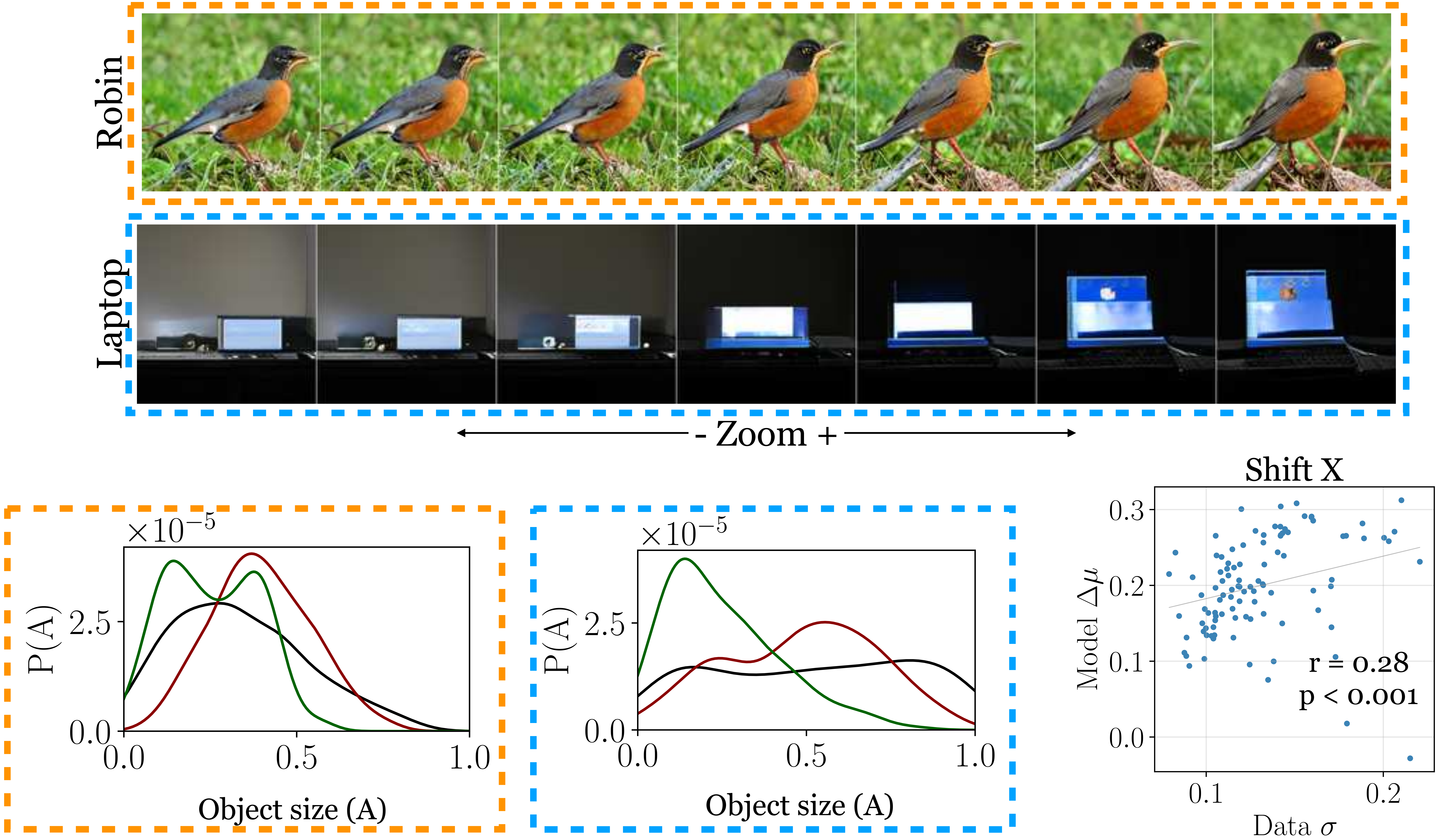
Yes, the transformations extrapolate! i.e. shift the data distribution

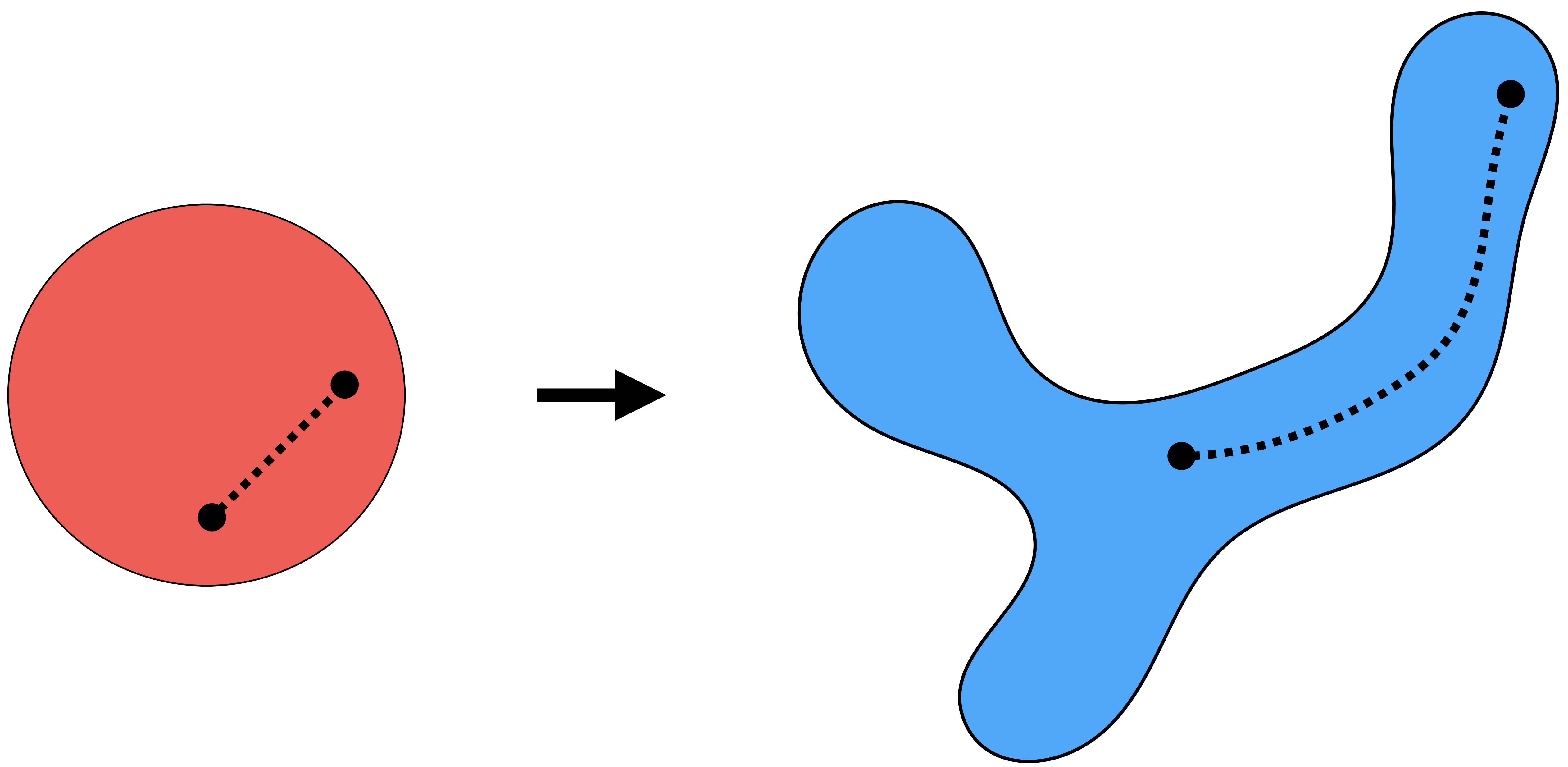
Ganbreeder

Create beautiful, wild and weird images



[https://ganbreeder.app/]


Robin
Laptop
- Zoom +
×10^−5
P(A)
2.5
0.0
0.0
0.5
1.0
Object size (A)
×10^−5
P(A)
2.5
0.0
0.0
0.5
1.0
Object size (A)
Shift X
Model Δμ
0.3
0.2
0.1
0.0
0.1
0.2
Data σ
r = 0.28
p < 0.001



— Generative models let us explore the data distribution as a continuum, interpolating between datapoints and even extrapolating beyond them

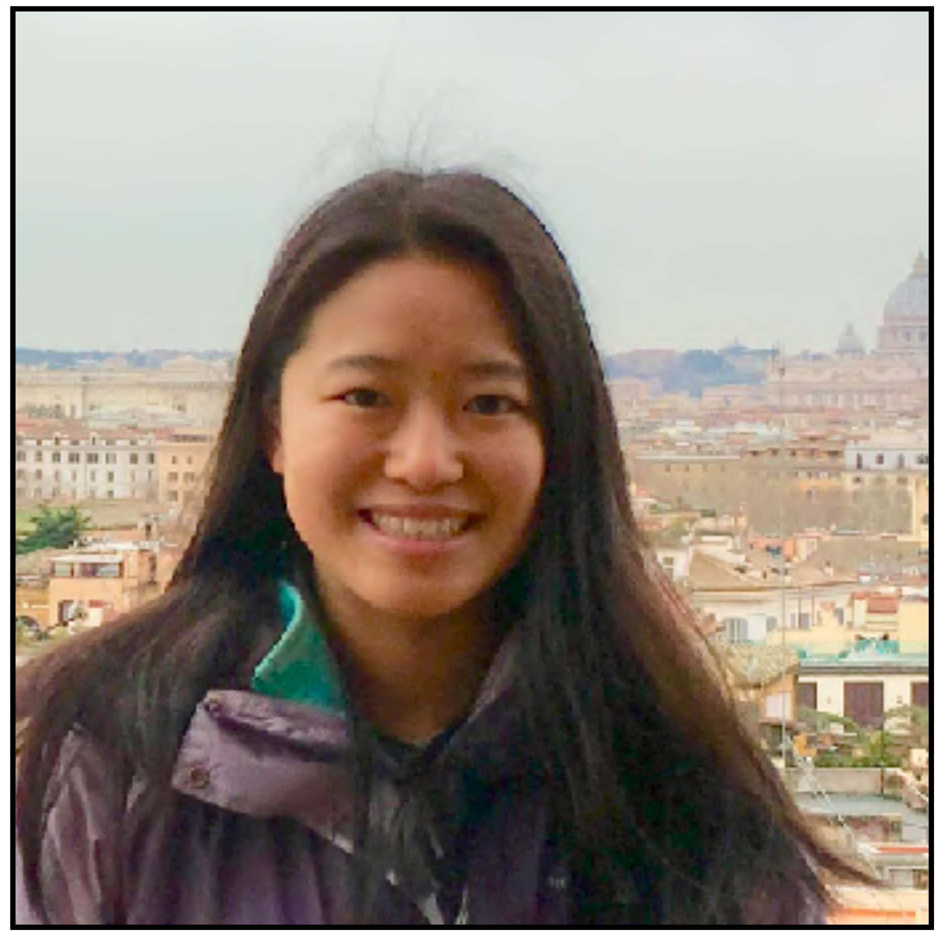
— Disentangled factors emerge, and we can learn to navigate the latent space to observe interesting transformations

Thanks!

Lore Goetschalckx

Alex Andonian

Aude Oliva



Lucy Chai

Ali Jahanian