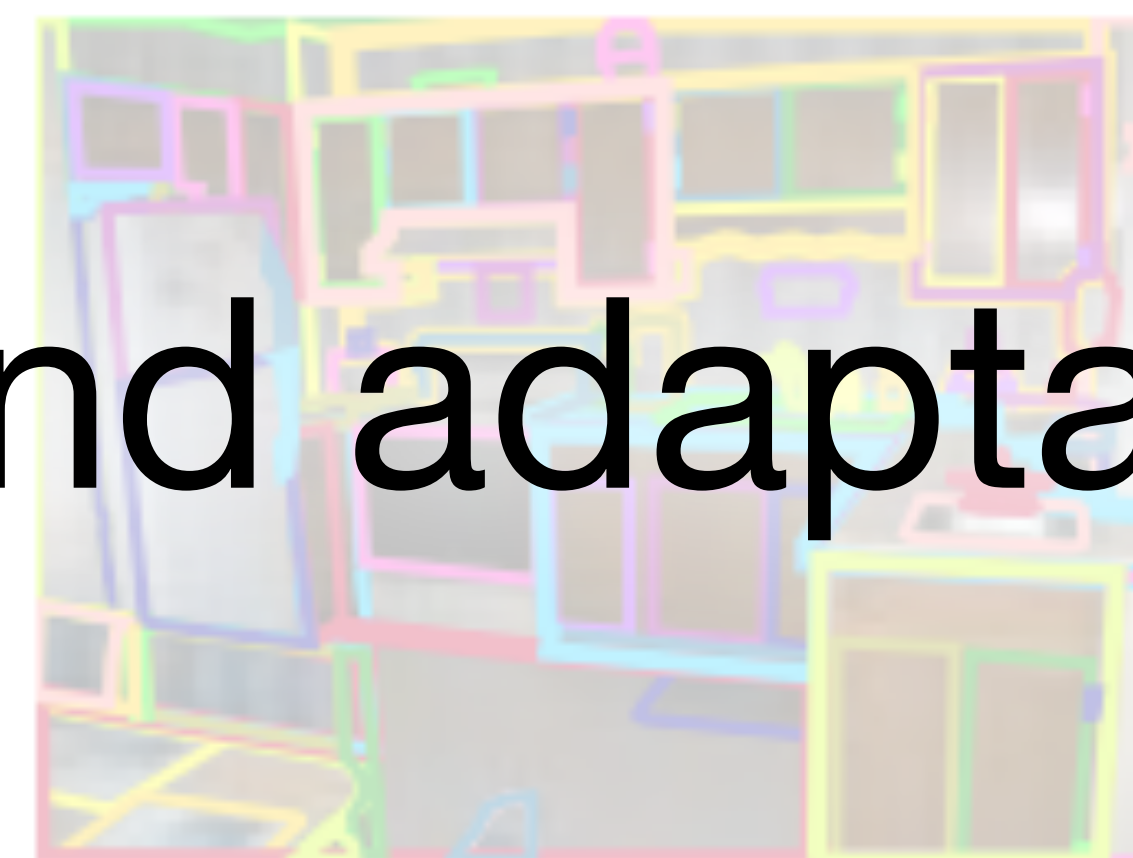
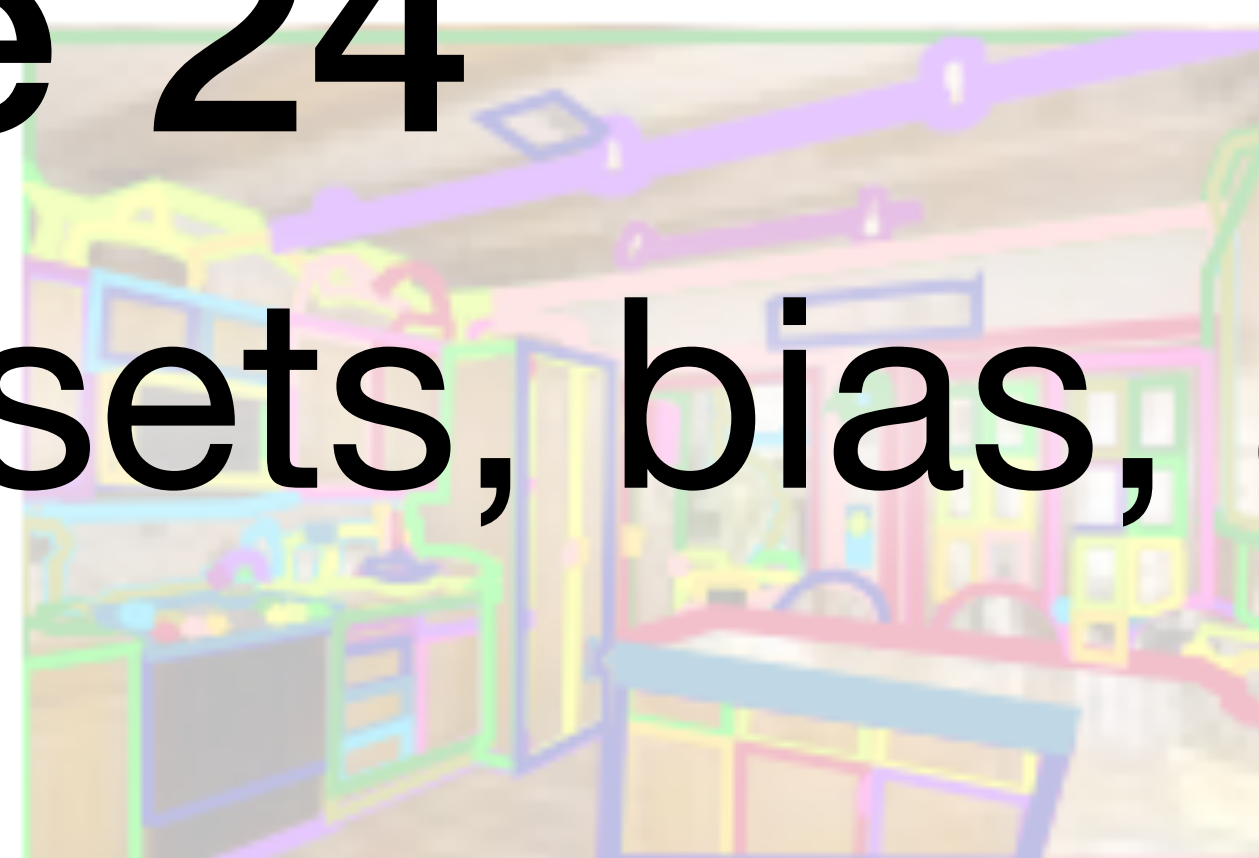


Lecture 24

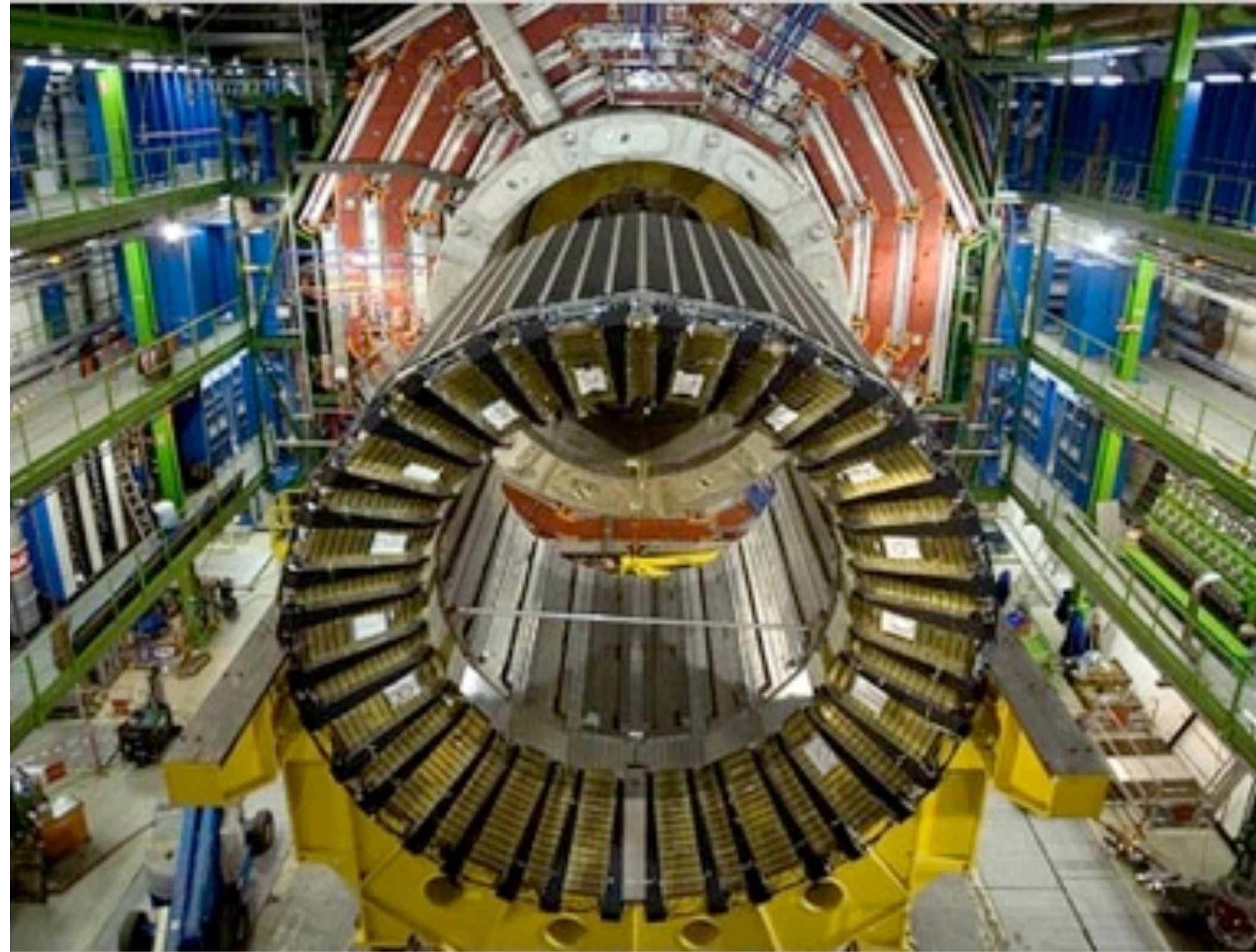
Datasets, bias, and adaptation



Once you decide on the model, architecture, ...

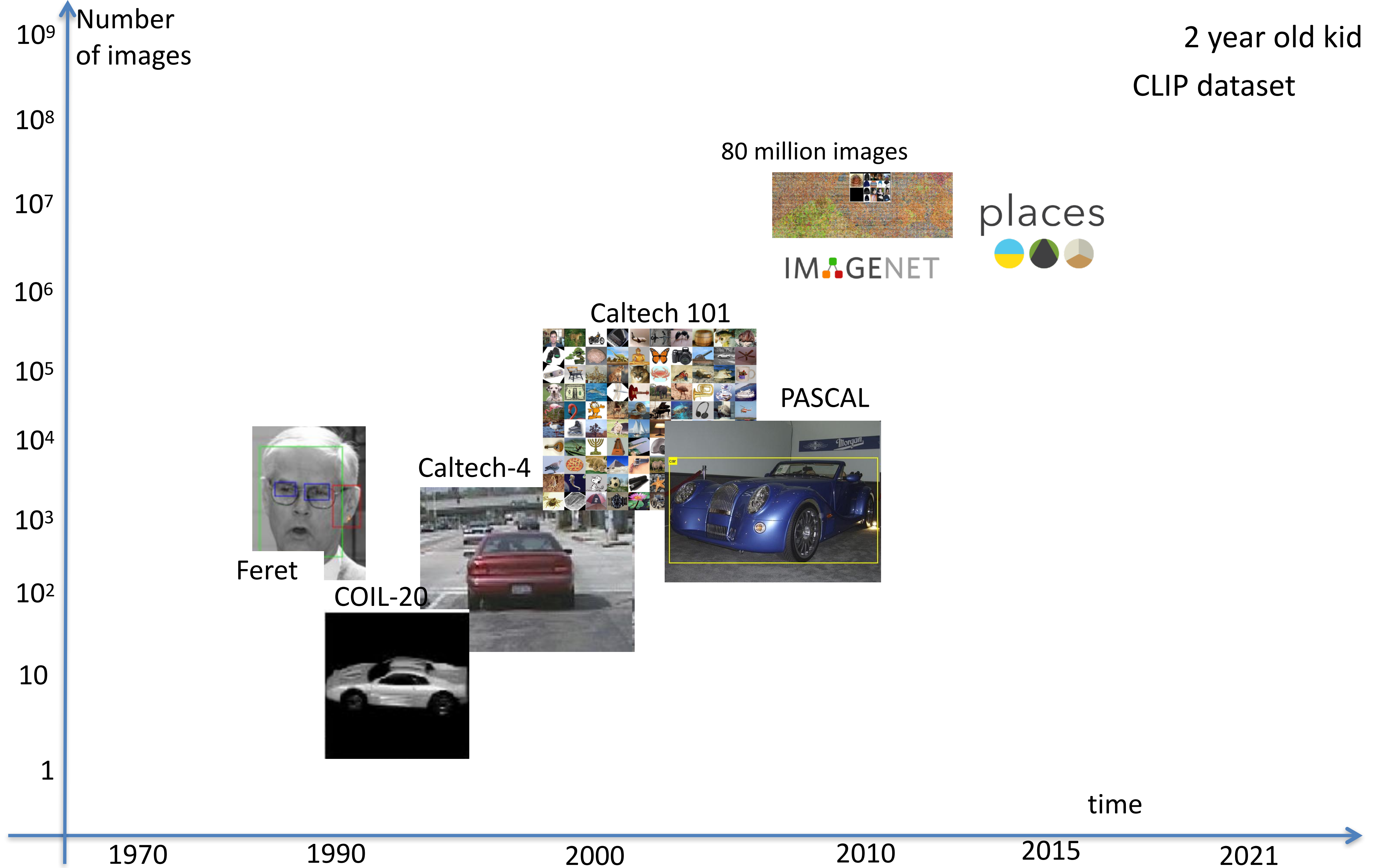
... it is all about the data

The value of data



The Large Hadron Collider

\$ 10¹⁰



The big data generation



WIKIPEDIA
The Free Encyclopedia

English

5 878 000+ articles

日本語

1 155 000+ 記事

Español

1 528 000+ artículos

Русский

1 551 000+ статей

Italiano

1 535 000+ voci

Português

1 008 000+ artigos

Deutsch

2 314 000+ Artikel

Français

2 116 000+ articles

中文

1 062 000+ 條目

Polski

1 342 000+ haseł



IMAGENET

A short story of image databases



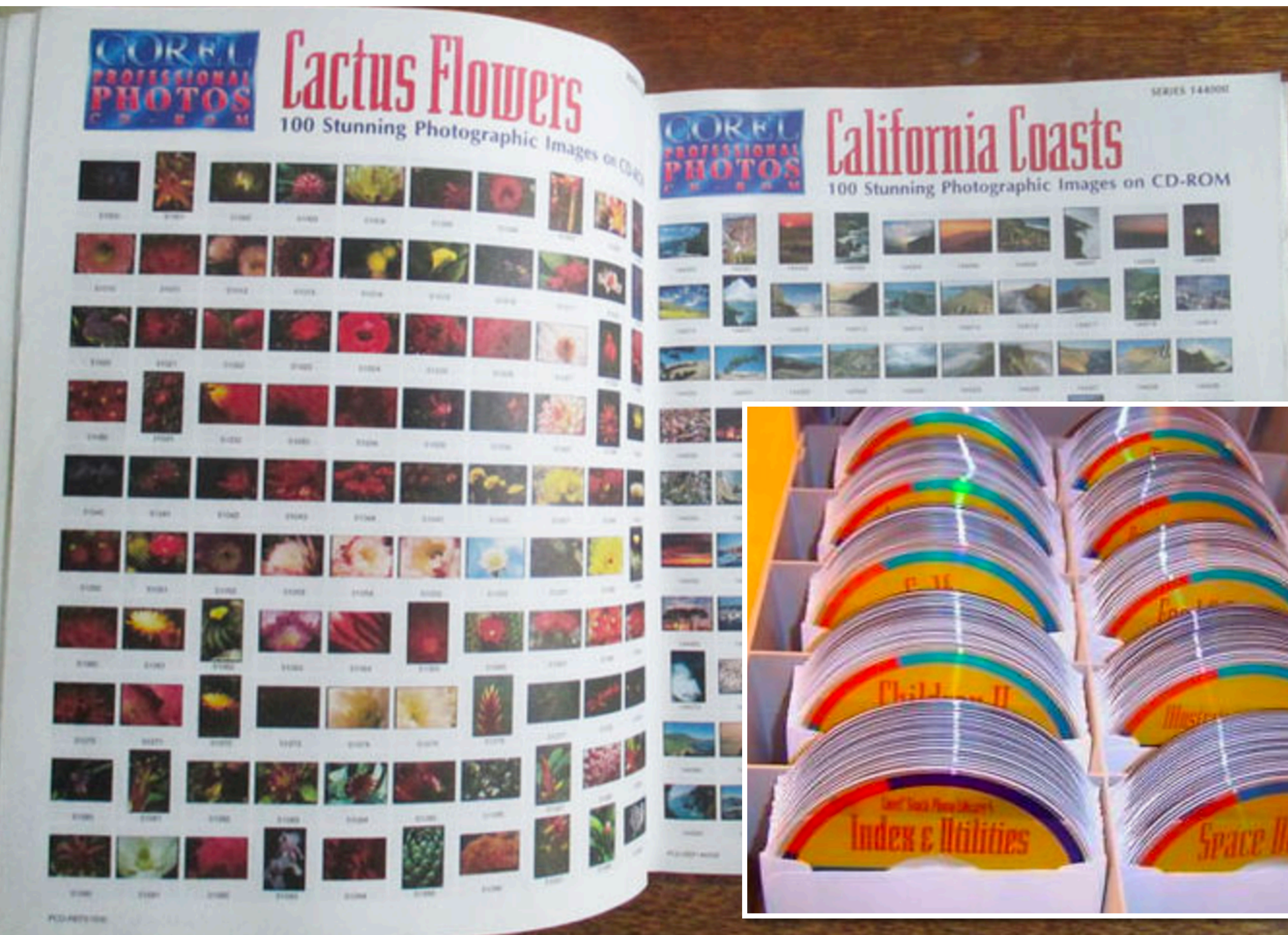
1996

Columbia University Image Library (COIL-100)

S. A. Nene, S. K. Nayar and H. Murase,



COREL image database



Caltech 101 and 256



Fei-Fei, Fergus, Perona, 2004



Griffin, Holub, Perona, 2007



Russell, Torralba, Freeman, 2005

The PASCAL Visual Object Classes



M. Everingham, Luc van Gool, C. Williams, J. Winn, A. Zisserman 2007

ImageNet classification and Neural nets



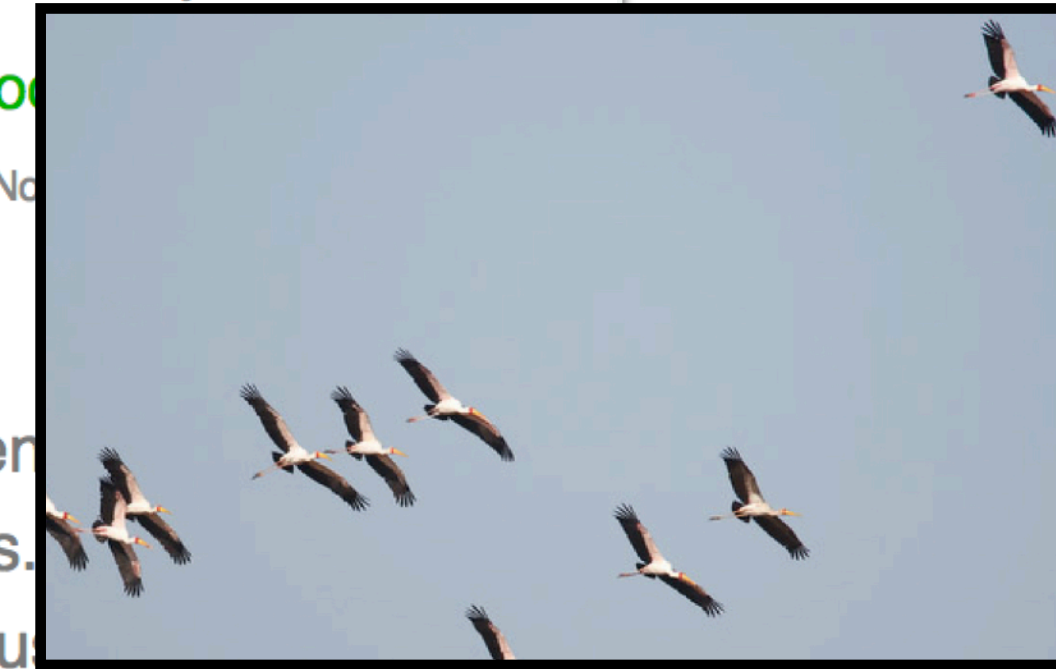
14,197,122 images, 21841 synsets indexed

Explore Download Challenges Publications Code

No

ImageNet is an image database organized according to the **WordNet** hierarchy (current in which each node of the hierarchy is depicted by hundreds and thousands of images. an average of over five hundred images per node. We hope ImageNet will become a u researchers, educators, students and all of you who share our passion for pictures.

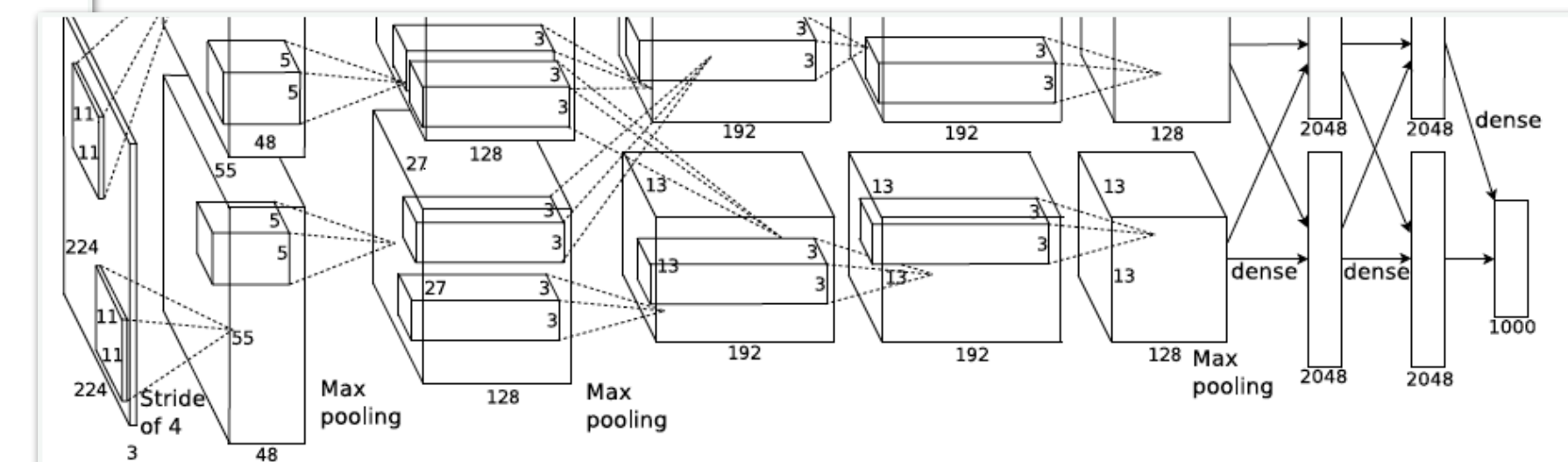
[Click here](#) to learn more about ImageNet, [Click here](#) to join the ImageNet mailing list.



“Birds”



What do these images have in common? *Find out!*



Crowdsourcing

amazonmechanical turk
Artificial Artificial Intelligence

Bryan C Russell | [Account Settings](#) | [Sign Out](#) | [Help](#)

Your Account | **HITS** | Qualifications | **56,035 HITS** available now

All HITS | HITS Available To You | HITS Assigned To You

Search for **HITS** containing that pay at least \$ **0.00** for which you are qualified ☐ **GO**

Timer: 00:00:13 of 60 minutes

Finished with this HIT? Let someone else do it?

☐ Automatically accept the next HIT

Total Earned: \$0.01
Total HITS Submitted: 12

LabelMe: Label objects in this image

Requester: Bryan C Russell
Qualifications Required: None

Reward: \$0.01 per HIT **HITS Available:** 269 **Duration:** 60 minutes

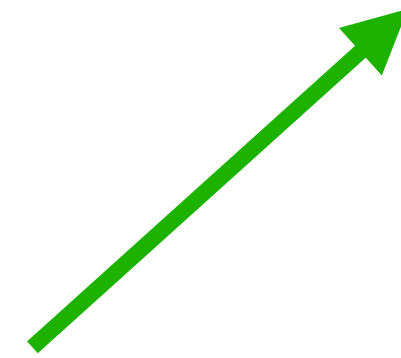
Please label as many objects as you want in this image. Scroll down to see the entire image.

Labeling for money
(Sorokin, Forsyth, 2008)

amazonmechanical turk
Artificial Artificial Intelligence



How can we collect good data?



- + Correctly labeled
- + Unbiased (good coverage of all relevant kinds of data)

Getting more humans in the annotation loop

Labeling to get a Ph.D.



Labeling for fun

Luis Von Ahn and Laura Dabbish 2004



Labeling for money
(Sorokin, Forsyth, 2008)



Labeling because it
gives you added value



Visipedia
(Belongie, Perona, et al)

Just for labeling



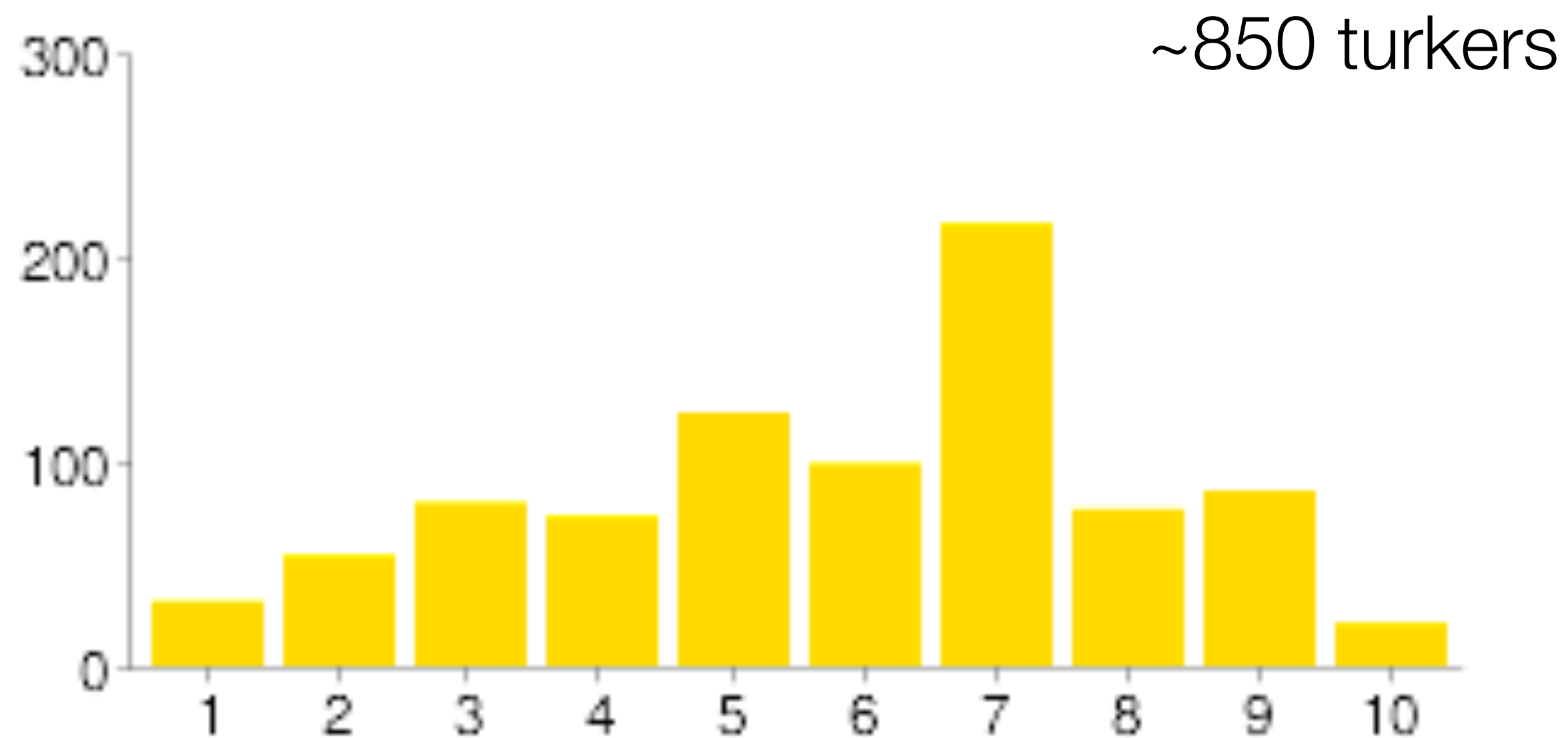
Beware of the human in your loop

- What do you know about them?
- Will they do the work you pay for?

Let's check a few simple experiments

People have biases...

Turkers were offered 1 cent to pick a number from 1 to 10.



Experiment by Greg Little

From <http://groups.csail.mit.edu/uid/deneme/>

Do humans have consistent biases?

Choose Item

Requester: SimpleSphere

Reward: \$0.01 per HIT

HITs Available: 1

Duration: 60 minutes

Qualifications Required: None

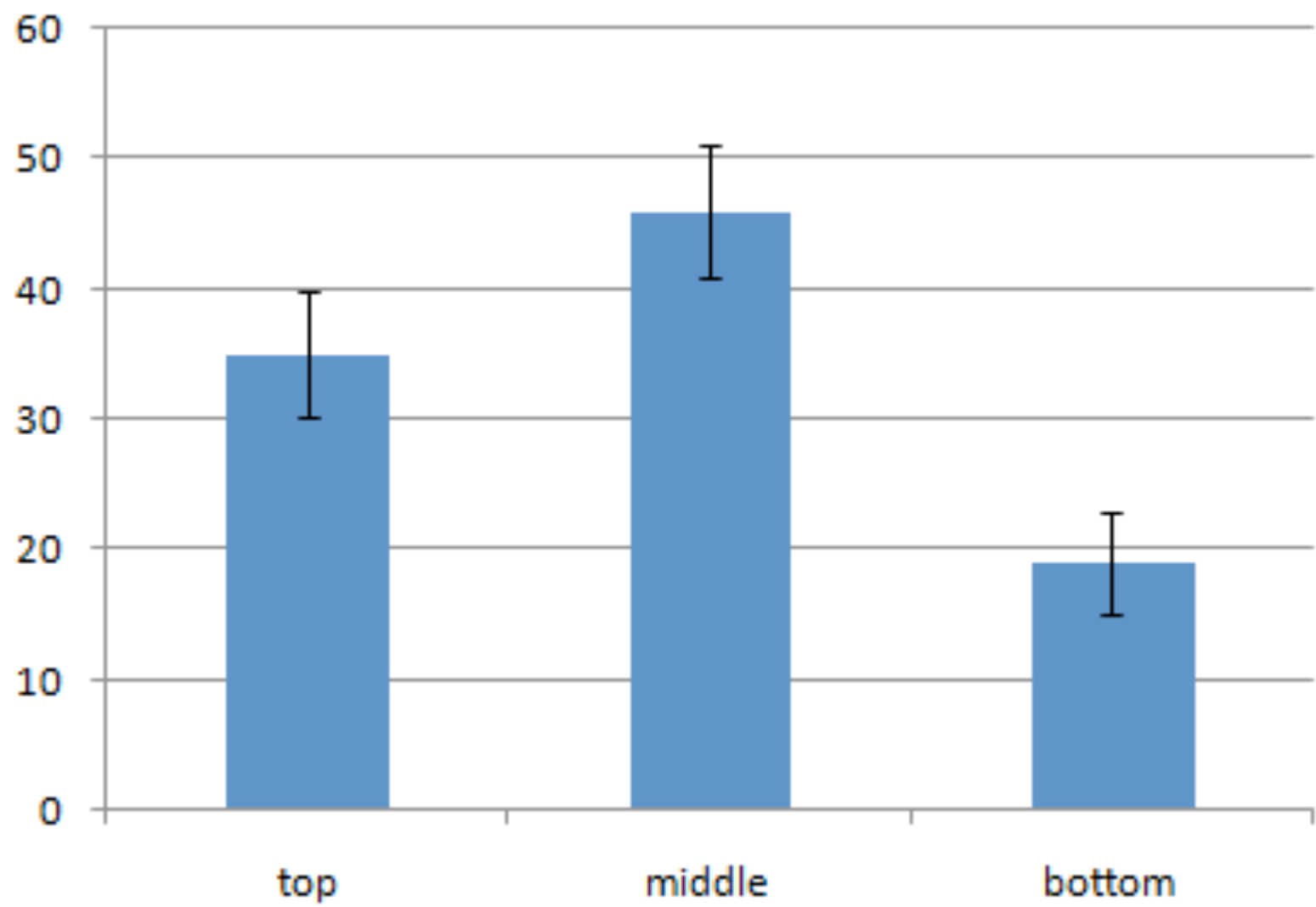
Please choose one of the following:

☐

☐

☐

Results form 100 HITS:



Experiment by Greg Little
From <http://groups.csail.mit.edu/uid/deneme/>

Do humans do what you ask for?

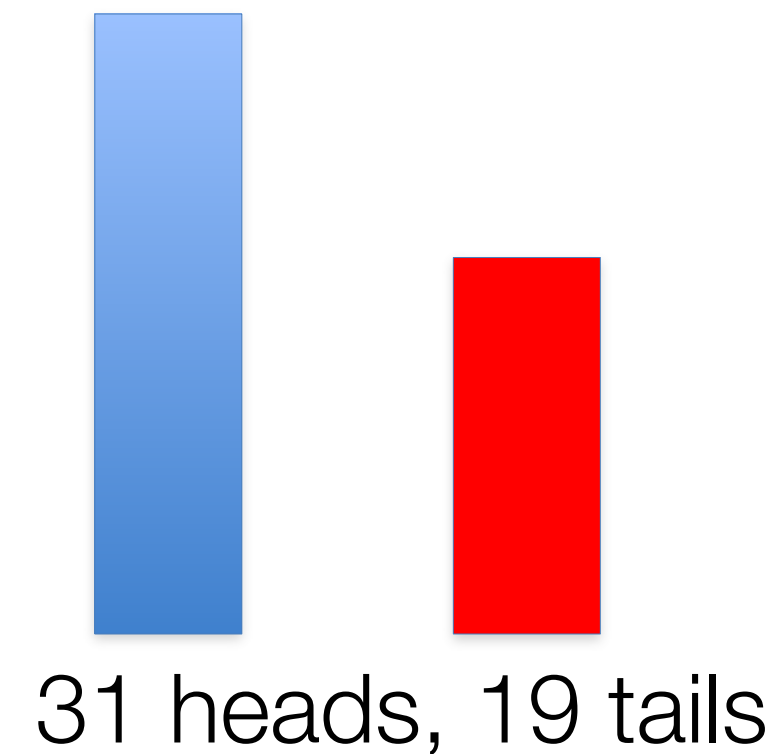
Flip a coin

Requester: ROBERT C MILLER Reward: \$0.01 per HIT HITs Available: 3 Duration: 5 minutes

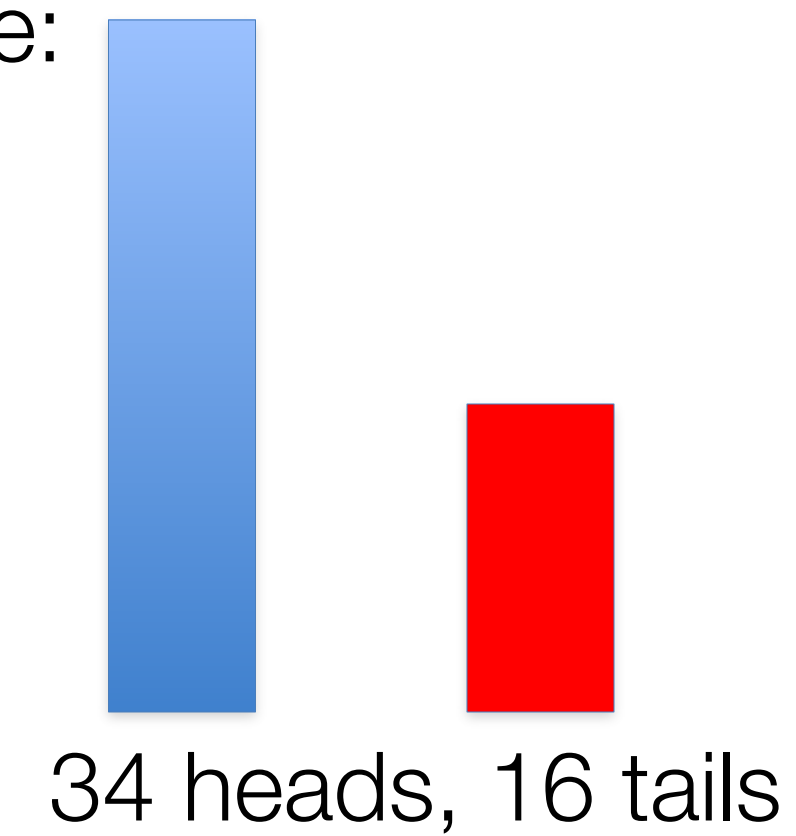
Qualifications Required: None

Please flip an actual coin and type either H or T below.

After 50 HITS:



And 50 more:



Experiment by Rob Miller

From <http://groups.csail.mit.edu/uid/deneme/>

Are humans reliable even in simple tasks?

Choose the given item.

Requester: SimpleSphere

Reward: \$0.01 per HIT

HITs Available: 1

Duration: 60 minutes

Qualifications Required: None

Please click button B:

B

C

A

Results of 100 HITS:

A: 2

B: 96

C: 2

Experiment by Greg Little

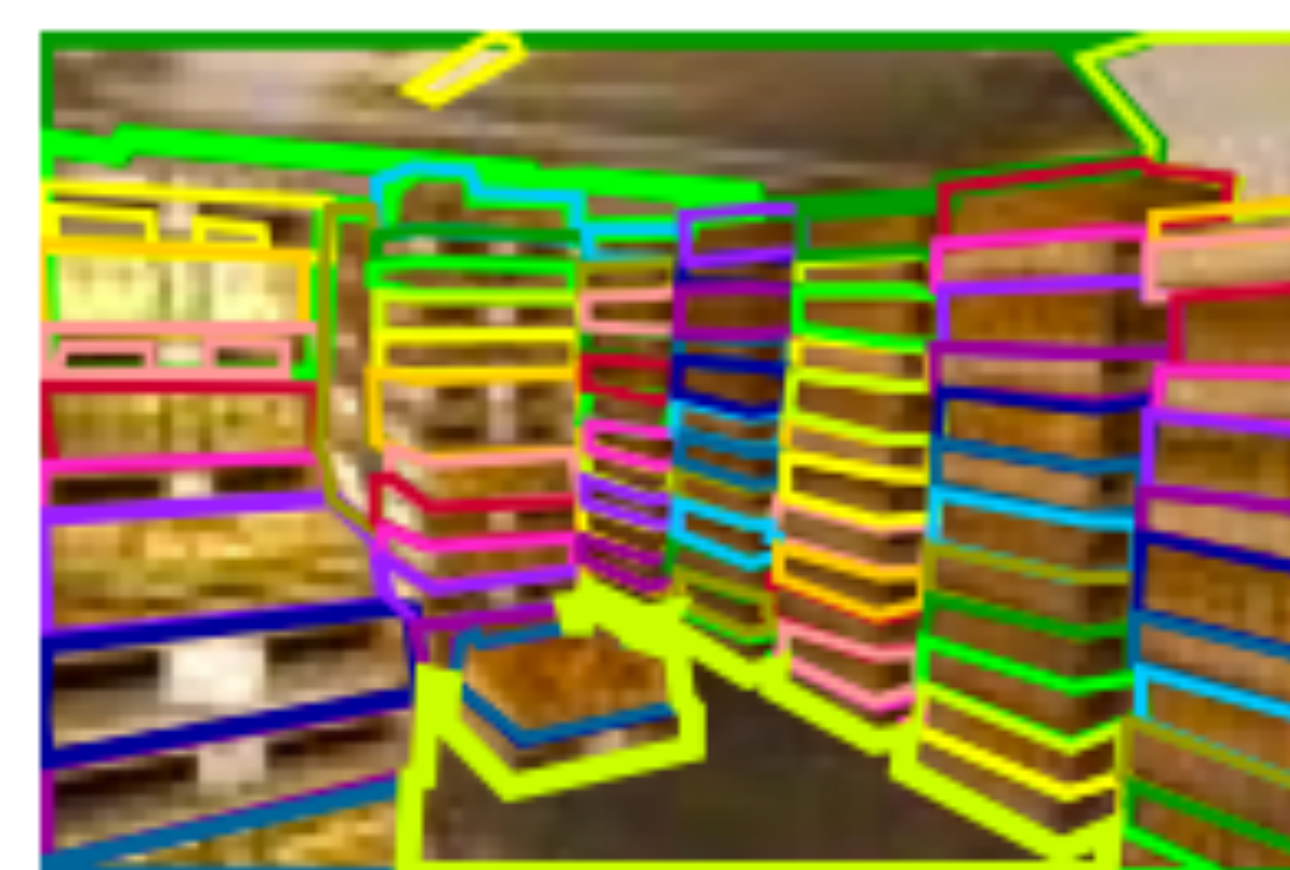
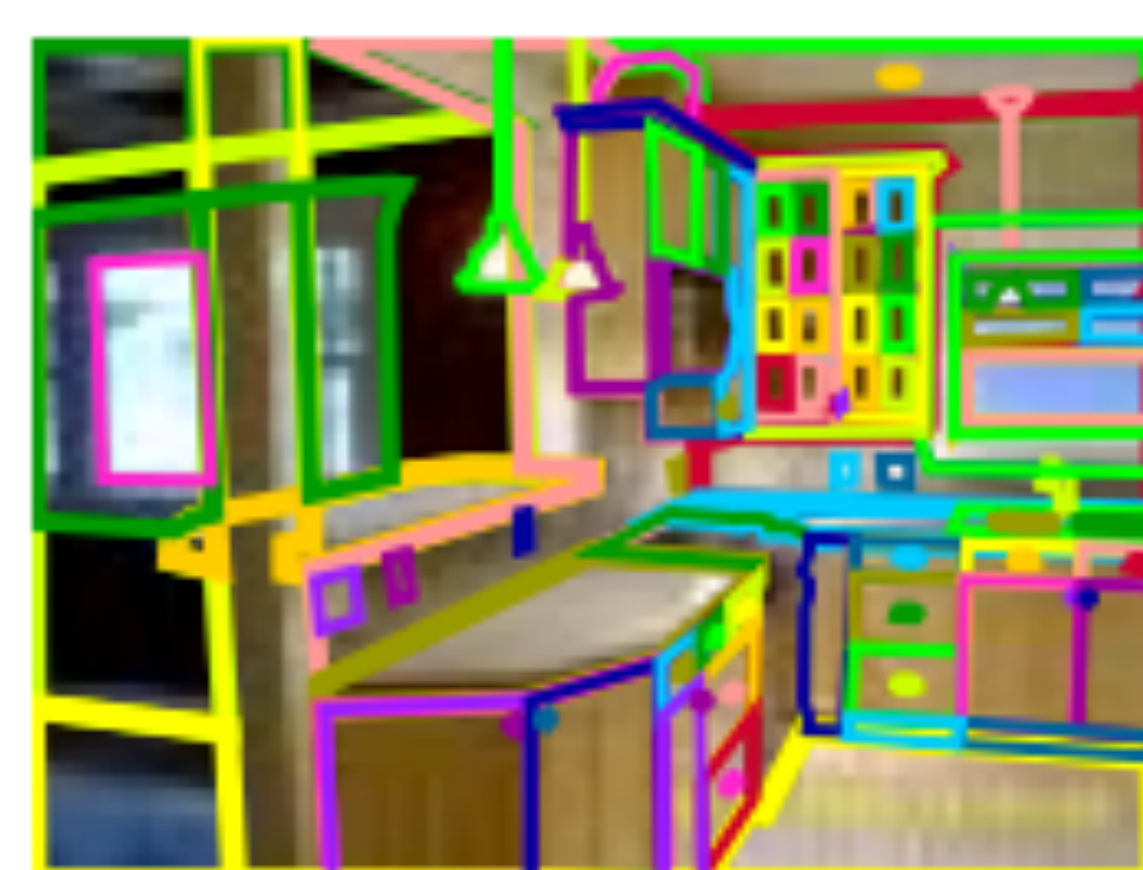
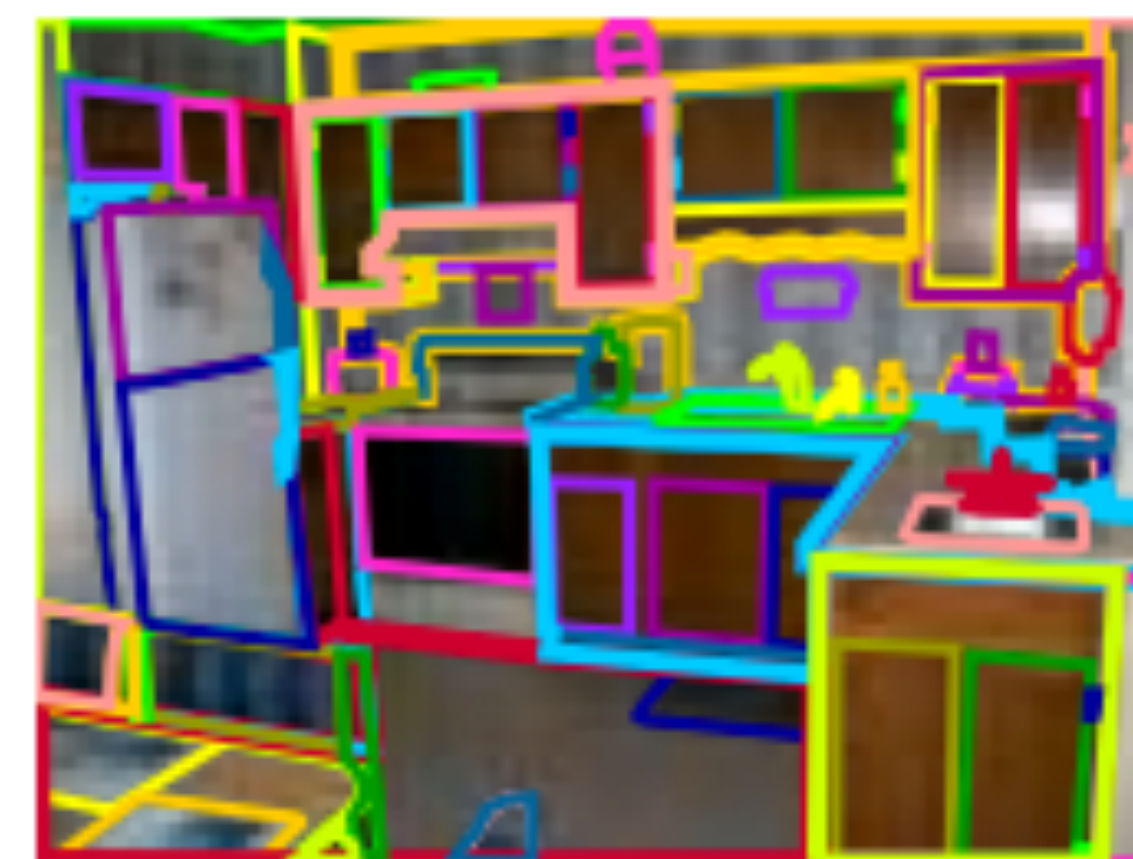
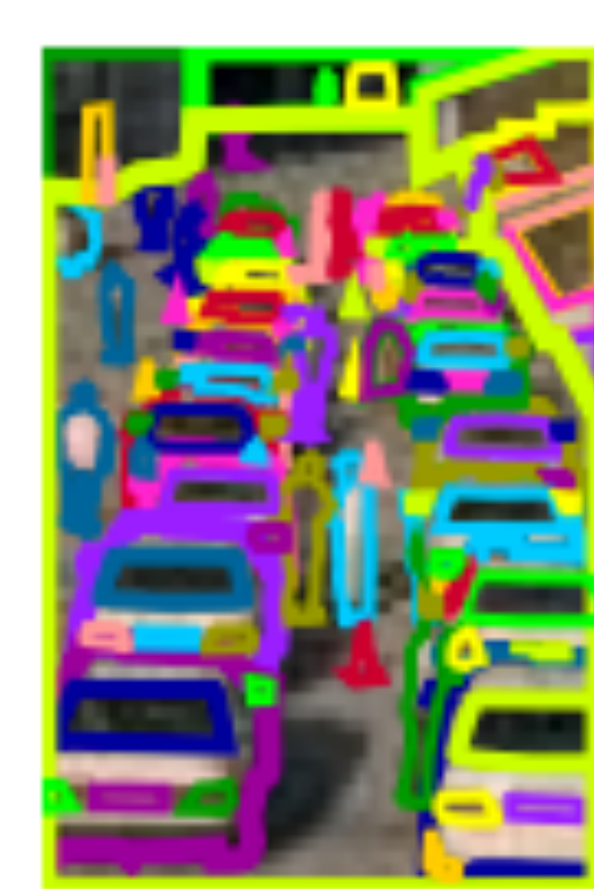
From <http://groups.csail.mit.edu/uid/deneme/>

Image annotation

<https://github.com/CSAILVision/LabelMeAnnotationTool>

<http://labelme.csail.mit.edu/Release3.0/>





...and 22.000 images

News

- Congratulations to the **winners** of the **ECCV 2018 Joint COCO and Mapillary Recognition Workshop**! Please visit the challenge website to view the winners and their talk slides.

What is COCO?



COCO is a large-scale object detection, segmentation, and captioning dataset. COCO has several features:

- ✓ **Object segmentation**
- ✓ **Recognition in context**
- ✓ **Superpixel stuff segmentation**
- ✓ **330K images (>200K labeled)**
- ✓ **1.5 million object instances**
- ✓ **80 object categories**
- ✓ **91 stuff categories**
- ✓ **5 captions per image**
- ✓ **250,000 people with keypoints**

Collaborators

Tsung-Yi Lin Google Brain

Genevieve Patterson MSR, Trash TV

Matteo R. Ronchi Caltech

Yin Cui Cornell Tech

Michael Maire TTI-Chicago

Serge Belongie Cornell Tech

Lubomir Bourdev WaveOne, Inc.

Ross Girshick FAIR

James Hays Georgia Tech

Pietro Perona Caltech

Deva Ramanan CMU

Larry Zitnick FAIR

Piotr Dollár FAIR

Sponsors

**CVDF****Microsoft****facebook****Mighty Ai**

Tasks

Image classification: assign one (or few) tags to each image.

Caltech 101

http://www.vision.caltech.edu/Image_Datasets/Caltech101/

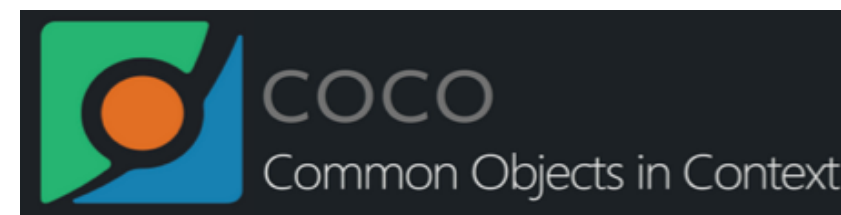
IMAGENET

<http://www.image-net.org/>



<http://places2.csail.mit.edu/>

Semantic segmentation: assign one (or few) tags to each pixel.



<http://cocodataset.org/#home>



<https://groups.csail.mit.edu/vision/datasets/ADE20K/>



<https://www.cityscapes-dataset.com/>

Action and Activity datasets

Action Datasets



Soomro et al. CoRR 2012

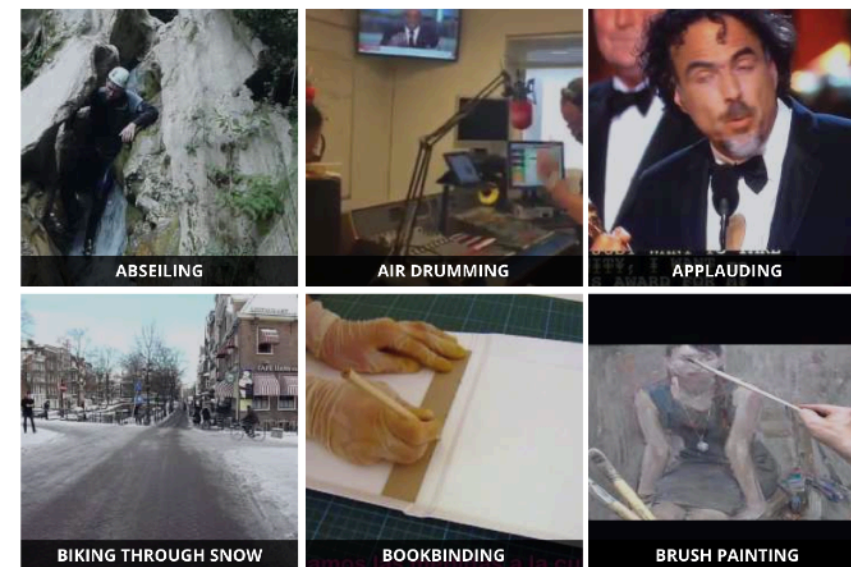


Left: Stand, Carry/Hold, Listen to; Middle: Stand, Carry/Hold, Talk to; Right: Sit, Write

Gu et al. Arxiv:1705.08421



Monfort et al. 2017

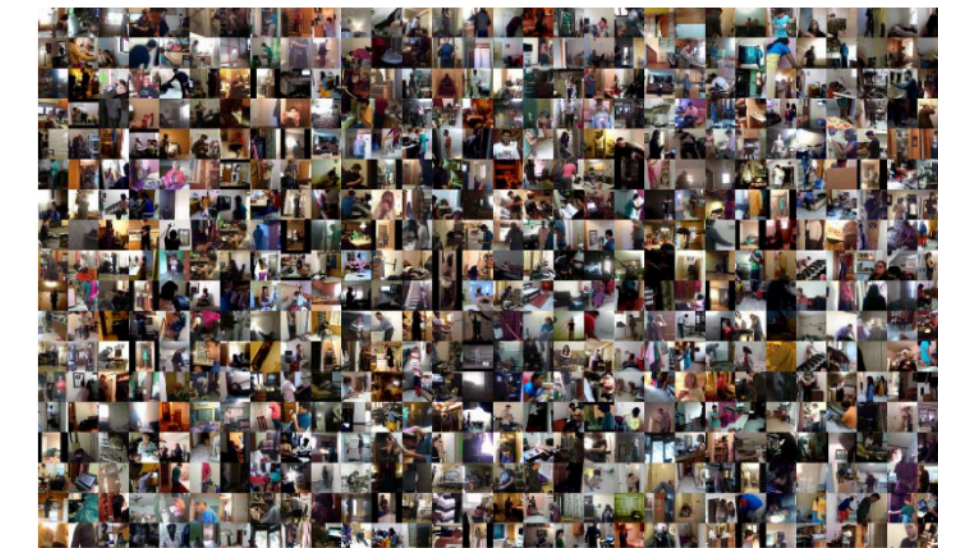


Carreira et al. CVPR 2017

Activity Datasets



Heilbron et al. CVPR 2015



Sigurdsson et al. ECCV 2016



Xu et al. CVPR 2013



Fouhey, et al. CVPR 2018

Beyond labeled datasets

A young child with light brown hair, wearing a white short-sleeved shirt with a small pattern and dark pants, is sitting on a light-colored carpeted floor. The child is looking down at their hands, which are resting on the carpet. The background shows a wooden floor and a dark object in the upper left corner.

Vision

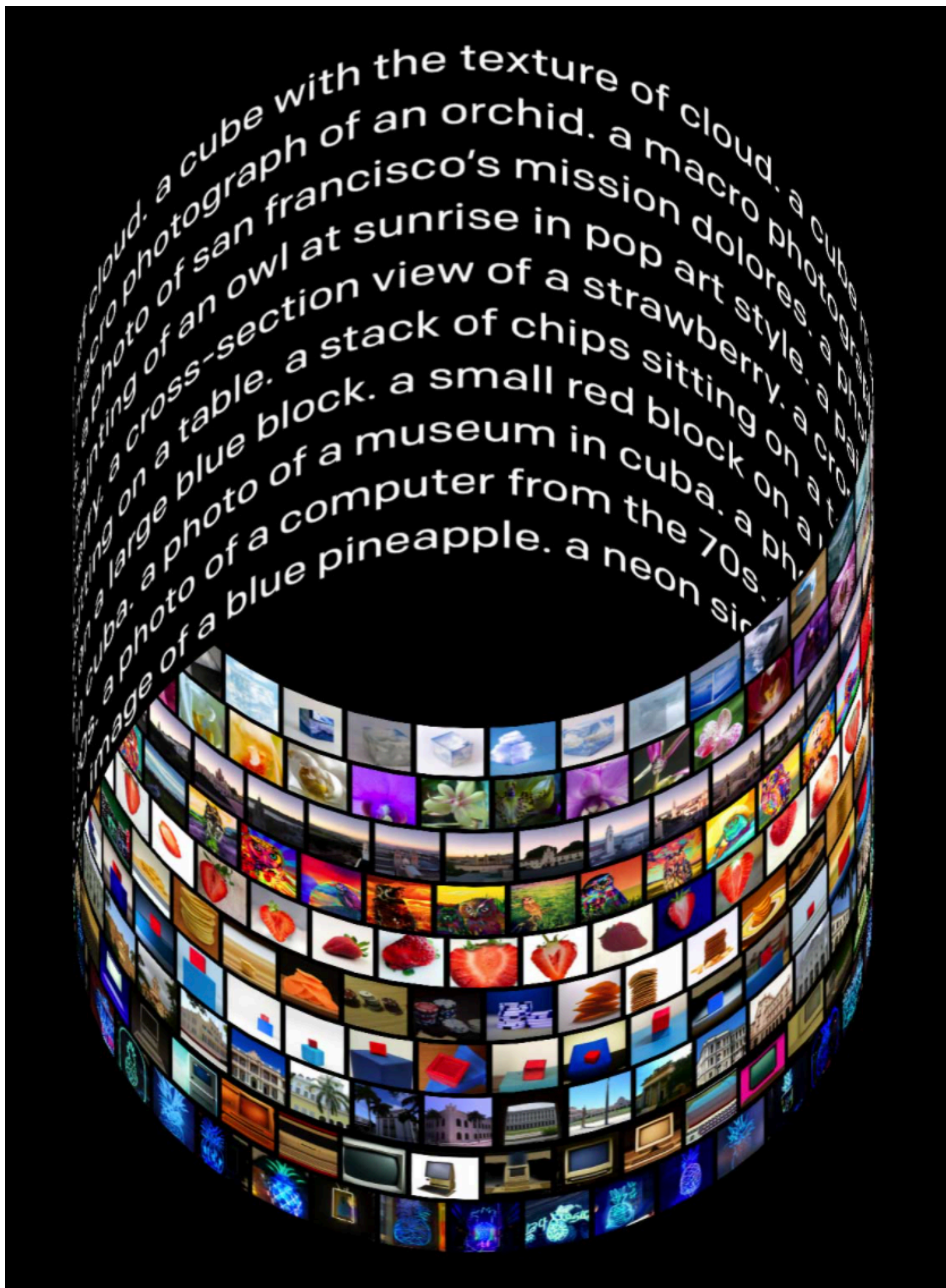
Audition

Smell

Taste

Self-supervised system

Touch



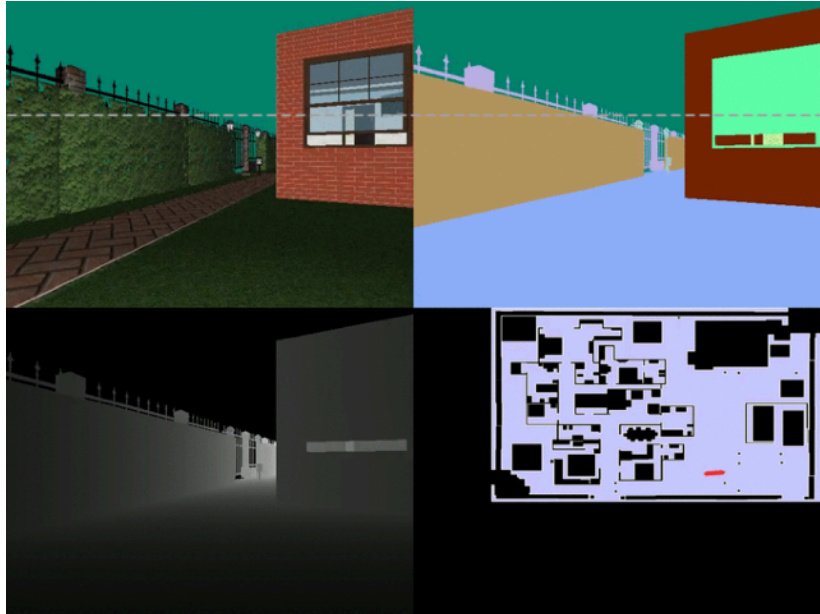
Recent trend has been to just scrape massive, raw data — images, text, etc — from the Internet, and distill knowledge from the web.

DALL-E [<https://openai.com/blog/dall-e/>]

CLIP [<https://openai.com/blog/clip/>]

Virtual Environments

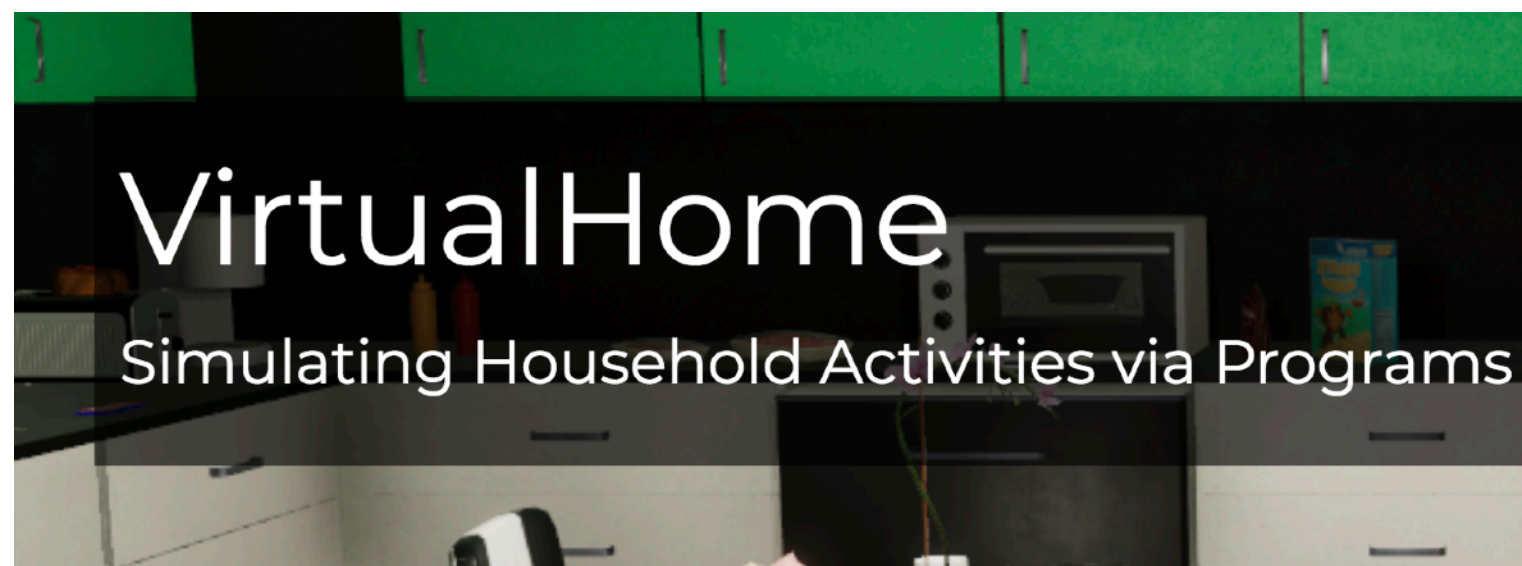
Navigation



Wu et al. 2018



Savva et al. 2017



Puig et al, 2018

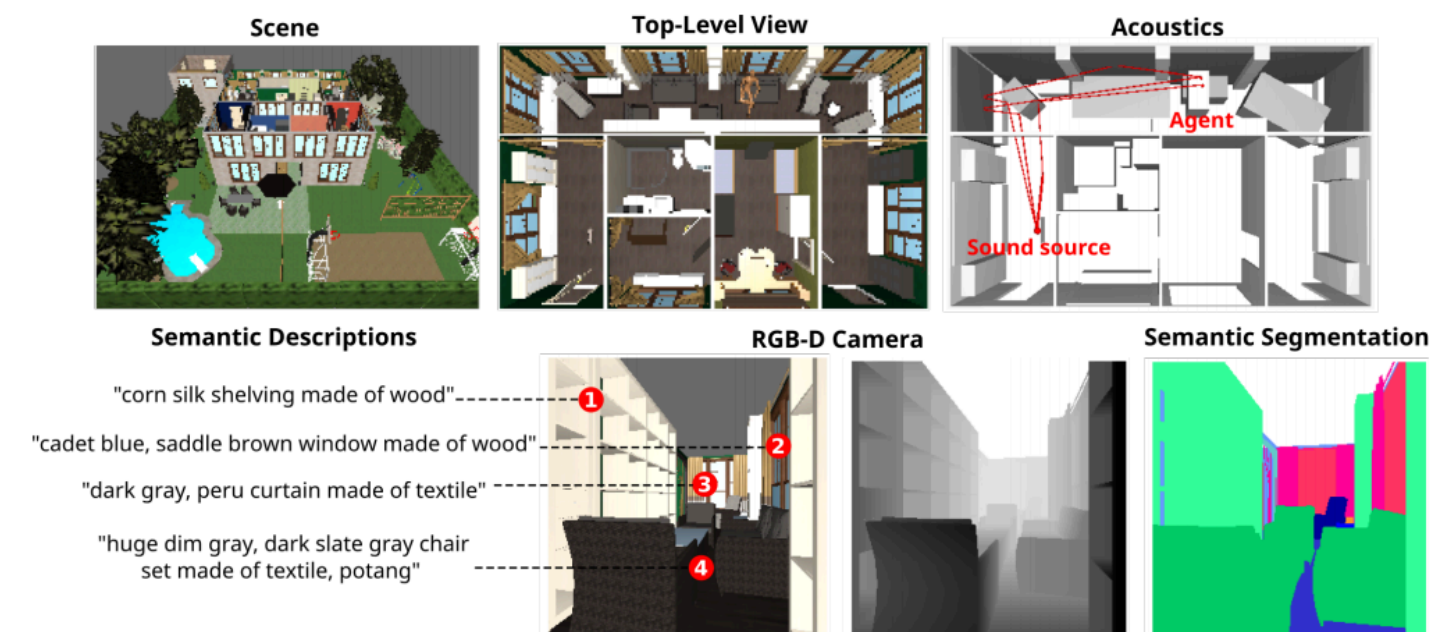


Savva et al. 2019

Interactions

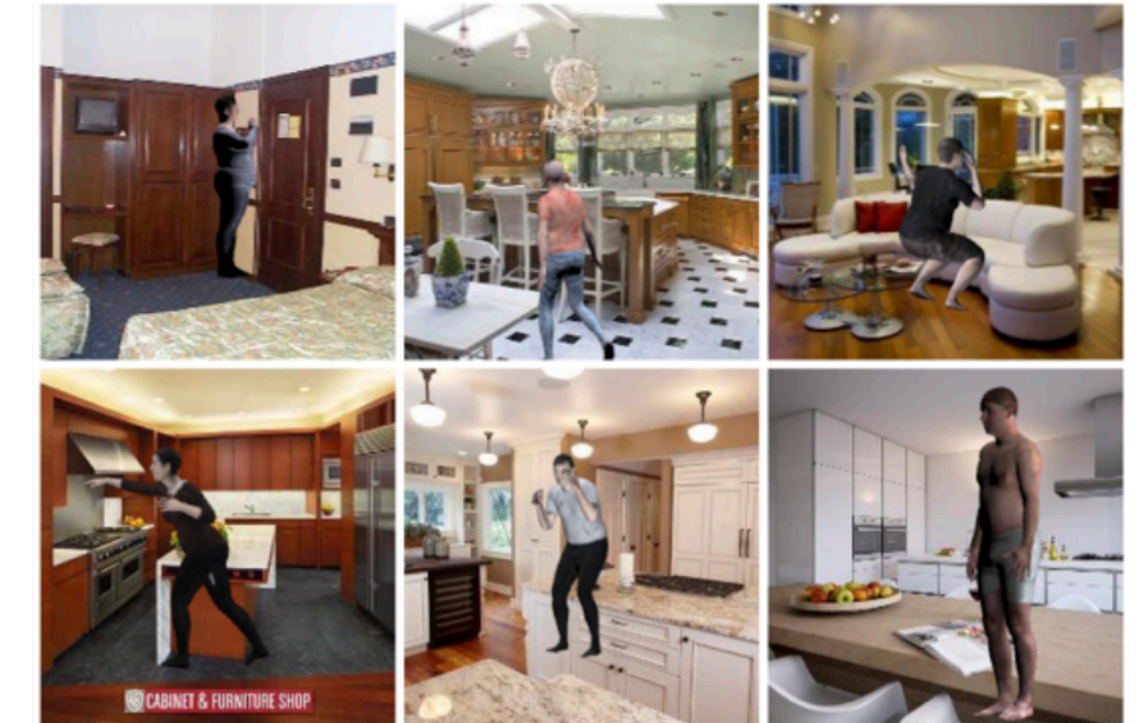


Kolve et al. 2017



Brodeur et al. 2017

Action Recognition



Varol et al. 2017



De Souza et al. 2017

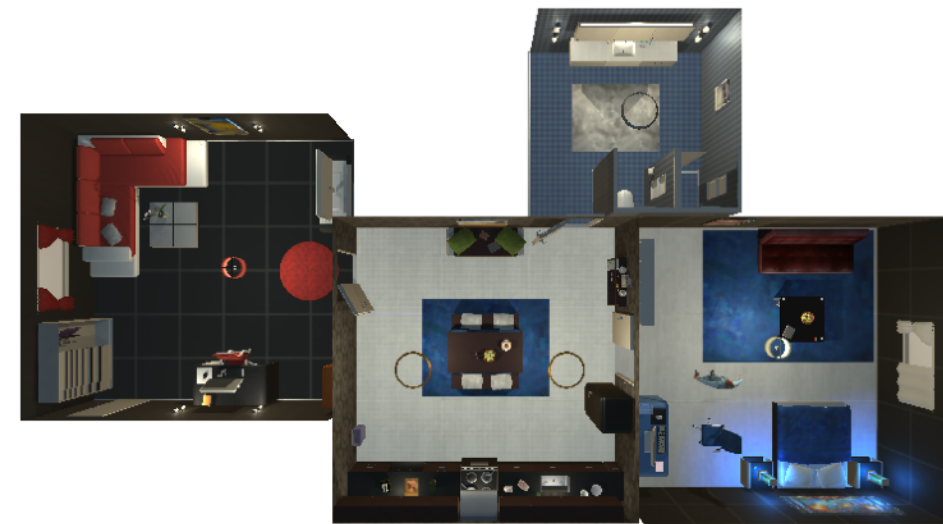
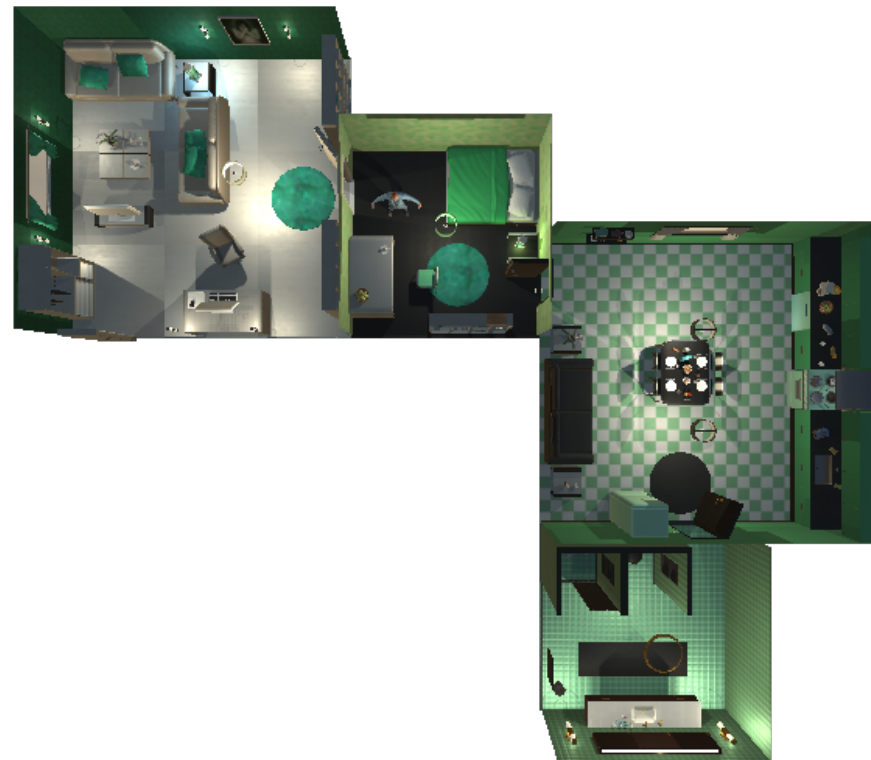
VirtualHome

Simulating Household Activities via Programs



VirtualHome: Simulating Household Activities via Programs. X. Puig*, K. Ra*, M. Boben*, J. Li, T. Wang, S. Fidler, A. Torralba. CVPR 2018.

Apartments



Animated characters



Interactive objects



I go grab a drink in the kitchen and then go watch the news on TV. I tweet about the news

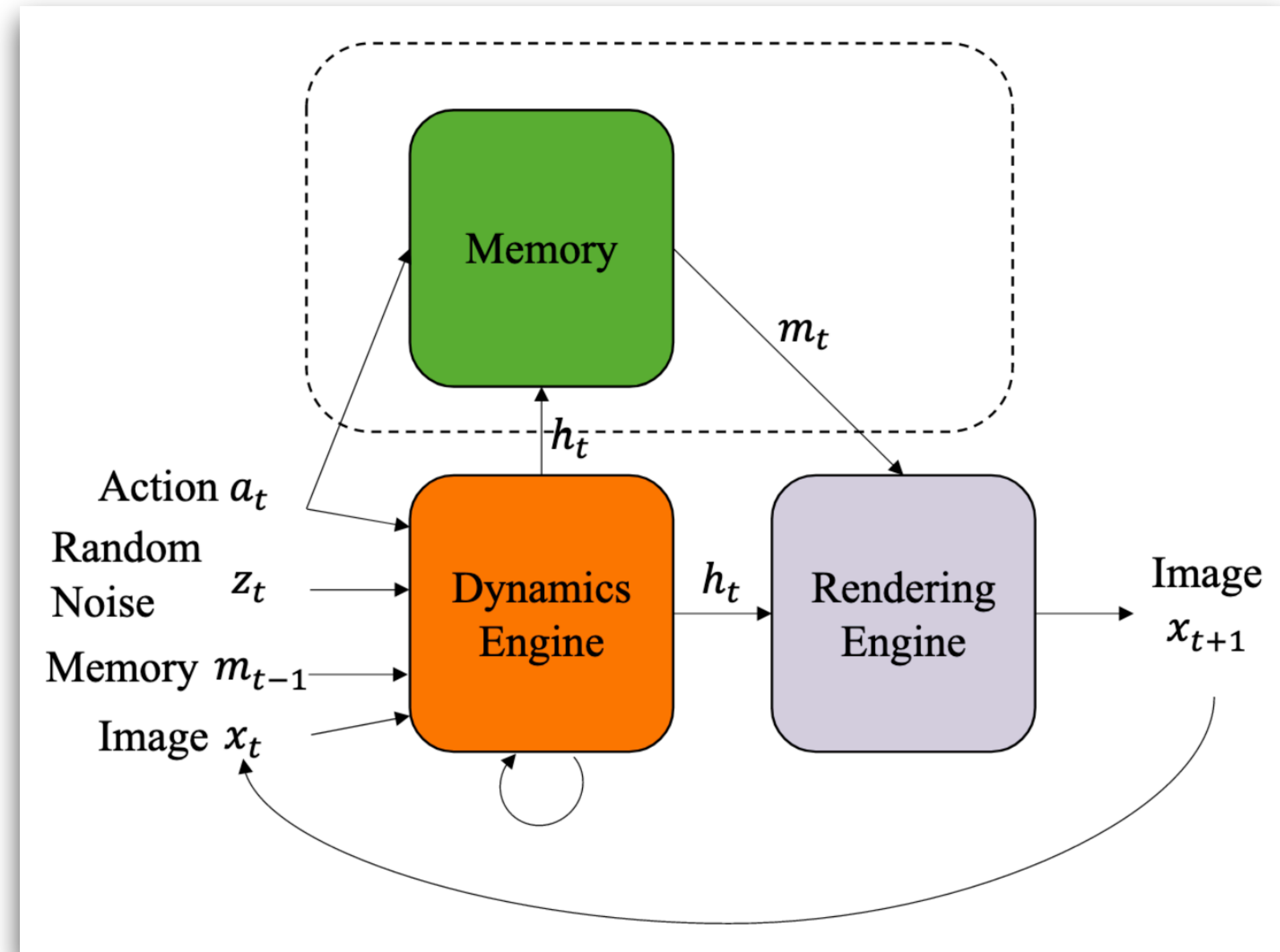
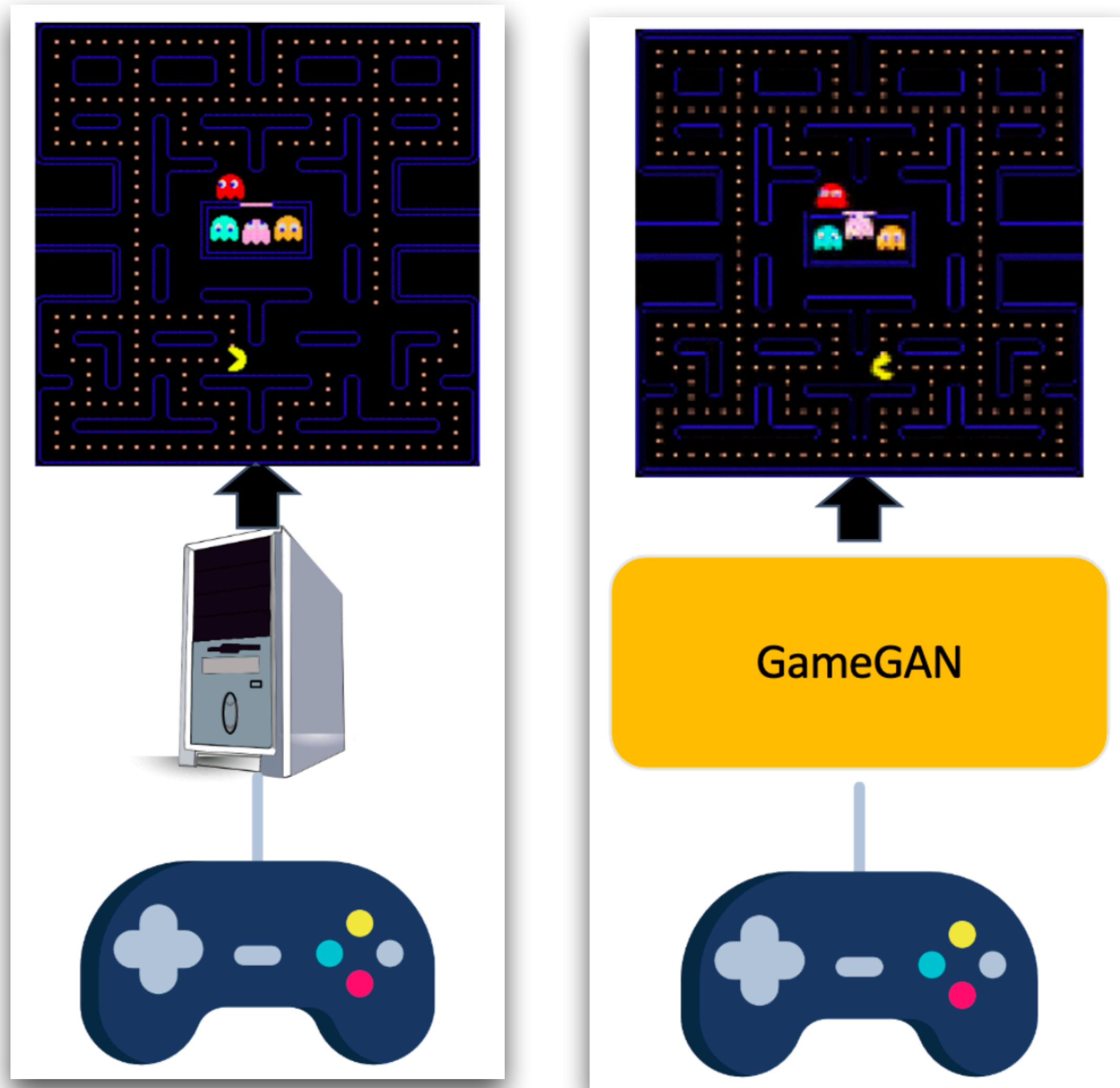




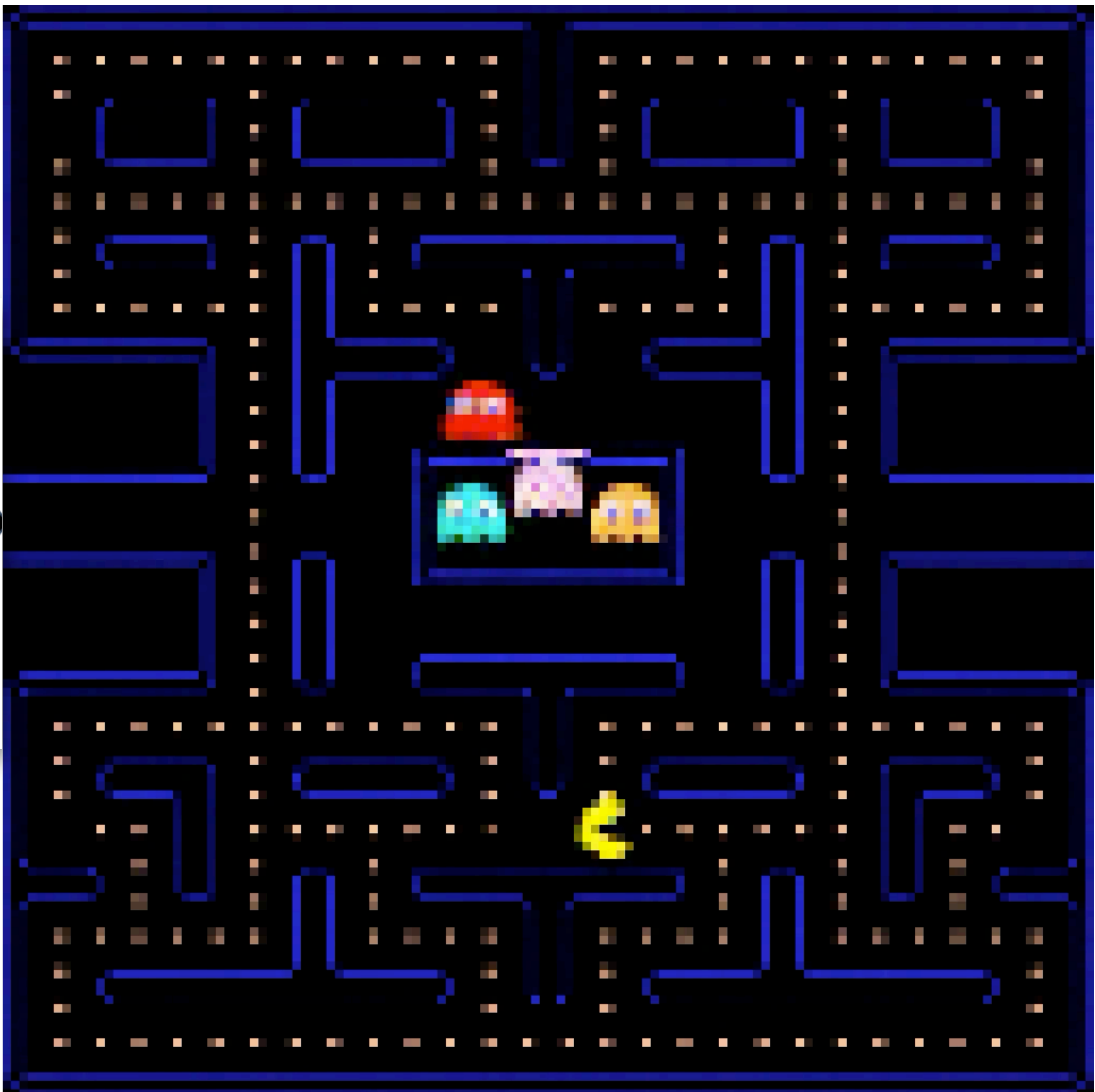
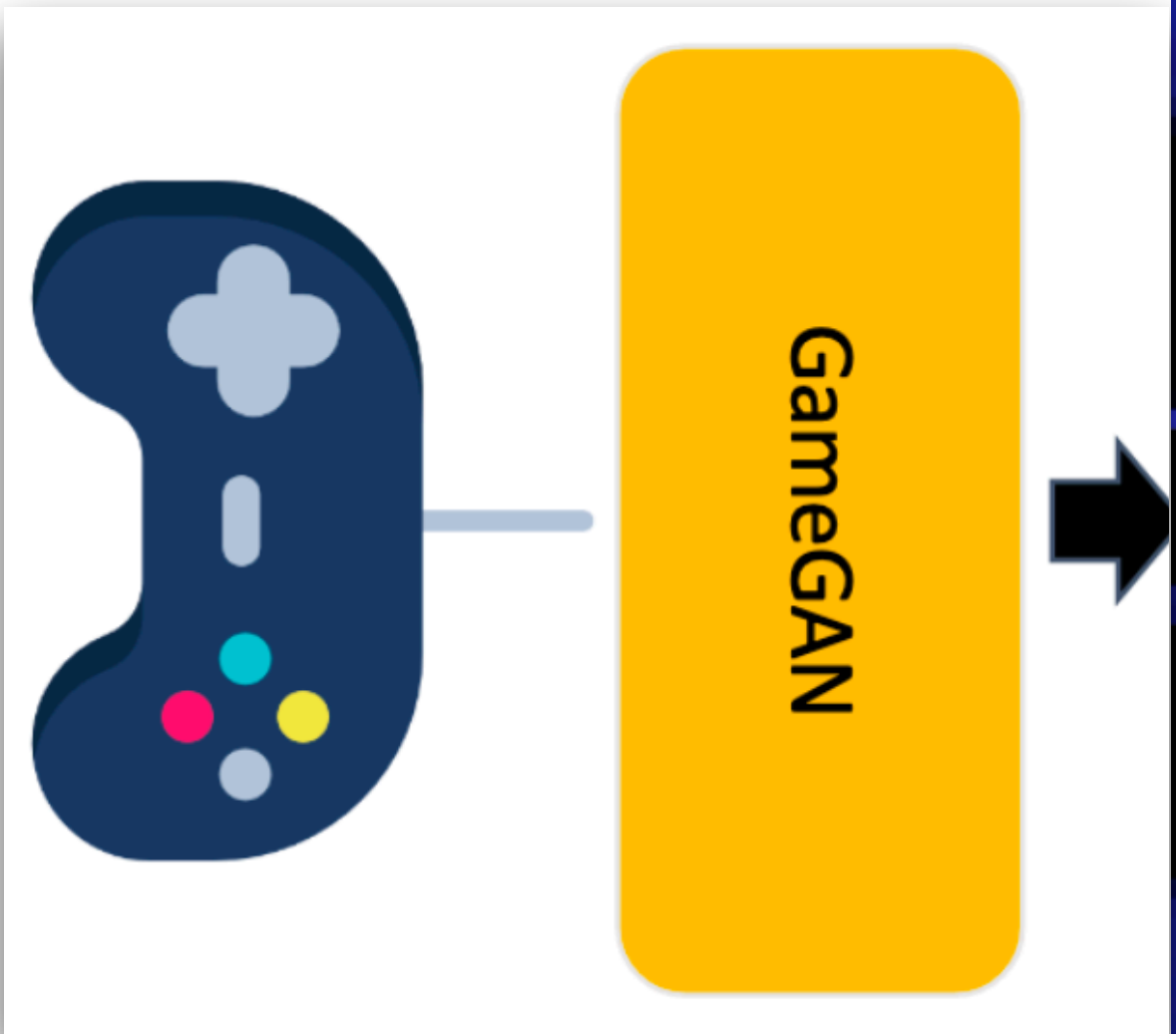
Need for Common Sense Knowledge



Learning with GANs



<https://arxiv.org/pdf/2005.12126.pdf>



Dataset bias

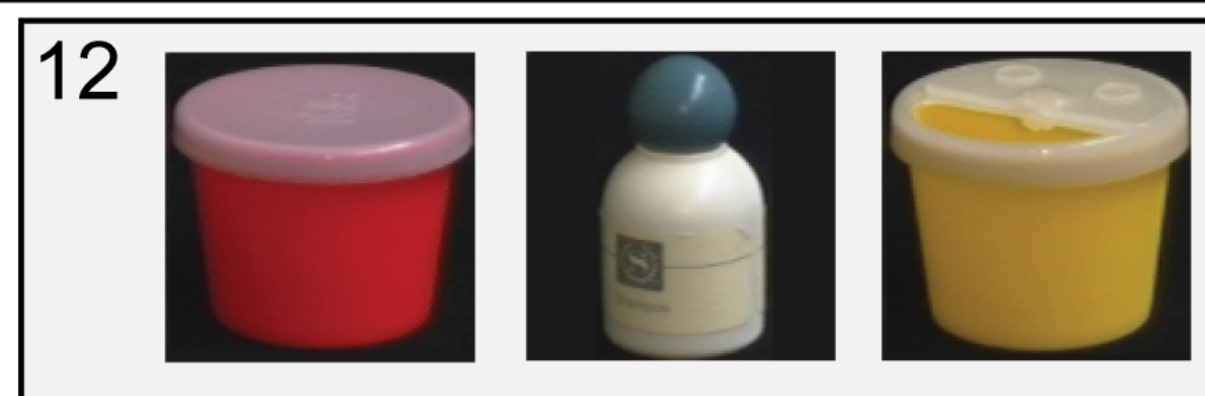
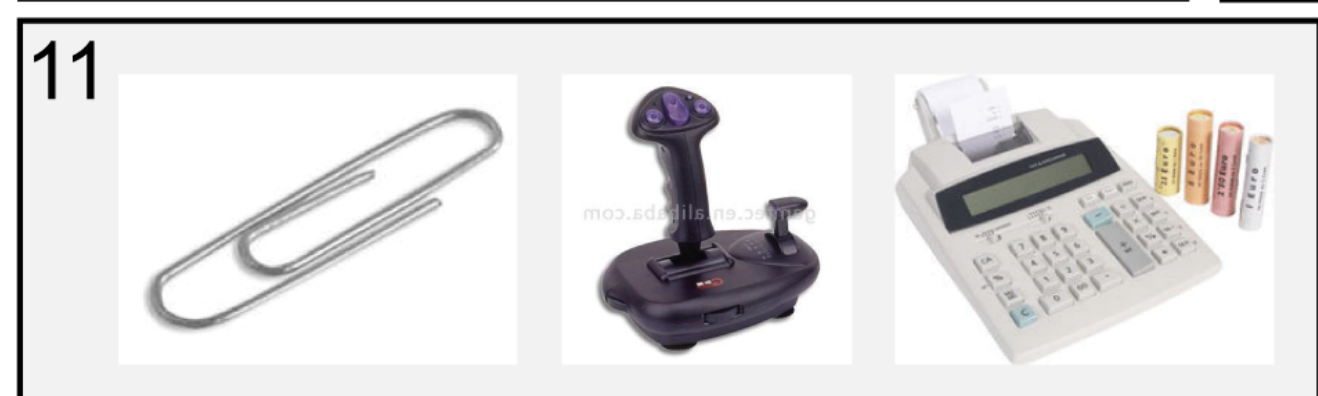
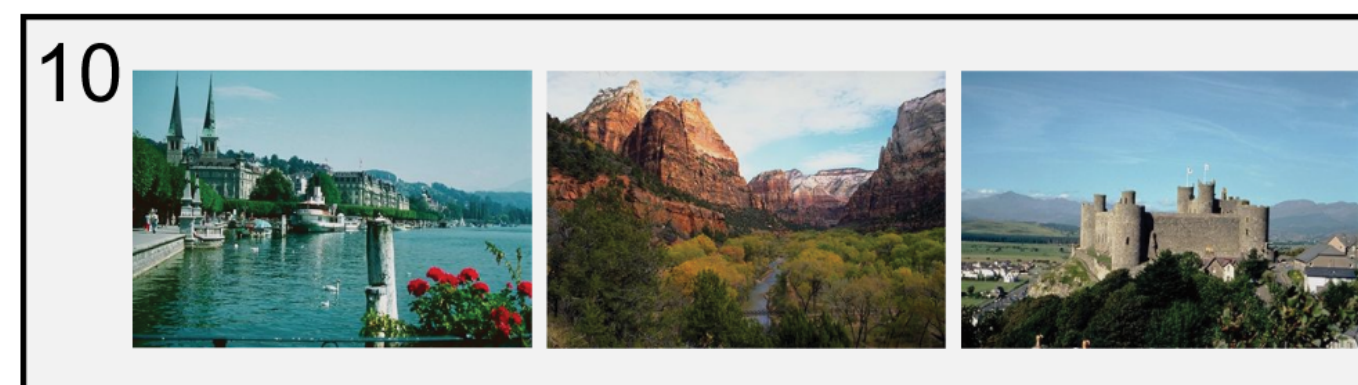
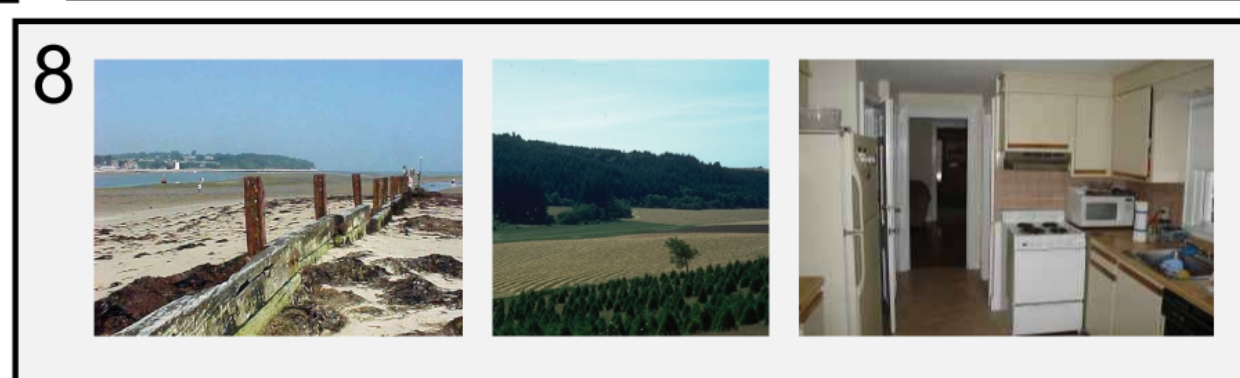
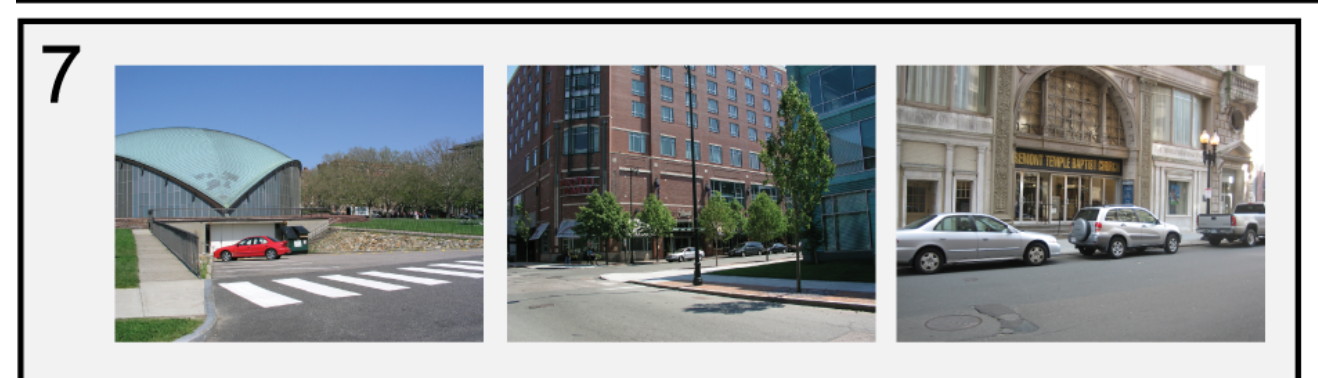
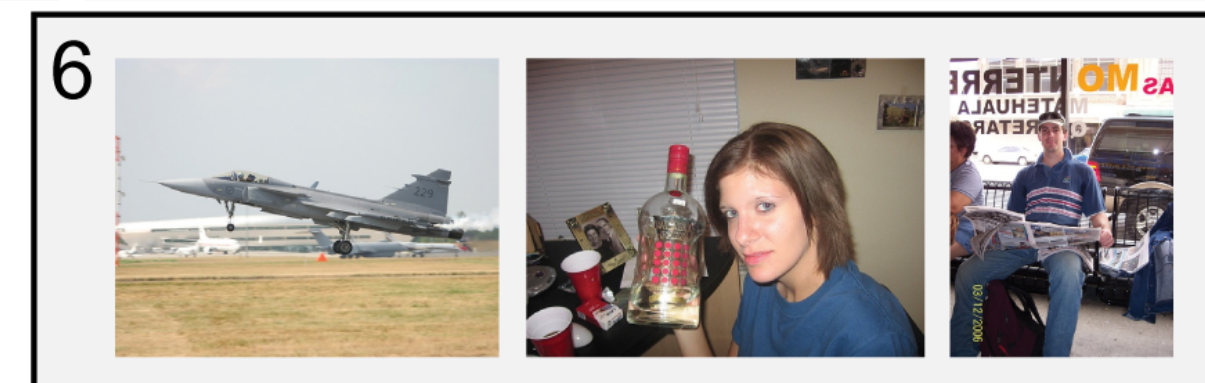
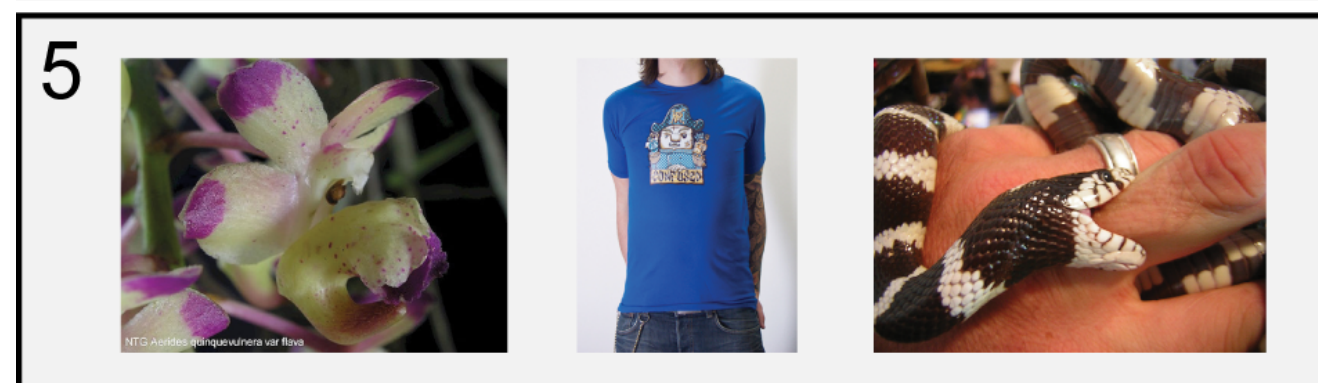
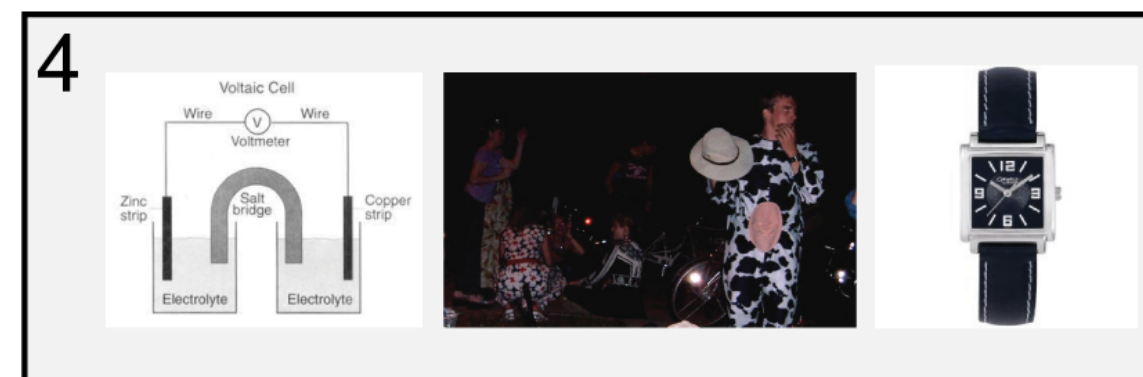
Unbiased Look at Dataset Bias

Alyosha Efros (CMU)
Antonio Torralba (MIT)



Disclaimer: no graduate students have been harmed in the production of this paper

“Name That Dataset!” game



- Caltech 101
- Caltech 256
- MSRC
- UIUC cars
- Tiny Images
- Corel
- PASCAL 2007
- LabelMe
- COIL-100
- ImageNet
- 15 Scenes
- SUN'09



bedroom





bedroom



Search

About 299,000,000 results (0.19 seconds)



SafeSearch off ▾



Everything

Images

Maps

Videos

News

Shopping

More

Related searches: [bedroom designs](#) [master bedroom](#) [modern bedroom](#) [simple bedroom](#) [small bedroom](#)



Any time

Past 24 hours

Past week

Custom range...

All results

By subject

Personal



Any size

Large

Medium

Icon

Larger than...

Exactly...





student bedroom



Search

About 66,700,000 results (0.15 seconds)



SafeSearch off ▾



Everything

Images

Maps

Videos

News

Shopping

More

Any time

Past 24 hours

Past week

Custom range...

All results

By subject

Personal

Any size

Large

Medium

Icon

Larger than...

Exactly...

Any color

Full color





www.bigstock.com - 7067629



Google

mug



Training data

What Google thinks are
student bedrooms

Google

student bedroom

Search

About 66,700,000 results (0.15 seconds)

Everything

Images

Maps

Videos

News

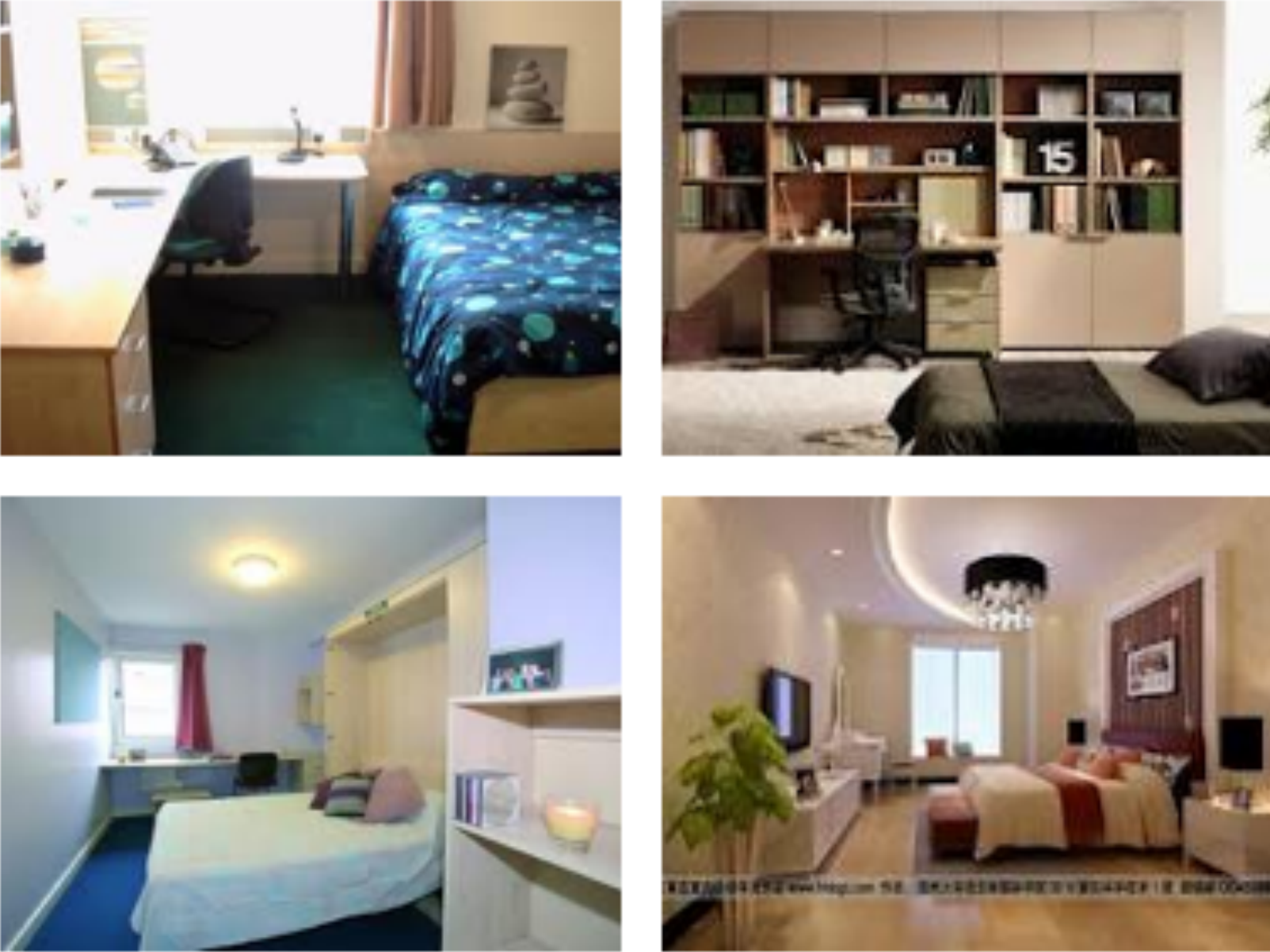
Shopping

More

Any time

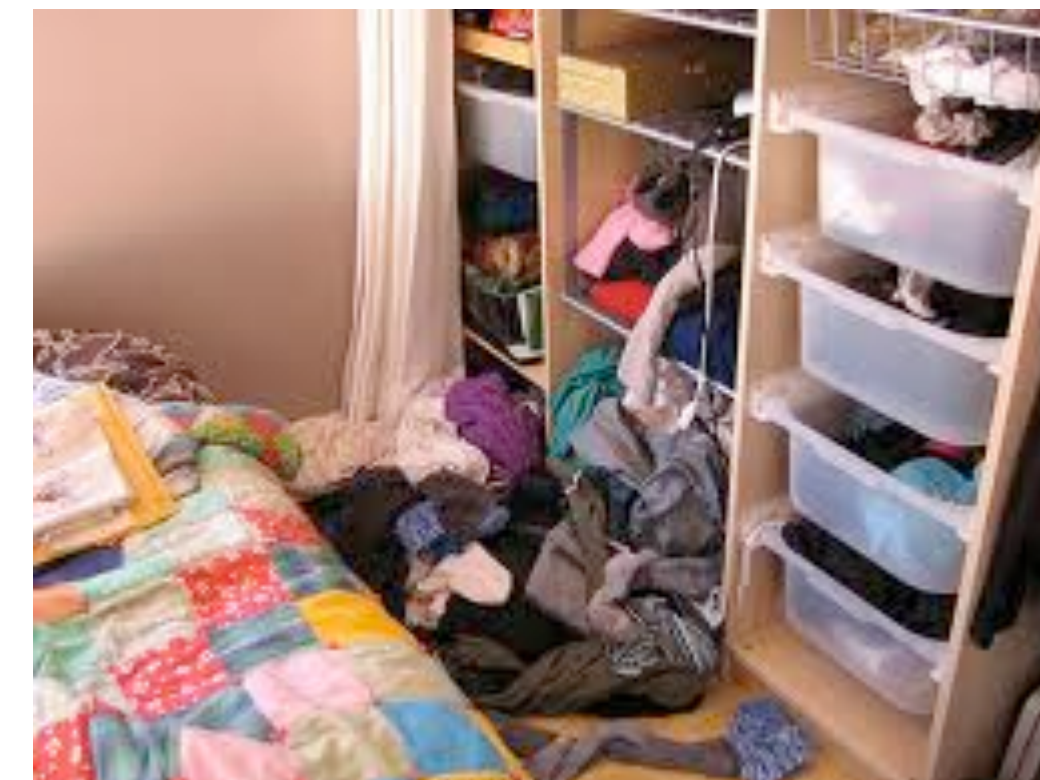
Past 24 hours

Past week



The training data consists of four images of student bedrooms. The top-left image shows a room with a desk, chair, and a bed with a blue patterned coverlet. The top-right image shows a room with a large bookshelf, a desk, and a bed. The bottom-left image shows a room with a desk, chair, and a bed with a white coverlet. The bottom-right image shows a room with a desk, chair, and a bed with a red and white coverlet.

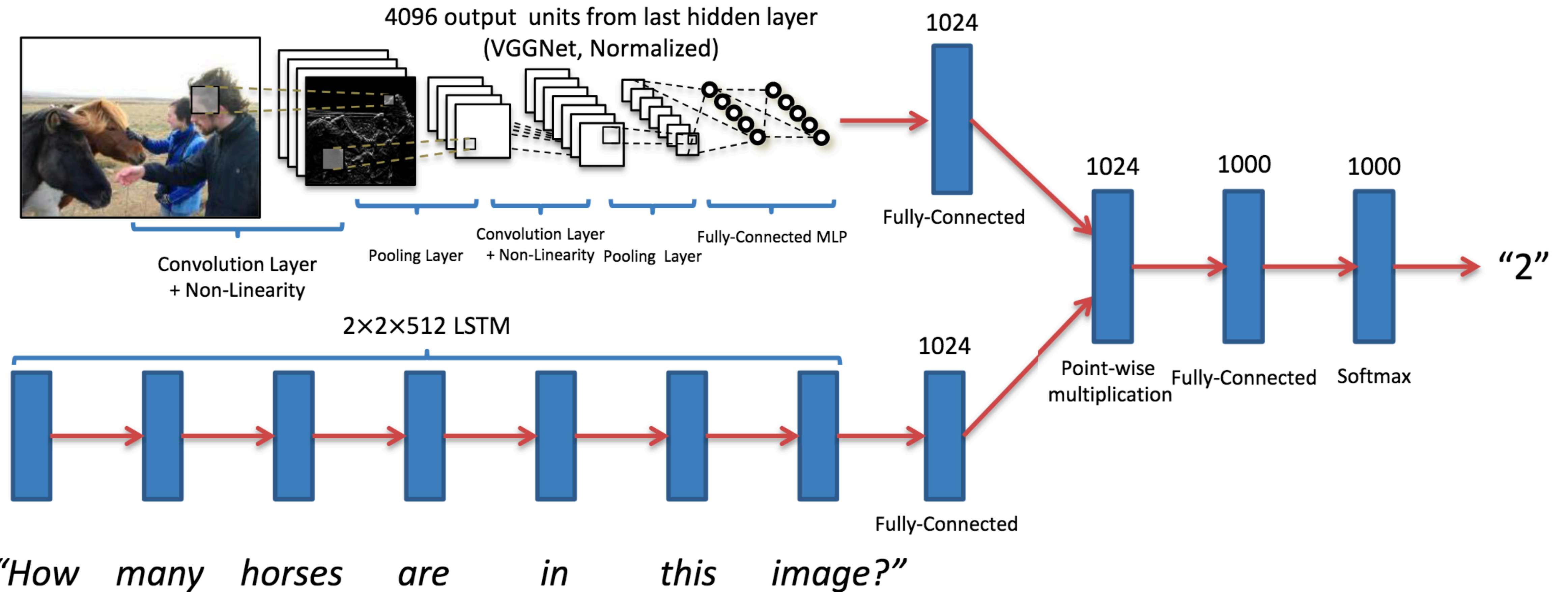
Test data



Visual Question Answering



Architecture



There are 1000 possible answers in this system. Questions are unlimited.



what is on the ground?

Submit

Predicted top-5 answers with confidence:

sand

90.748%

snow

2.858%

beach

1.418%

surfboards

0.677%

water

0.528%



what color is the umbrella?

Submit

Predicted top-5 answers with confidence:

yellow

95.090%

white

1.811%

black

0.663%

blue

0.541%

gray

0.362%



are we alone in the universe?

Submit

Predicted top-5 answers with confidence:

no

78.234%

yes

21.763%

people

0.001%

birds

0.000%

out

0.000%



what is the meaning of life?

Submit

Predicted top-5 answers with confidence:

beach

15.262%

sand

8.537%

seagull

4.708%

tower

2.393%

rocks

1.746%



what is the yellow thing?

Submit

Predicted top-5 answers with confidence:

frisbee

79.844%

surfboard

7.319%

banana

2.844%

lemon

2.438%

surfboards

1.252%



how many trains are in the picture?

Submit

Predicted top-5 answers with confidence:

3

30.233%

5

18.270%

4

17.000%

2

11.343%

6

7.806%

Of number questions (e.g. “how many...”), 26.04% of the time, the answer is 2

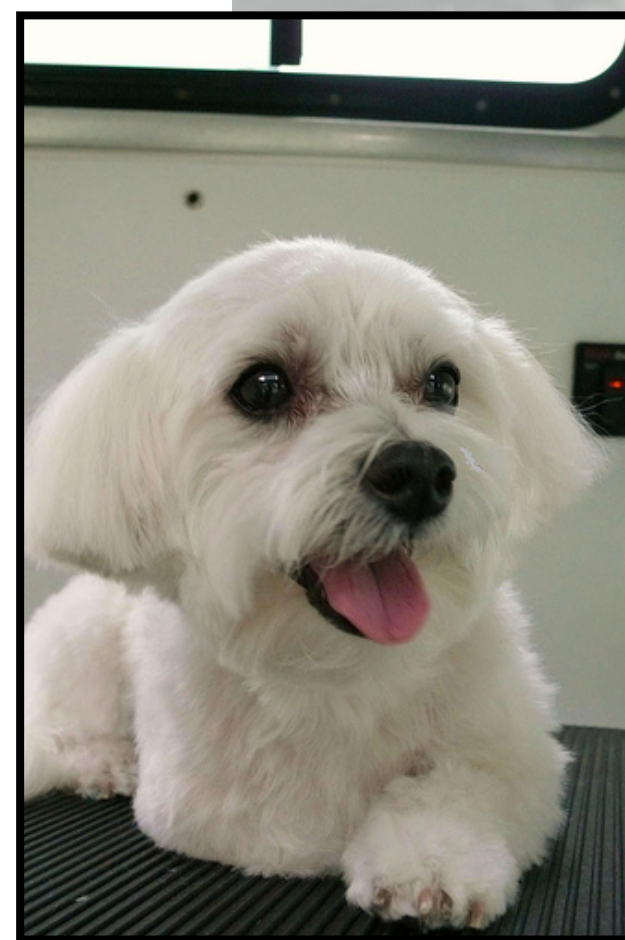
Of yes/no questions, 58.83% of the time, the answer is yes



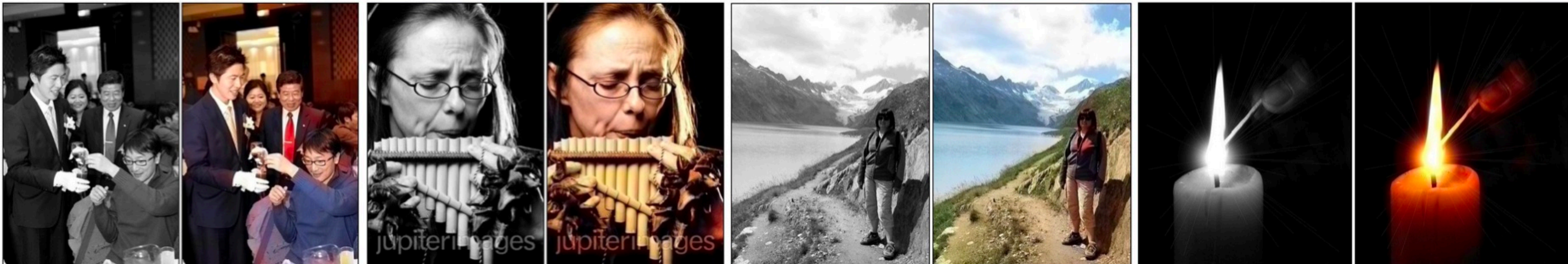
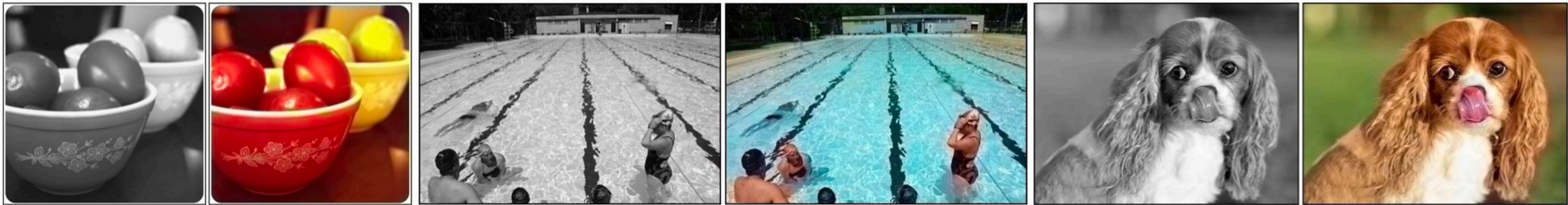
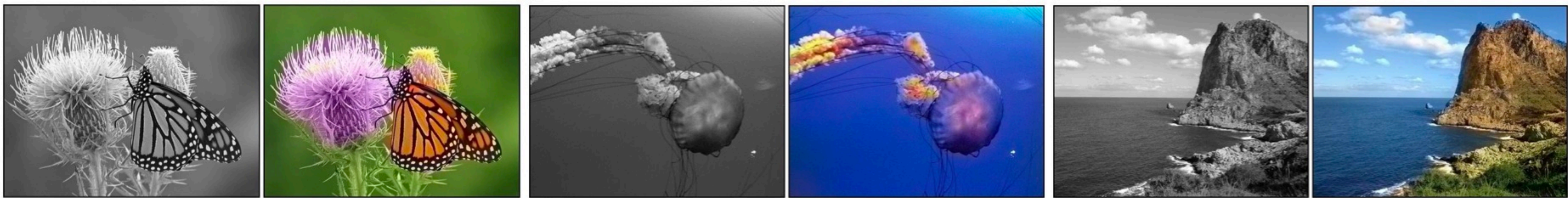
["Colorful image colorization", Zhang et al., ECCV 2016]



["Colorful image colorization", Zhang et al., ECCV 2016]



["Colorful image colorization", Zhang et al., ECCV 2016]





u/Rafael_P_S



Thylacine



Chopin

So we can sometimes collect good training data.

But suppose we messed up. Our test setting doesn't look like the training data!

How can we bridge the domain gap?

CARLA

Open-source simulator for autonomous driving research.

GET STARTED

Introduction

CARLA has been developed from the ground up to support development, training, and validation of autonomous driving systems. In addition to open-source code and protocols, CARLA provides open digital assets (urban layouts, buildings, vehicles) that were created for this purpose and can be used freely. The simulation platform supports flexible specification of sensor suites, environmental conditions, full control of all static and dynamic actors, maps generation and much more.

<http://carla.org/>

Training data

Driving simulator (GTA)

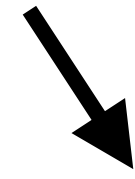


Test data

Driving in the real world



source domain



target domain

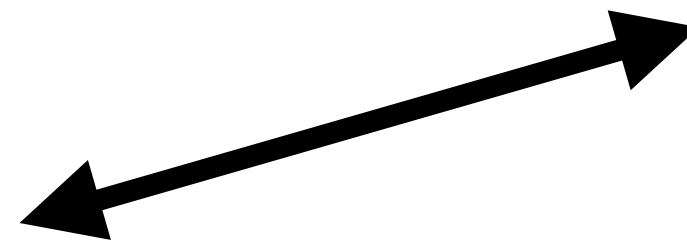
(where we actual use our model)



Domain gap between p_{source} and p_{target} will cause us to fail to generalize.

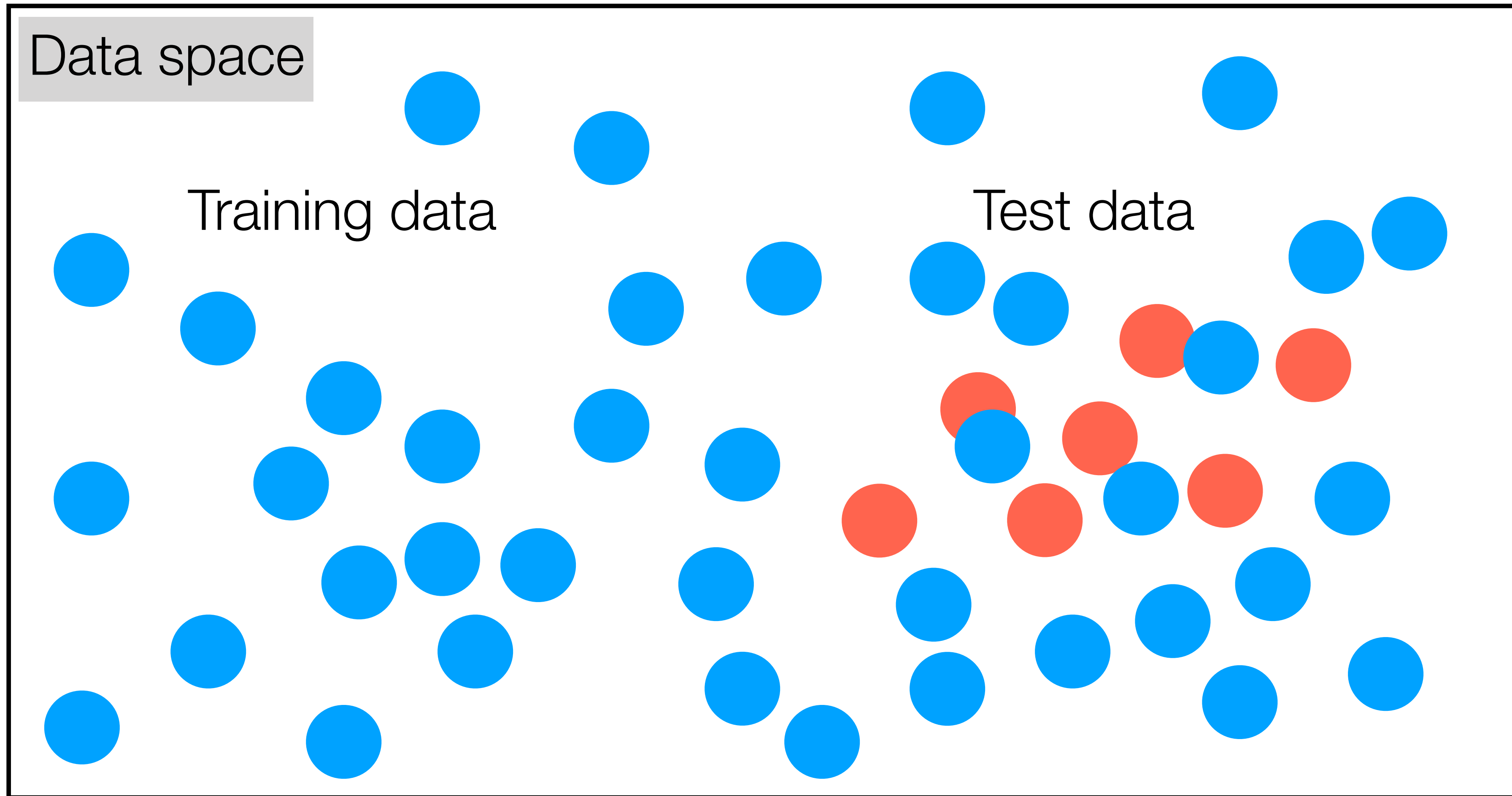
Space of images

Source data



Target data

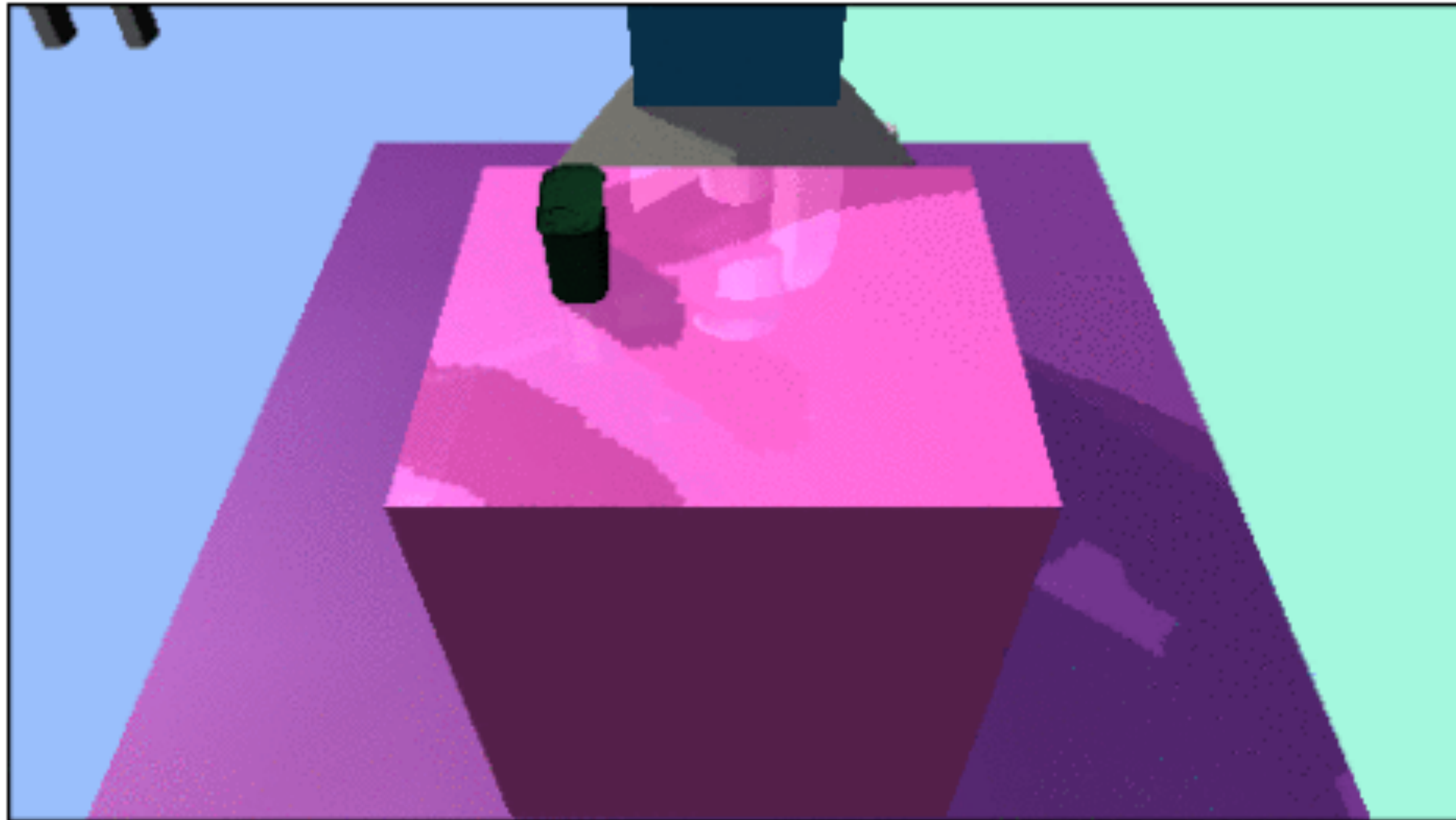
Idea #1: Train on randomly perturbed data, so that test set just looks like another random perturbation



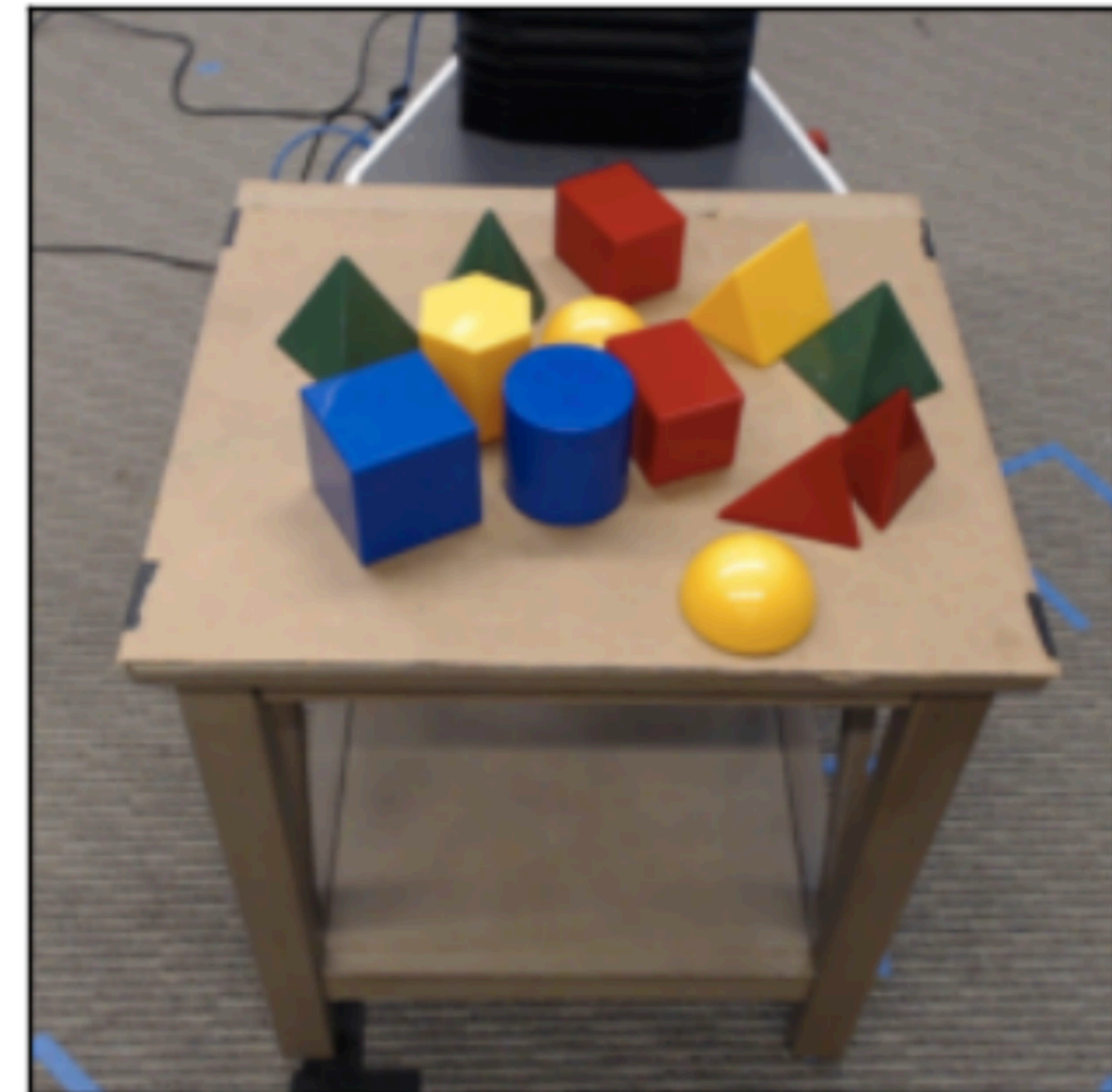
This is called **domain randomization** or **data augmentation**

Domain randomization

Training data



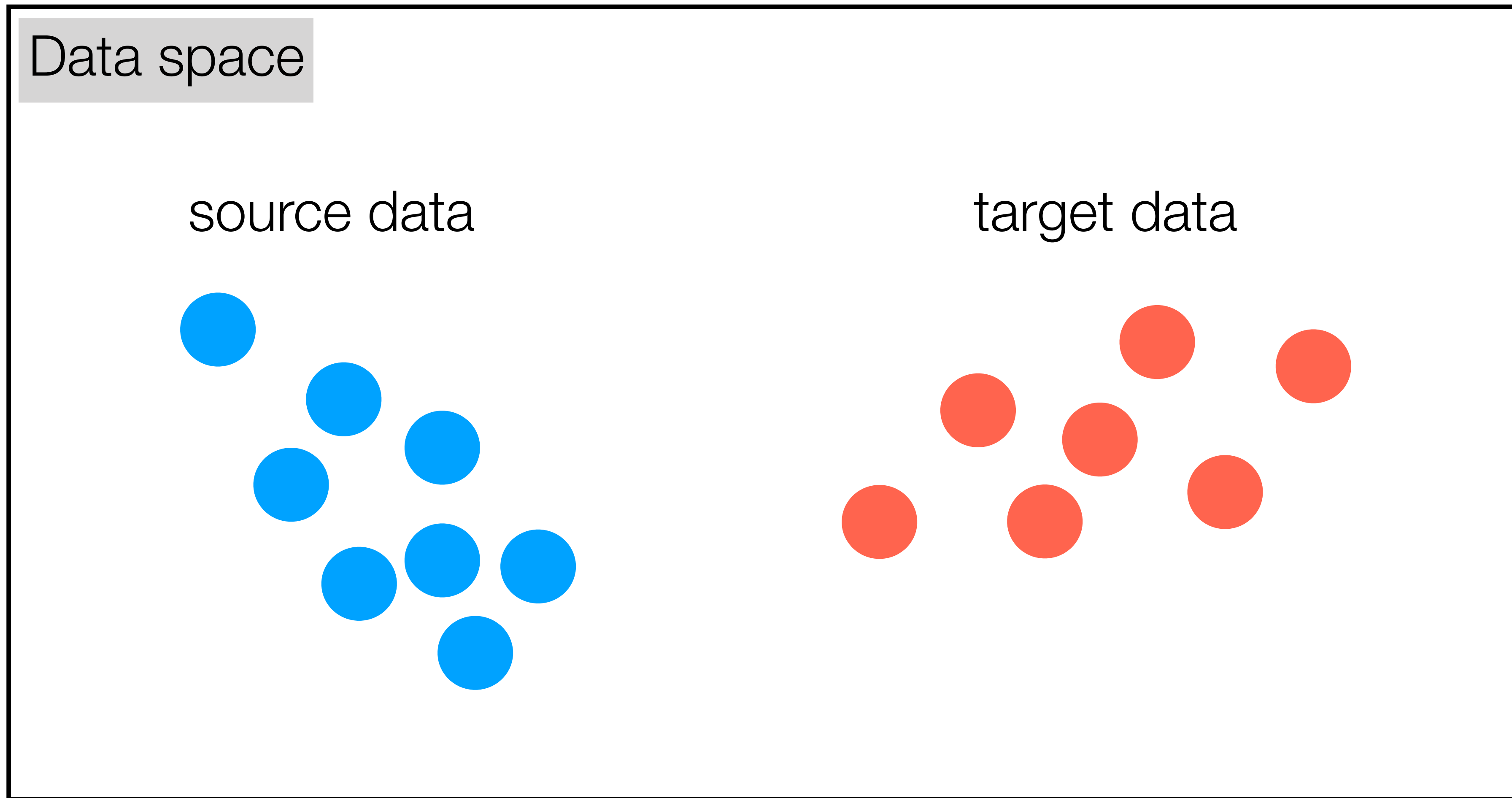
Test data



[Sadeghi & Levine 2016]

Above example is from [Tobin et al. 2017]

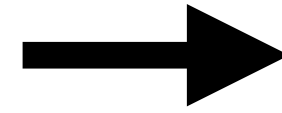
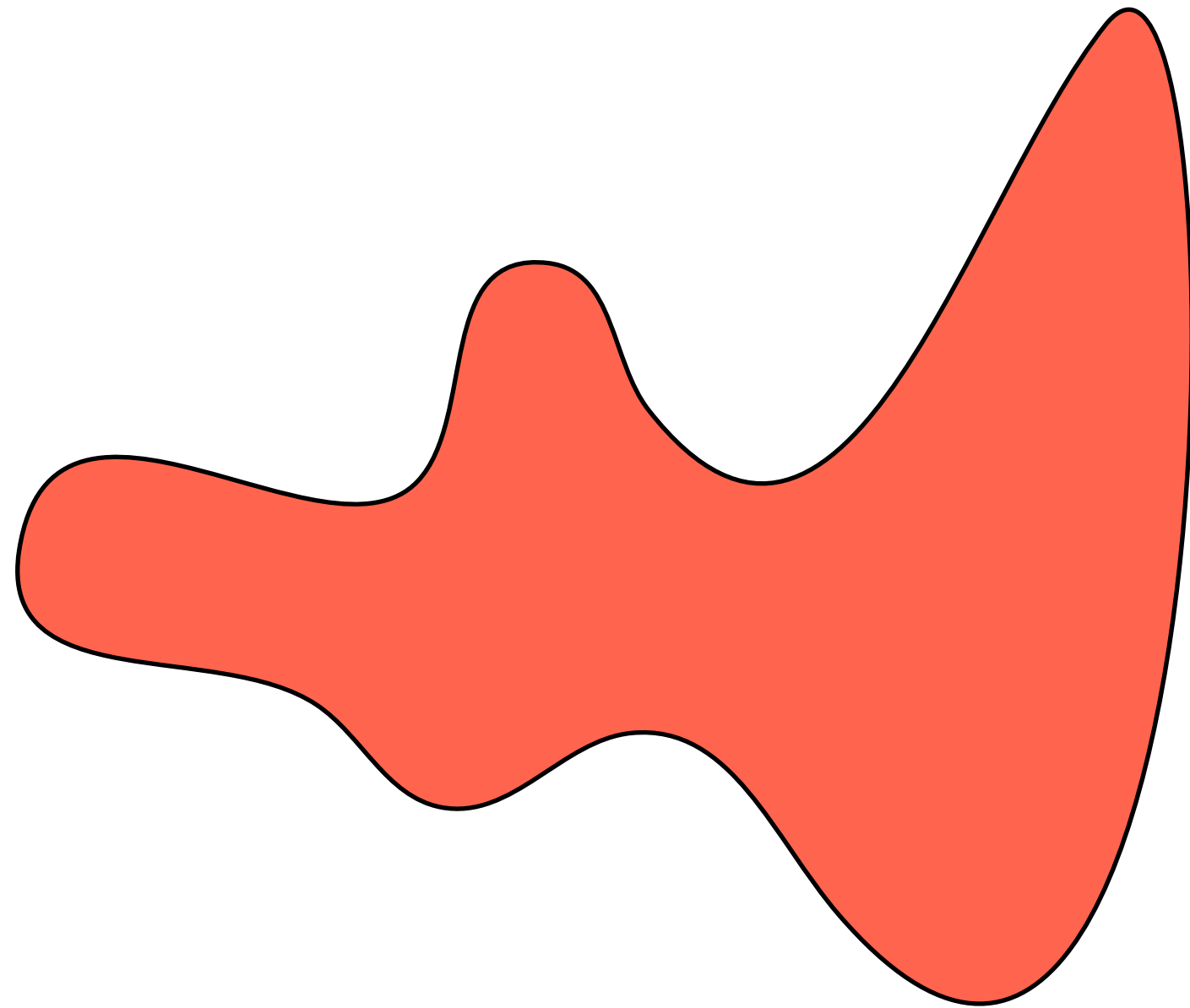
Idea #2: transform the target domain to look like the source domain



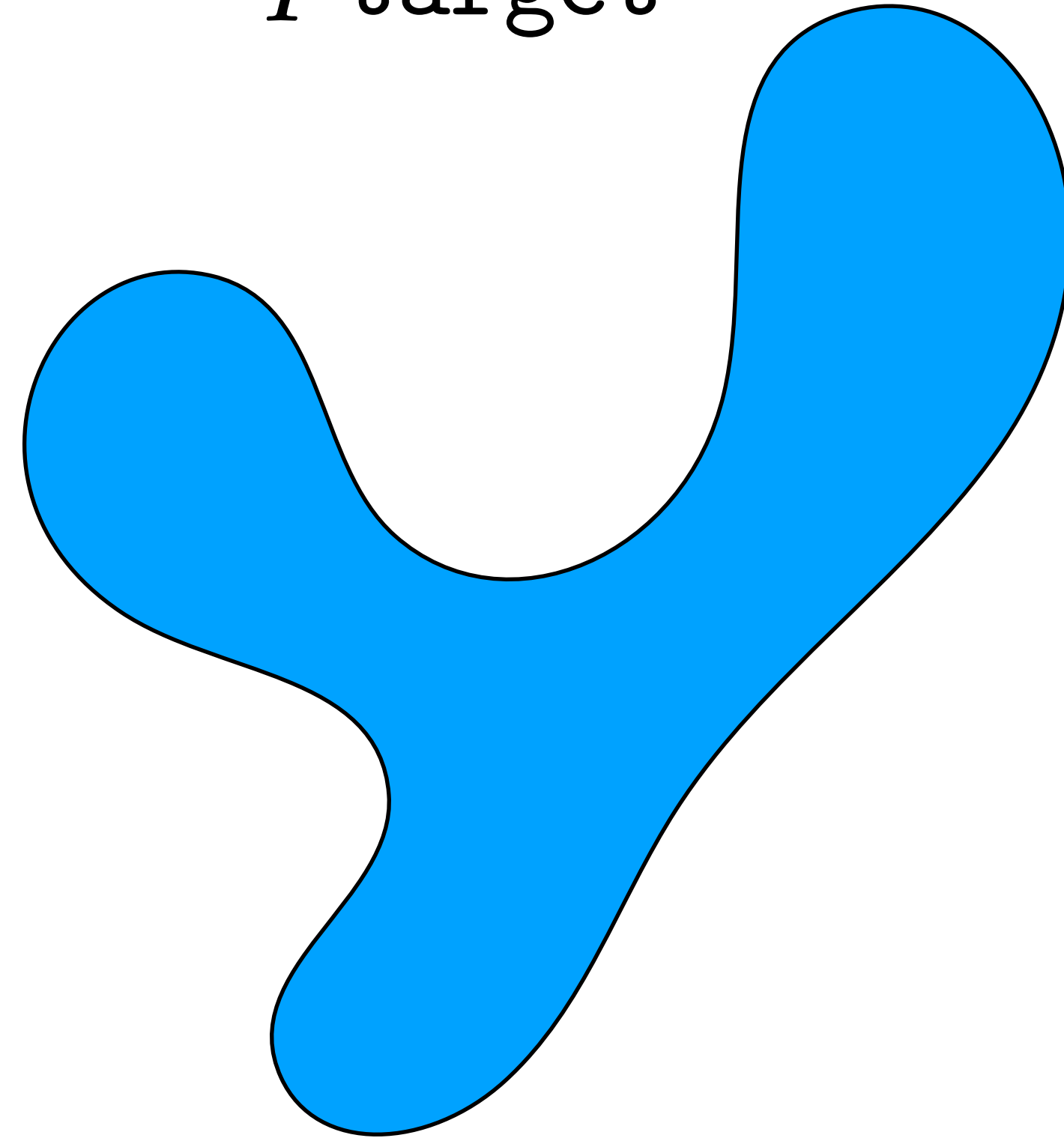
(Or vice versa)

This is called **domain adaptation**

p_{source}



p_{target}

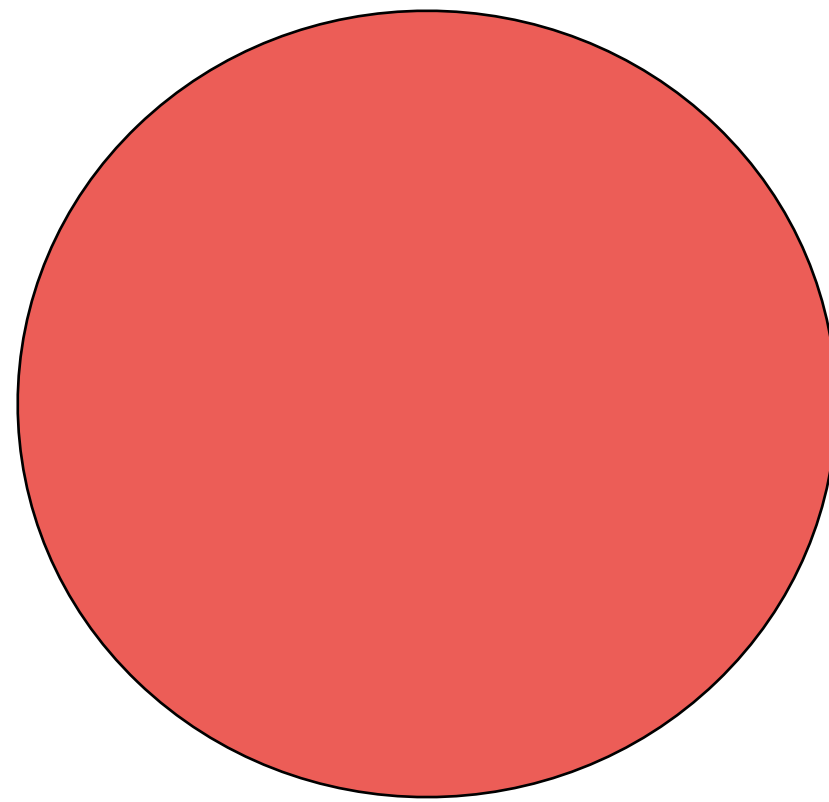


It's a just another distribution mapping problem!

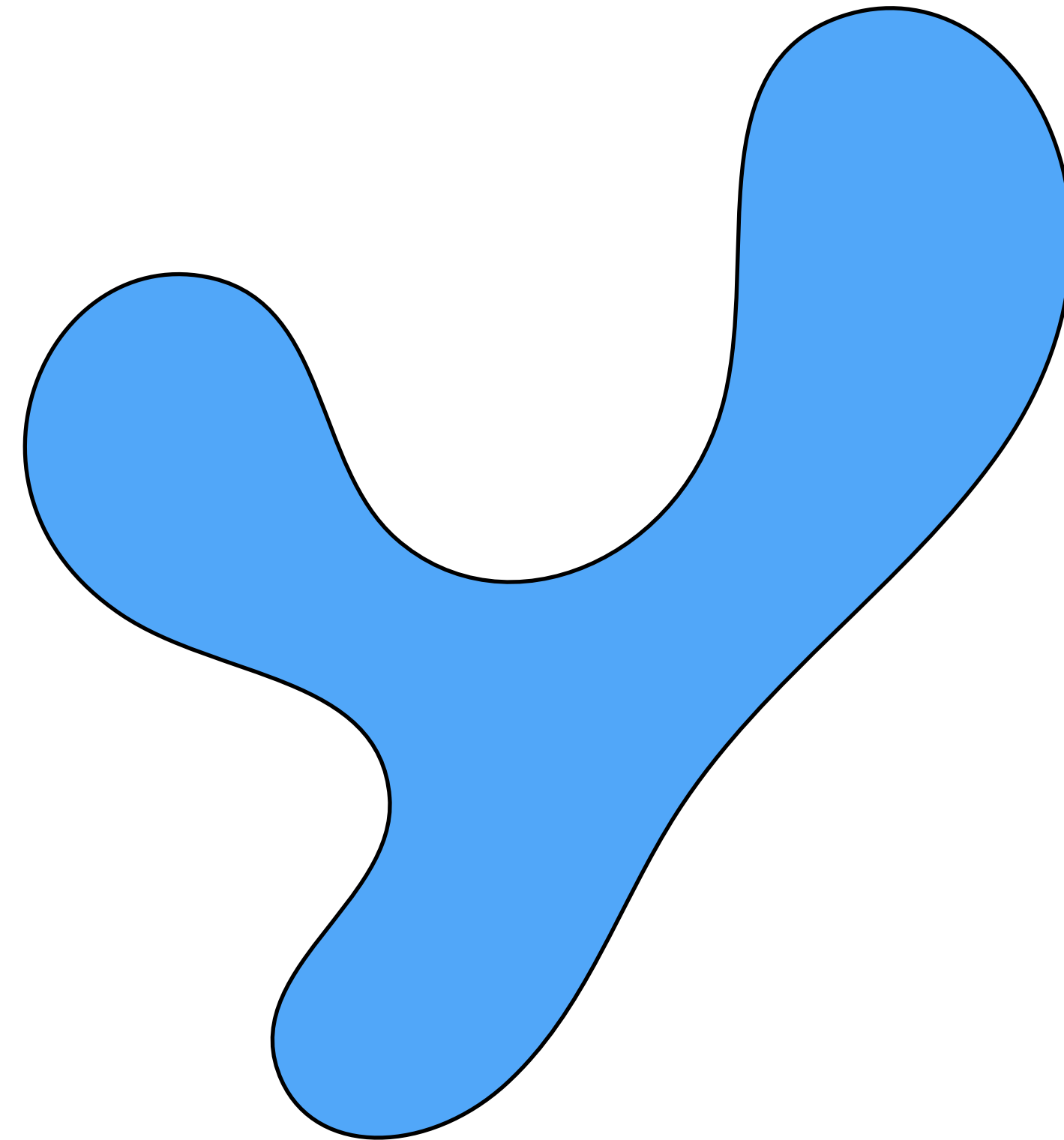
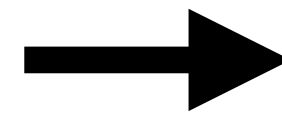
GANs

Gaussian

Target distribution



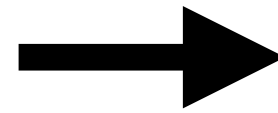
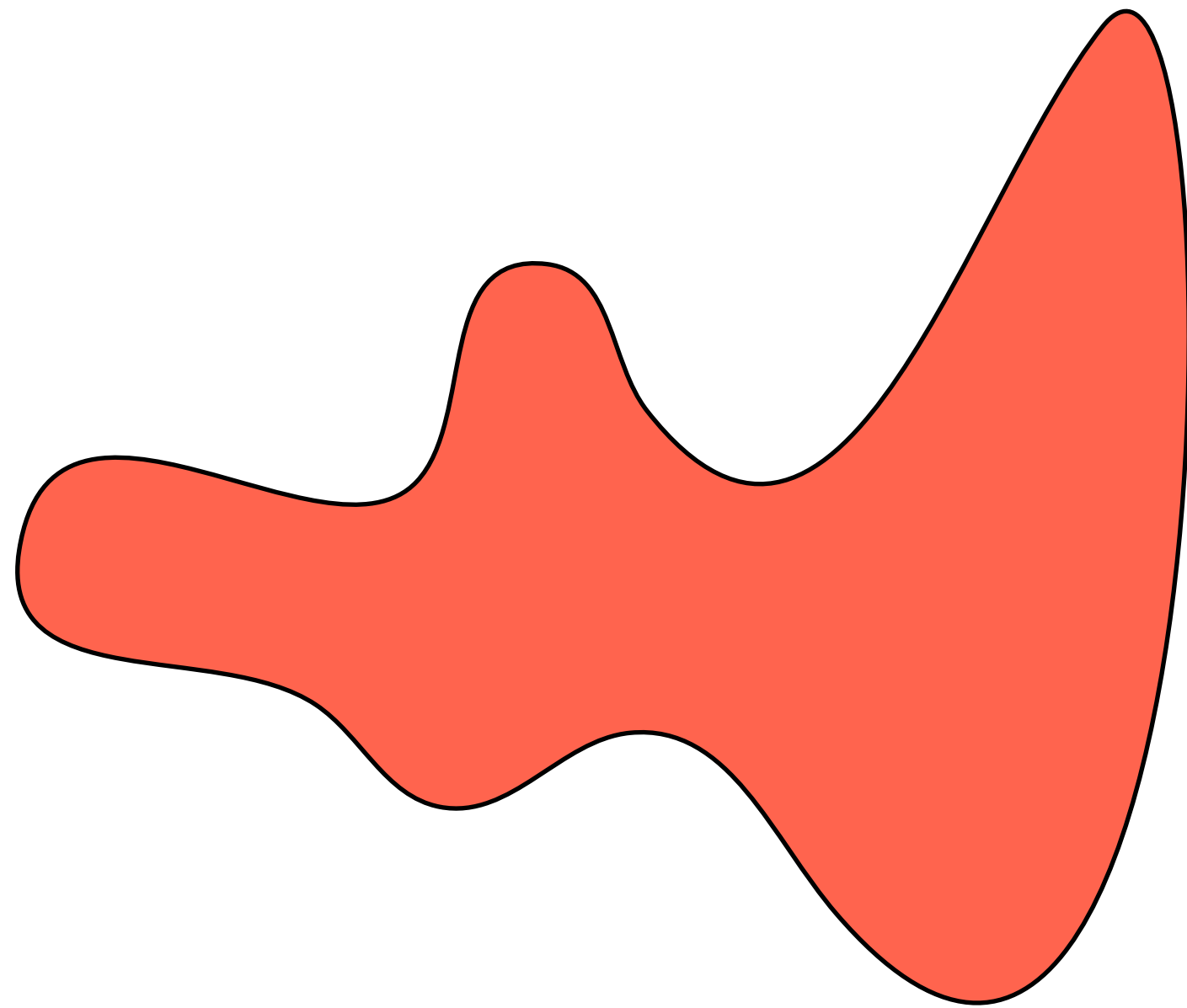
Z



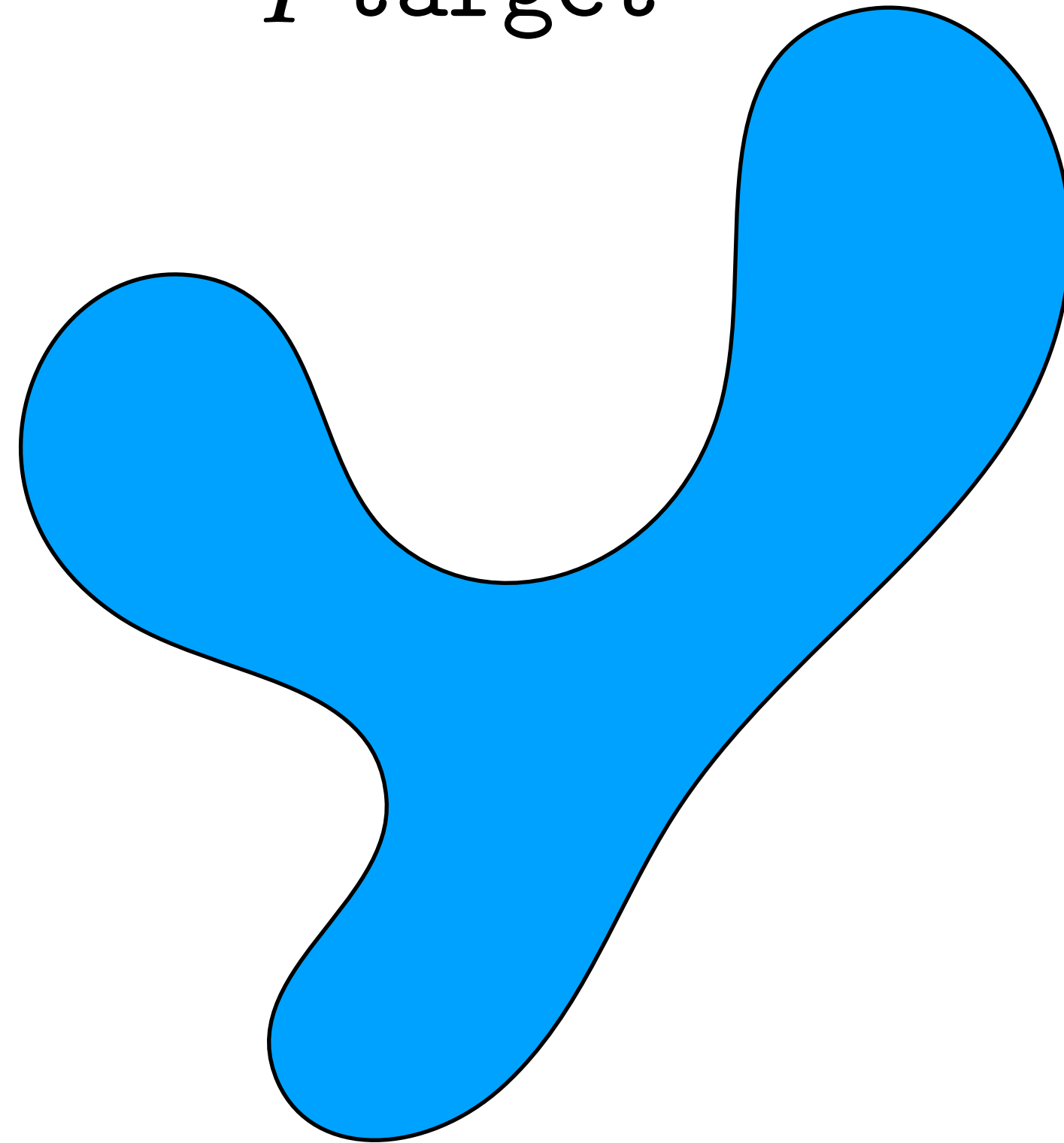
Y

Domain adaptation

p_{source}



p_{target}





GTA5 CG Input

Output

[Hoffman, Tzeng, Park, Zhu, Isola, Saenko, Darrell, Efros, ICML 2018]



Cityscape Input

Output

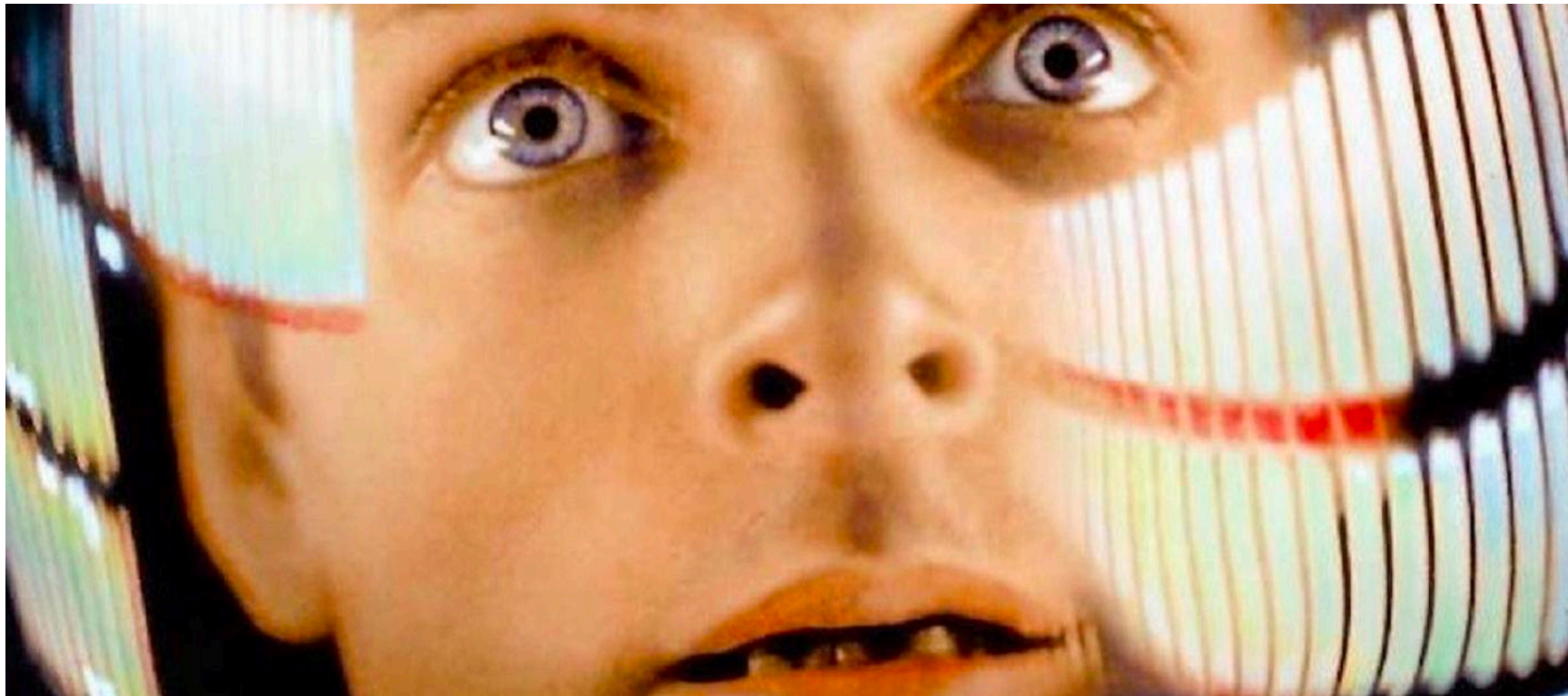
[Hoffman, Tzeng, Park, Zhu, Isola, Saenko, Darrell, Efros, ICML 2018]

Domain adaptation

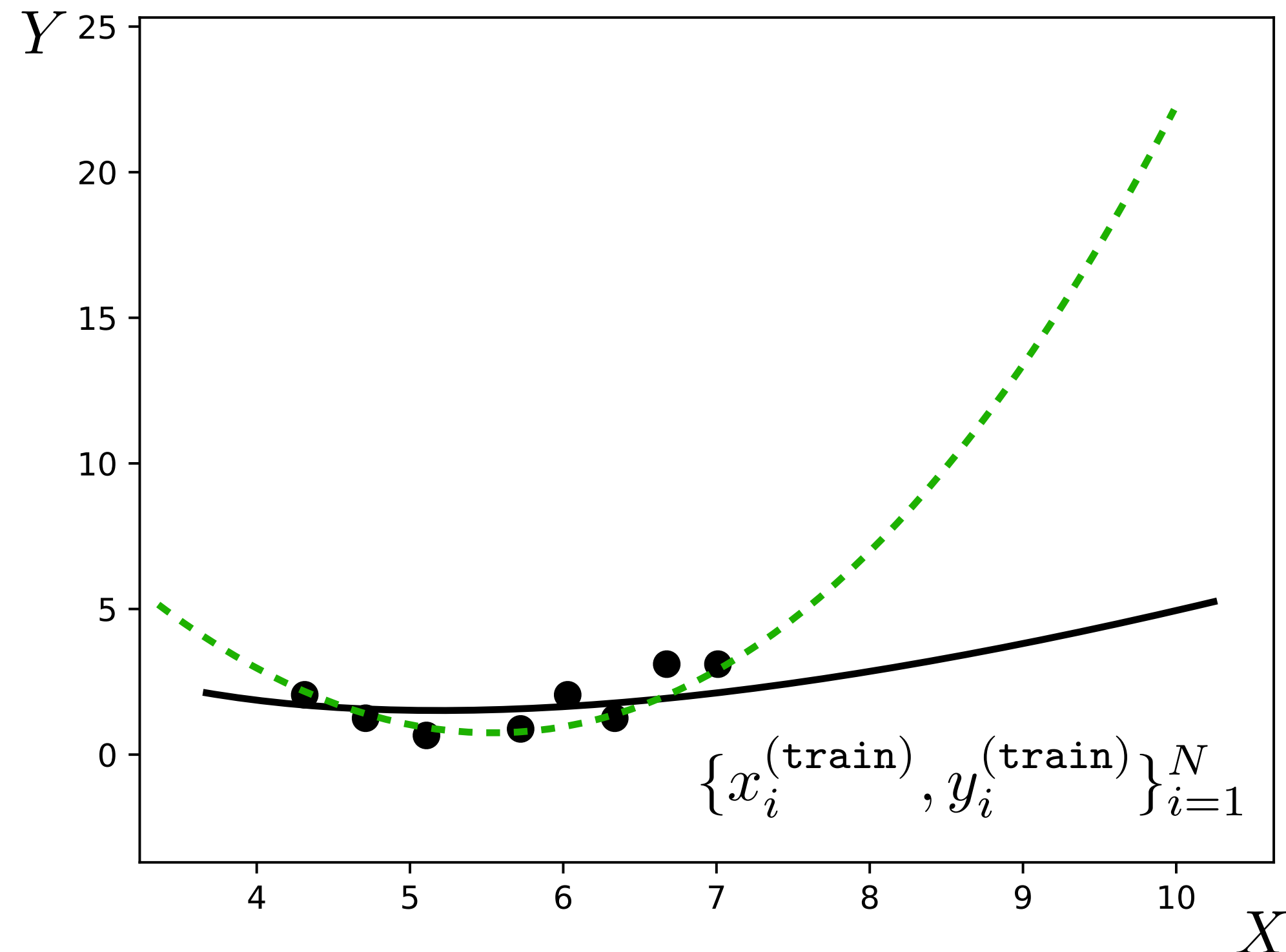
- We have source domain pairs $\{\mathbf{x}^{\text{source}}, \mathbf{y}^{\text{source}}\}$
- Learn a mapping $F: \mathbf{x}^{\text{source}} \rightarrow \mathbf{y}^{\text{source}}$
- We want to apply F to target domain data $\mathbf{x}^{\text{target}}$
- Find transformation $T: \mathbf{x}^{\text{target}} \rightarrow \mathbf{x}^{\text{source}}$
- Now apply $F(T(\mathbf{x}^{\text{target}}))$ to predict $\mathbf{y}^{\text{target}}$

Robustness

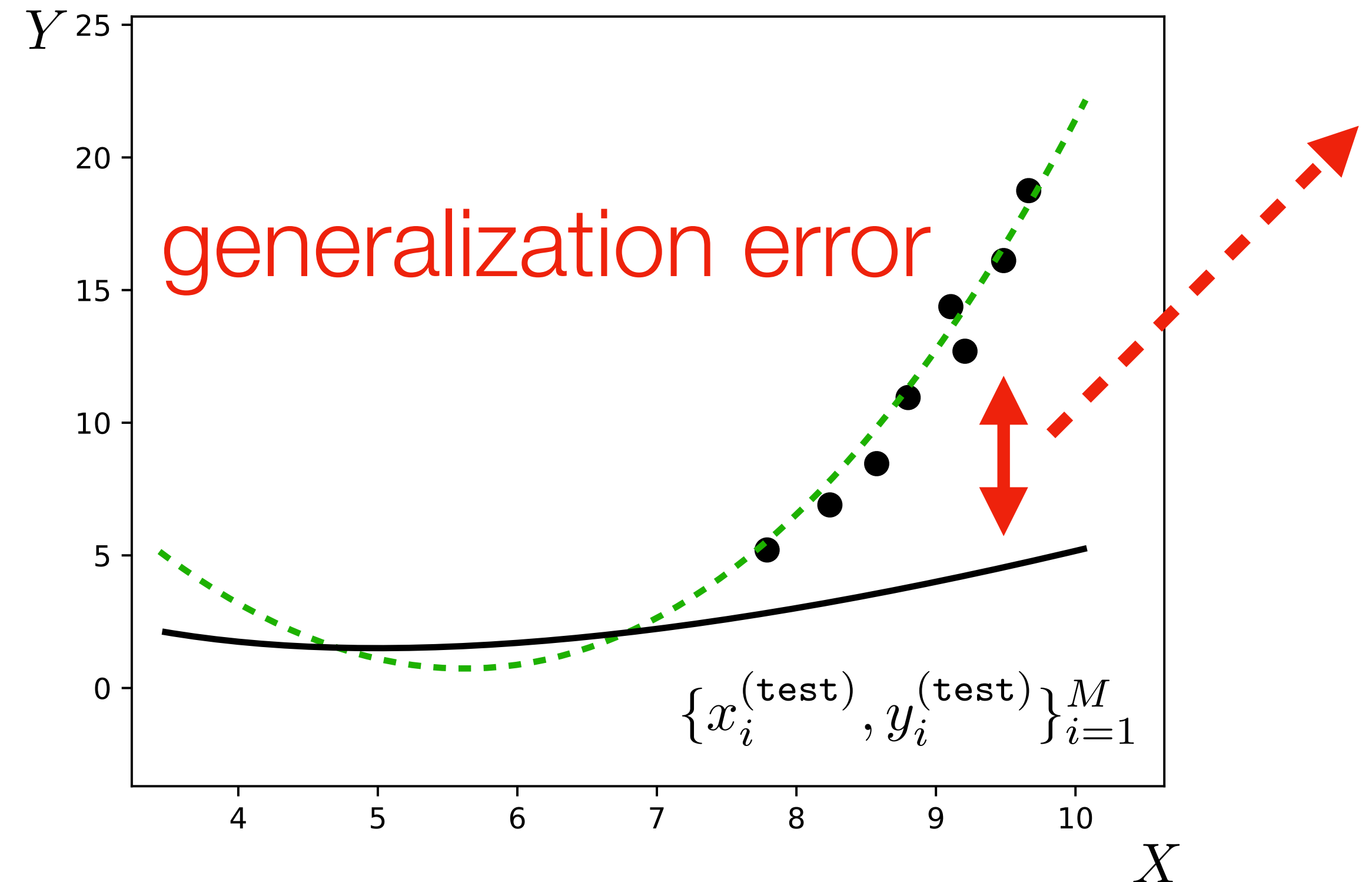
What if we go waaaaay outside of the training distribution?



Training data



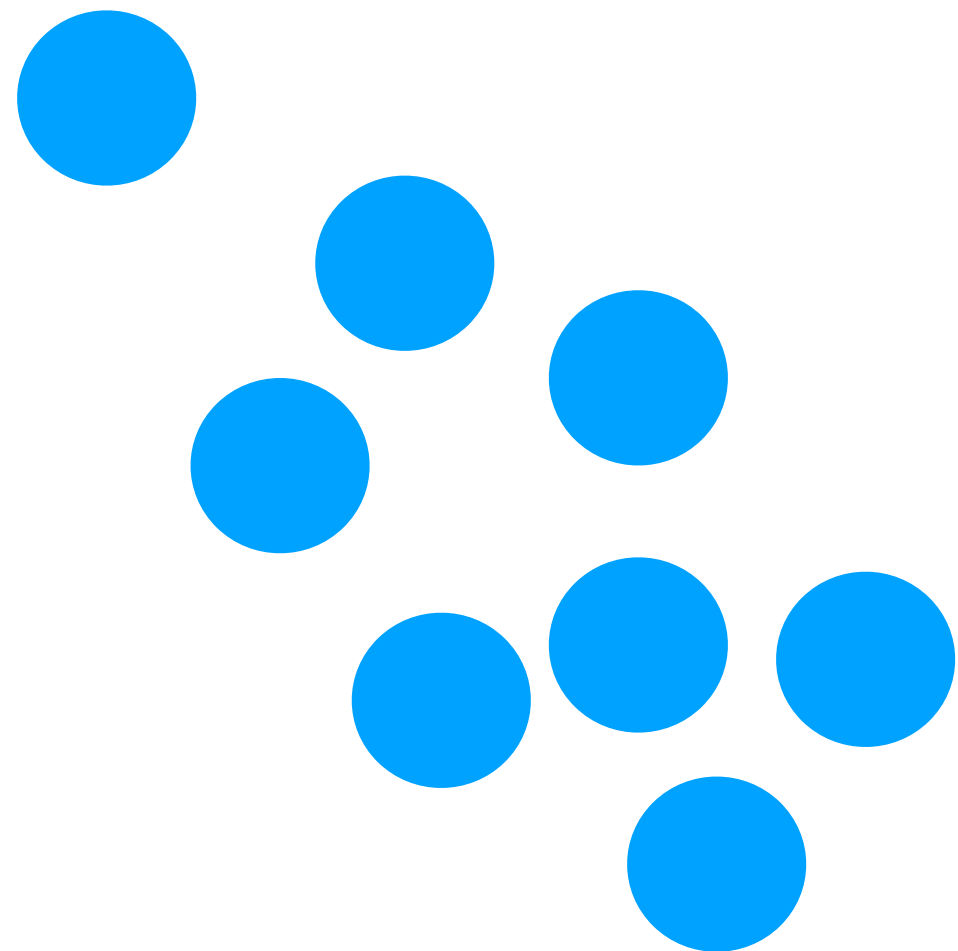
Test data



Our training data did not cover the part of the distribution that was tested
(biased data)

Data space

Training data

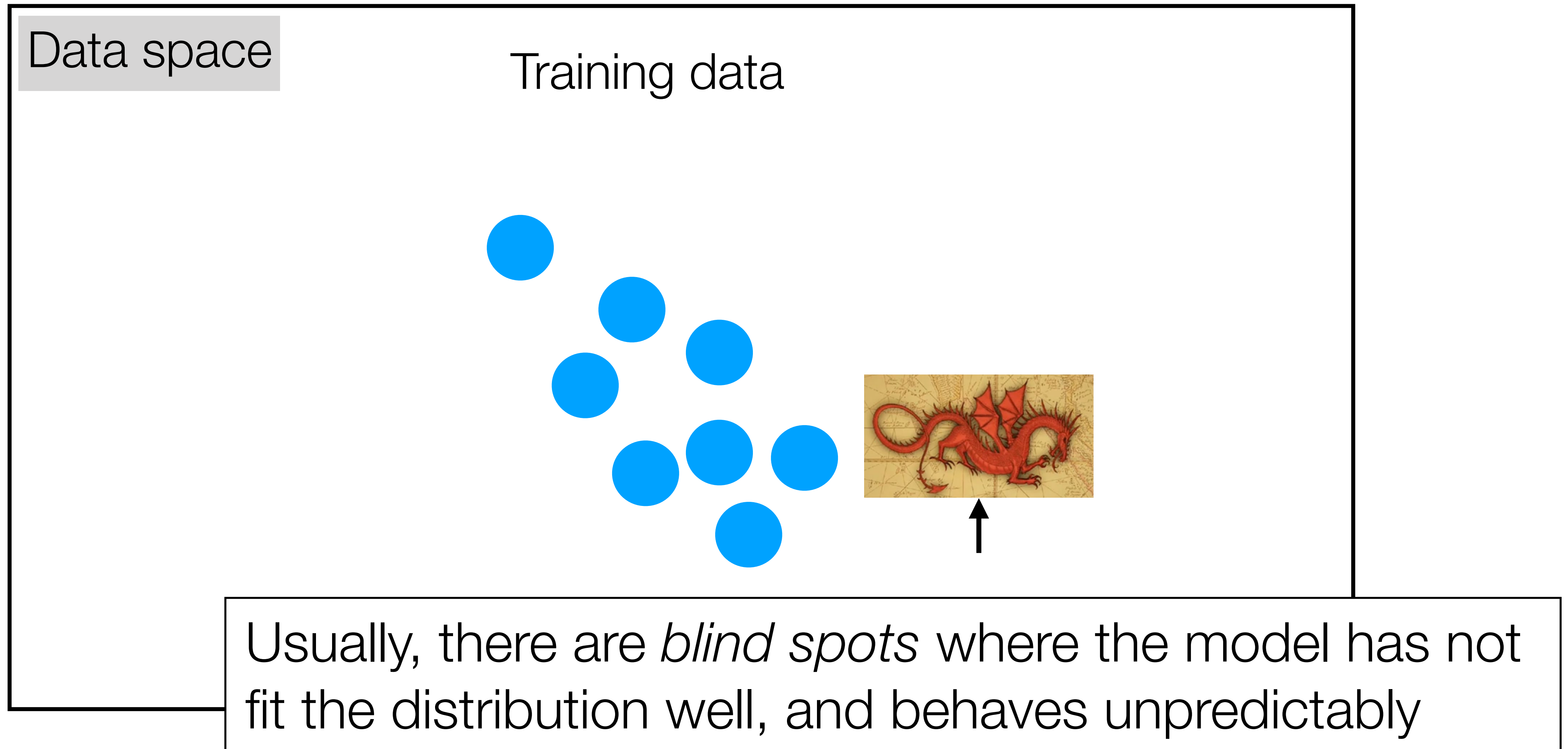


Test data

*Out here, model response
is highly unpredictable*



Weirdness of high-dimensional space:

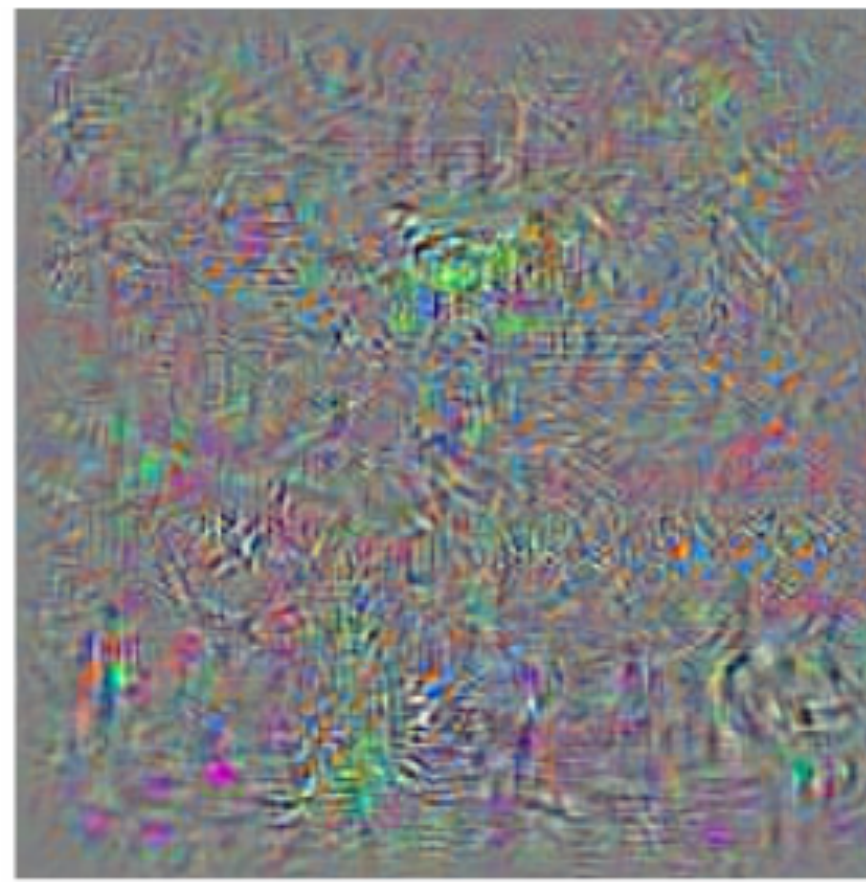


Adversarial noise

\mathbf{x}



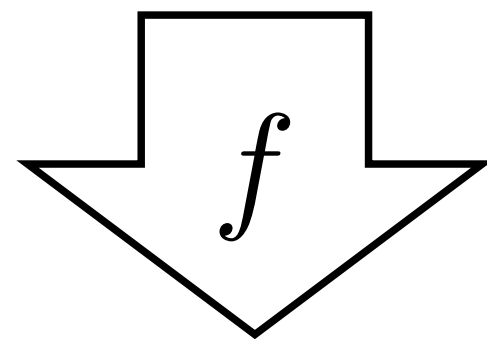
\mathbf{r}



+

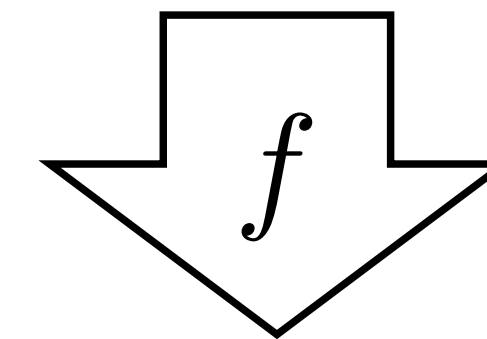
=

$\mathbf{x} + \mathbf{r}$



y

“School bus”



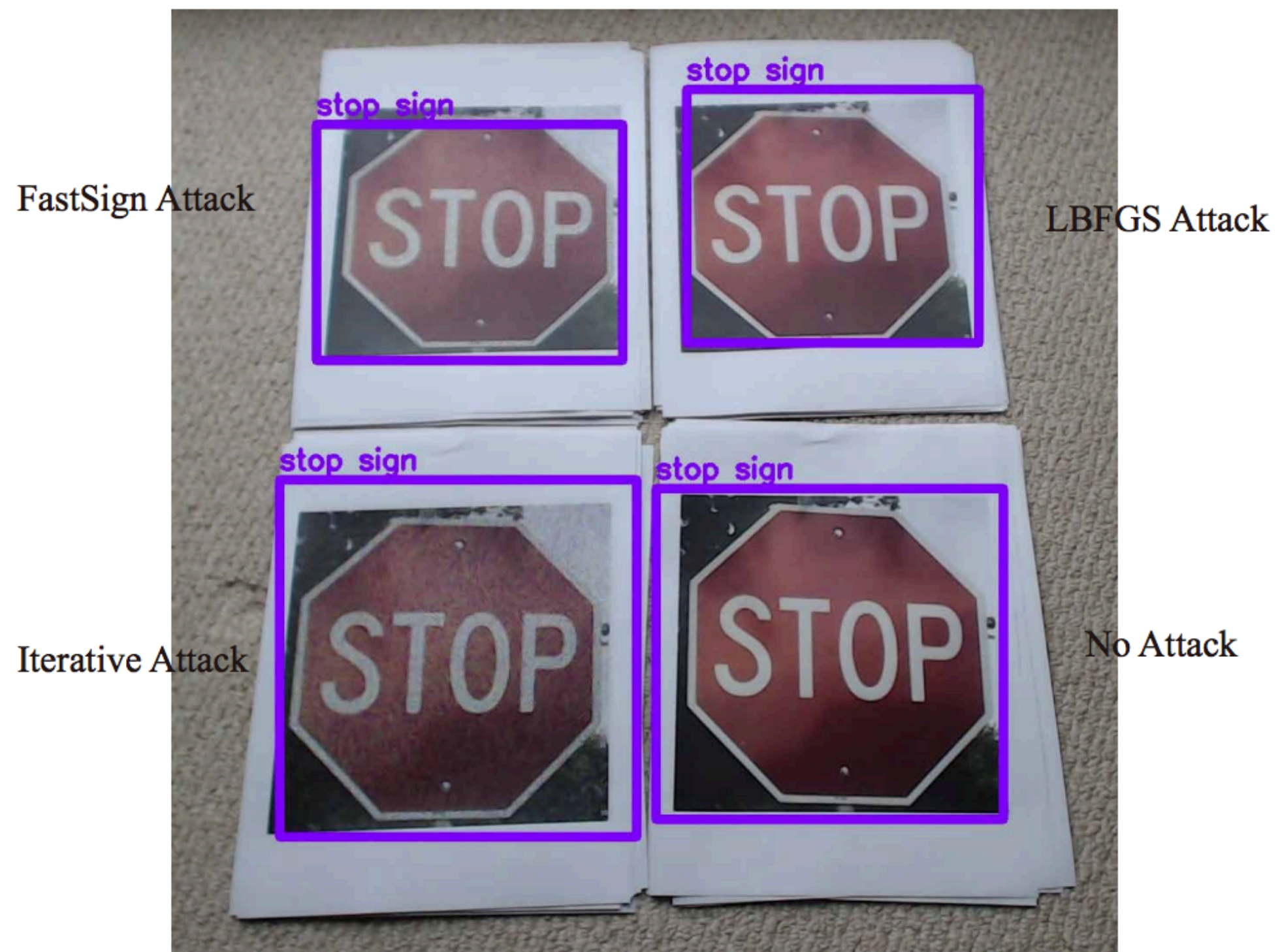
“Ostrich”

$$\arg \max_{\mathbf{r}} p(y = \text{ostrich} | \mathbf{x} + \mathbf{r}) \quad \text{subject to} \quad \|\mathbf{r}\| < \epsilon$$

[“Intriguing properties of neural networks”, Szegedy et al. 2014]

Anything to worry about?

“NO Need to Worry about Adversarial Examples in Object Detection in Autonomous Vehicles”, Lu et al. 2017



(Early) 2017's attacks fail on physical objects, since they are optimized to attack a single view!

Anything to worry about?

Later in 2017...

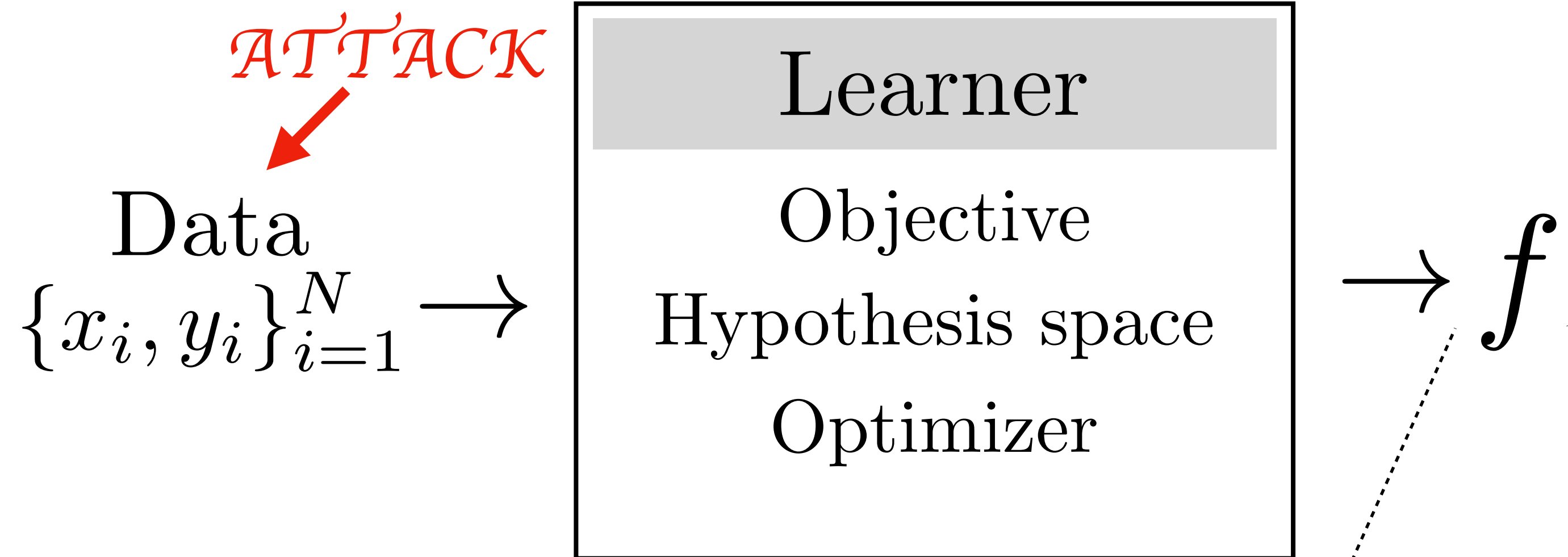
“Synthesizing Robust Adversarial Examples”, Athalye, Engstrom, Ilyas, Kwok, 2017

3D-printed **turtle** model classified as **rifle** from most viewpoints

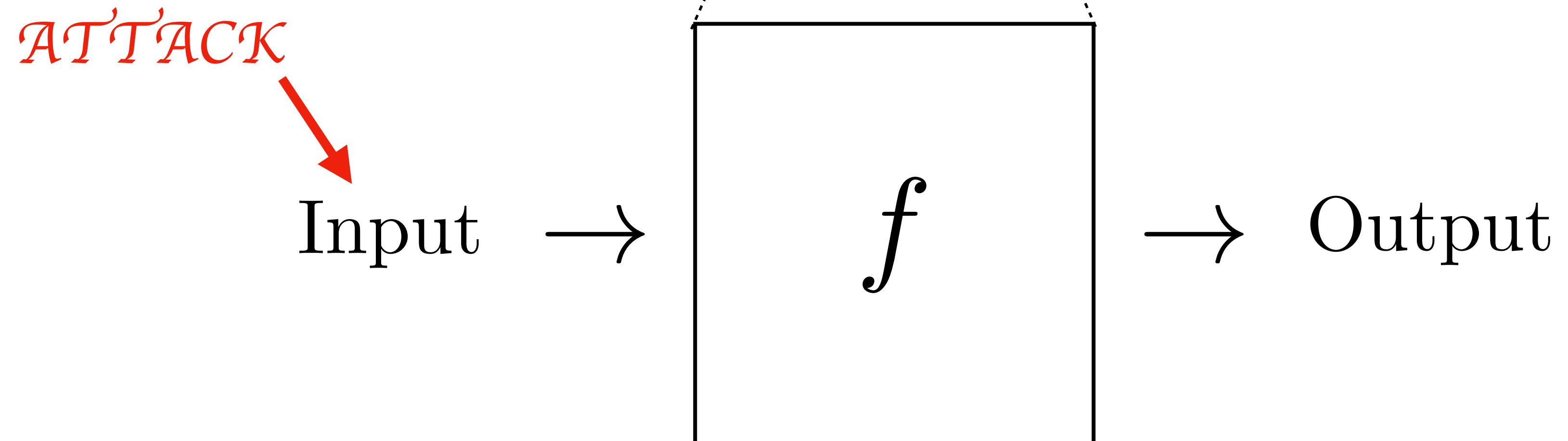


Attacks

Learning



Inference

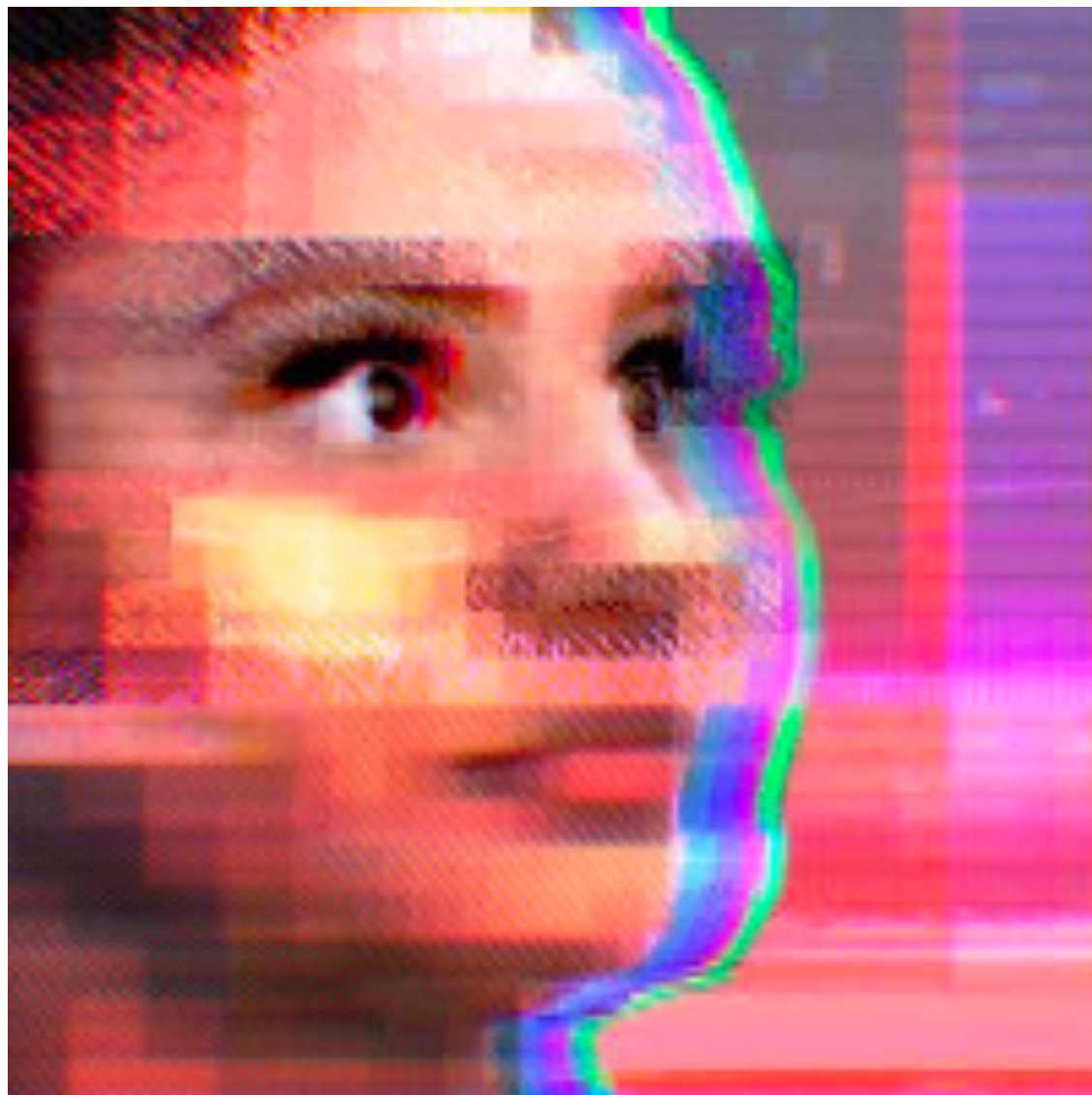


Garbage in, garbage out

A machine learning algorithm will do whatever the training data tells it to do.

If the data is bad or biased, the learned algorithm will be too.

Microsoft's Tay chatbot






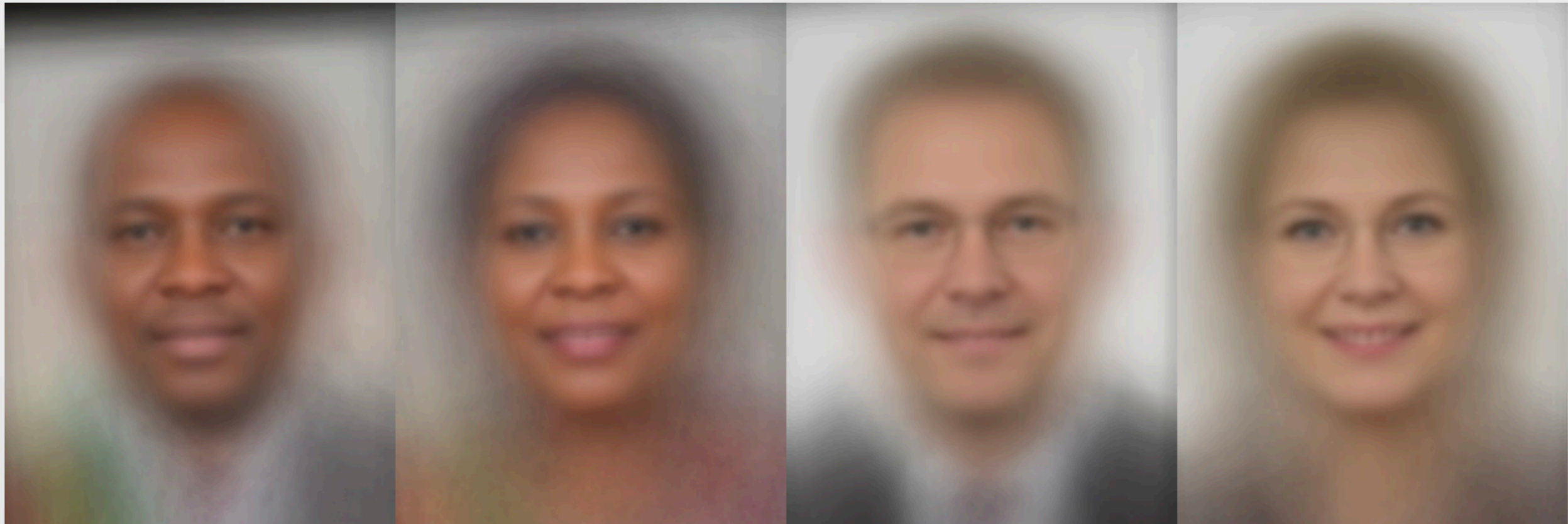
Chatbot released on twitter.

Learned from interactions with users (?)

Started mimicking offensive language, was shut down.

Algorithmic Bias

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
 Microsoft	94.0% <div><div></div></div>	79.2% <div><div></div></div>	100% <div><div></div></div>	98.3% <div><div></div></div>	20.8% <div><div></div></div>
 FACE++	99.3% <div><div></div></div>	65.5% <div><div></div></div>	99.2% <div><div></div></div>	94.0% <div><div></div></div>	33.8% <div><div></div></div>
 IBM	88.0% <div><div></div></div>	65.3% <div><div></div></div>	99.7% <div><div></div></div>	92.9% <div><div></div></div>	34.4% <div><div></div></div>



<http://gendershades.org/overview.html>

Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*

Joy Buolamwini
MIT Media Lab 75 Amherst St. Cambridge, MA 02139

JOYAB@MIT.EDU

Timnit Gebru
Microsoft Research 641 Avenue of the Americas, New York, NY 10011

TIMNIT.GEBRU@MICROSOFT.COM

Editors: Sorelle A. Friedler and Christo Wilson

Abstract

Recent studies demonstrate that machine learning algorithms can discriminate based on classes like race and gender. In this work, we present an approach to evaluate bias present in automated facial analysis algorithms and datasets with respect to phenotypic subgroups. Using the dermatologist approved Fitzpatrick Skin Type classification system, we characterize the gender and skin type distribution of two facial analysis benchmarks, IJB-A and Adience. We find that these datasets are overwhelmingly composed of lighter-skinned subjects (79.6% for IJB-A and 86.2% for Adience) and introduce a new facial analysis dataset which is balanced by gender and skin type. We evaluate 3 commercial gender classification systems using our dataset and show that darker-skinned females are the most misclassified group (with error rates of up to 34.7%). The maximum error rate for lighter-skinned males is 0.8%. The substantial disparities in the accuracy of classifying darker females, lighter females, darker males, and lighter males in gender classification systems require urgent attention if commercial companies are to build genuinely fair, transparent and accountable facial analysis algorithms.

Keywords: Computer Vision, Algorithmic Audit, Gender Classification

1. Introduction

Artificial Intelligence (AI) is rapidly infiltrating every aspect of society. From helping determine

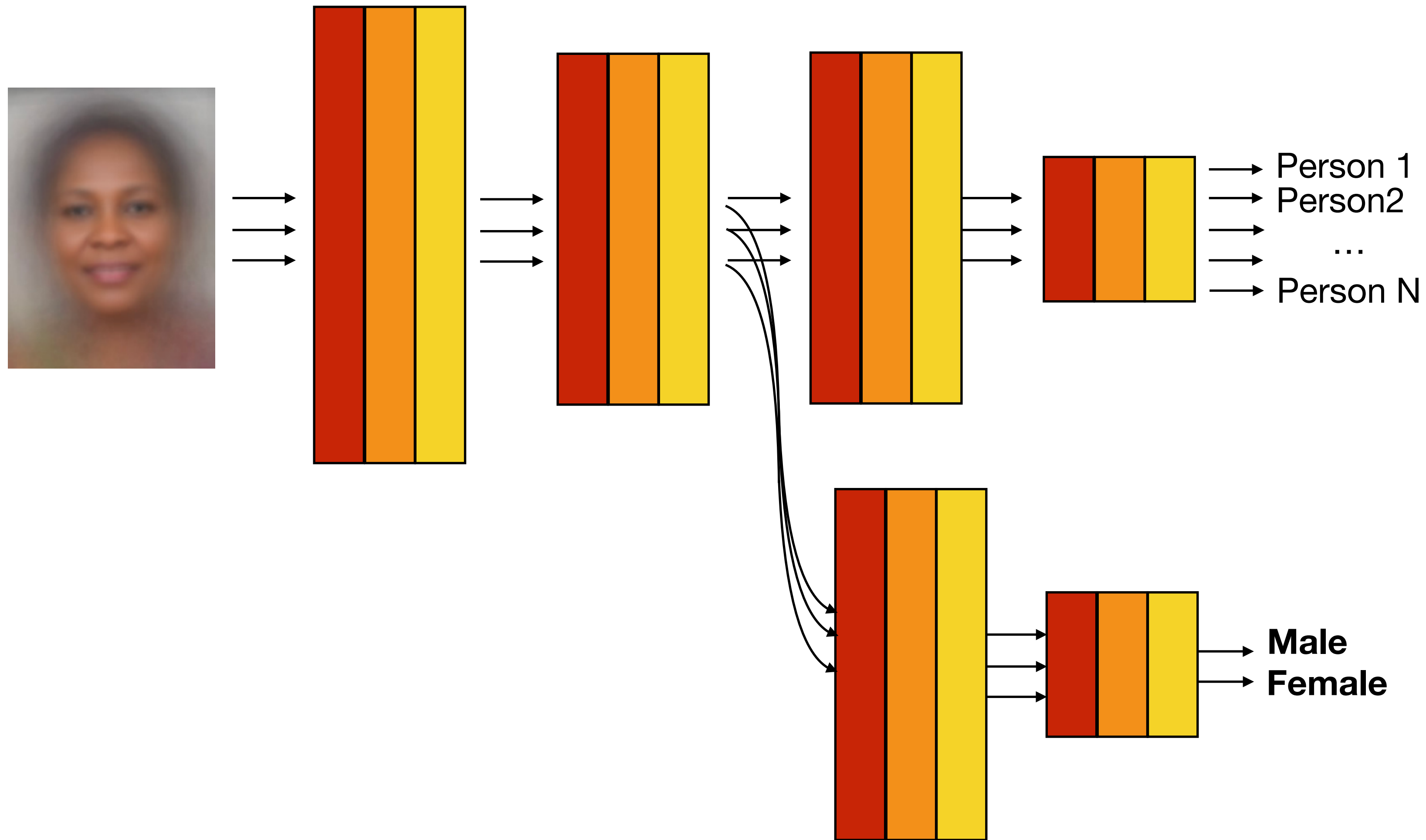
* Download our gender and skin type balanced PPB dataset at gendershades.org

who is hired, fired, granted a loan, or how long an individual spends in prison, decisions that have traditionally been performed by humans are rapidly made by algorithms (O’Neil, 2017; Citron and Pasquale, 2014). Even AI-based technologies that are not specifically trained to perform high-stakes tasks (such as determining how long someone spends in prison) can be used in a pipeline that performs such tasks. For example, while face recognition software by itself should not be trained to determine the fate of an individual in the criminal justice system, it is very likely that such software is used to identify suspects. Thus, an error in the output of a face recognition algorithm used as input for other tasks can have serious consequences. For example, someone could be wrongfully accused of a crime based on erroneous but confident misidentification of the perpetrator from security video footage analysis.

Many AI systems, e.g. face recognition tools, rely on machine learning algorithms that are trained with labeled data. It has recently been shown that algorithms trained with biased data have resulted in algorithmic discrimination (Bolukbasi et al., 2016; Caliskan et al., 2017). Bolukbasi et al. even showed that the popular word embedding space, Word2Vec, encodes societal gender biases. The authors used Word2Vec to train an analogy generator that fills in missing words in analogies. The analogy man is to computer programmer as woman is to “X” was completed with “homemaker”, conforming to the stereotype that programming is associated with men and homemaking with women. The biases in Word2Vec are thus likely to be propagated throughout any system that uses this embedding.

<http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>

ML Fairness, Domain adaptation



Overview of domain adaptation techniques: <https://arxiv.org/pdf/1802.03601.pdf>

Fairness and machine learning

Table of contents

[About this book](#)

1. [Introduction](#)

2. [Demographic classification criteria](#)

We introduce formal non-discrimination criteria, establish their relationships, and illustrate their limitations.

3. Legal background and normative questions

We survey the literature on discrimination in law, sociology, and philosophy. We then discuss the challenges that arise in translating these ideas of fairness to the statistical decision-making setting.

4. Causal inference

We dive into the rich technical repertoire of causal inference and how it helps articulate and address shortcomings of the prediction paradigm.

5. Measurement

different categories defined along the lines of different (conditional) independence⁷⁰ statements between the involved random variables.

Independence	Separation	Sufficiency
$R \perp A$	$R \perp A \mid Y$	$Y \perp A \mid R$

Below we will introduce and discuss each of these conditions in detail. Variants of these criteria arise from different ways of relaxing them.

As an exercise, think about why we omitted the conditional independence statement $R \perp Y \mid A$ from our discussion here.

Independence

Our first formal criterion simply requires the sensitive characteristic to be statistically independent of the score.

Definition 1. *The random variables (A, R) satisfy independence if $A \perp R$.*

Independence has been explored through many equivalent terms or variants, referred to as *demographic parity*, *statistical parity*, *group fairness*, *disparate impact* and others. In the case of binary classification, independence simplifies to the condition

$$\mathbb{P}\{R = 1 \mid A = a\} = \mathbb{P}\{R = 1 \mid A = b\},$$

for all groups a, b . Thinking of the event $R = 1$ as “acceptance”, the condition requires the acceptance rate to be the same in all groups. A relaxation of the constraint introduces a positive amount of slack $\epsilon > 0$ and requires that

$$\mathbb{P}\{R = 1 \mid A = a\} \geq \mathbb{P}\{R = 1 \mid A = b\} - \epsilon.$$

Note that we can swap a and b to get an inequality in the other direction. An alternative relaxation is to consider a ratio condition, such as,

$$\frac{\mathbb{P}\{R = 1 \mid A = a\}}{\mathbb{P}\{R = 1 \mid A = b\}} \geq 1 - \epsilon.$$

Some have argued⁷¹ that, for $\epsilon = 0.2$, this condition relates to the *80 percent rule* in disparate impact law.

Yet another way to state the independence condition in full generality is to require that A and R must have zero mutual information⁷² $I(A; R) = 0$. The characterization in terms of mutual information leads to useful relaxations of the constraint. For example, we could require $I(A; R) \leq \epsilon$.

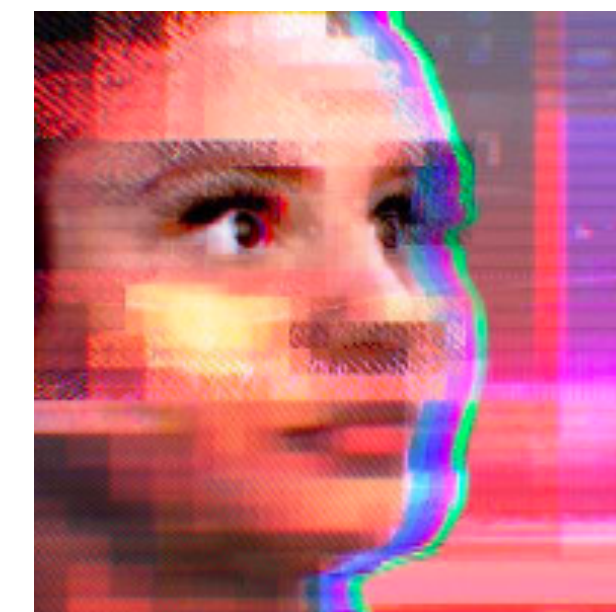
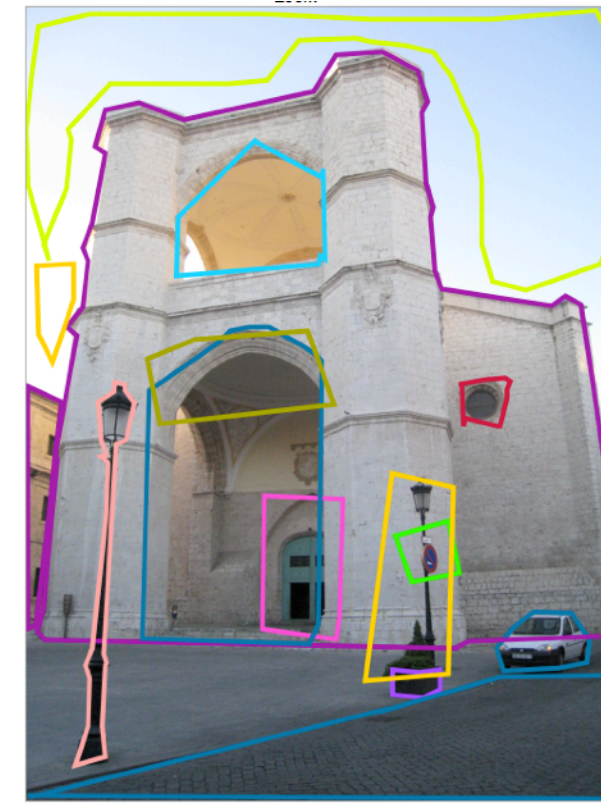
⁷⁰ Learn more about conditional independence here.

⁷¹ Michael Feldman et al., “Certifying and Removing Disparate Impact,” in *Proc. 21th ACM SIGKDD*, 2015.

⁷² Mutual information is defined as $I(A; R) = H(A) + H(R) - H(A, R)$, where H denotes the entropy.

Anything else to worry about?

- Our datasets are often poorly labeled
- And usually biased (overrepresent certain categories)
- ML methods perform beautifully on laboratory data, but often generalize poorly to real-world data
- Can have negative social consequences



But we can make progress, by

- Carefully curating and balancing our training data
- Using domain randomization and adaptation to bridge the gap between train and test
- Using simulators and generative models to create every richer and more compelling data to learn from
- Testing our methods to expose their limitations and biases

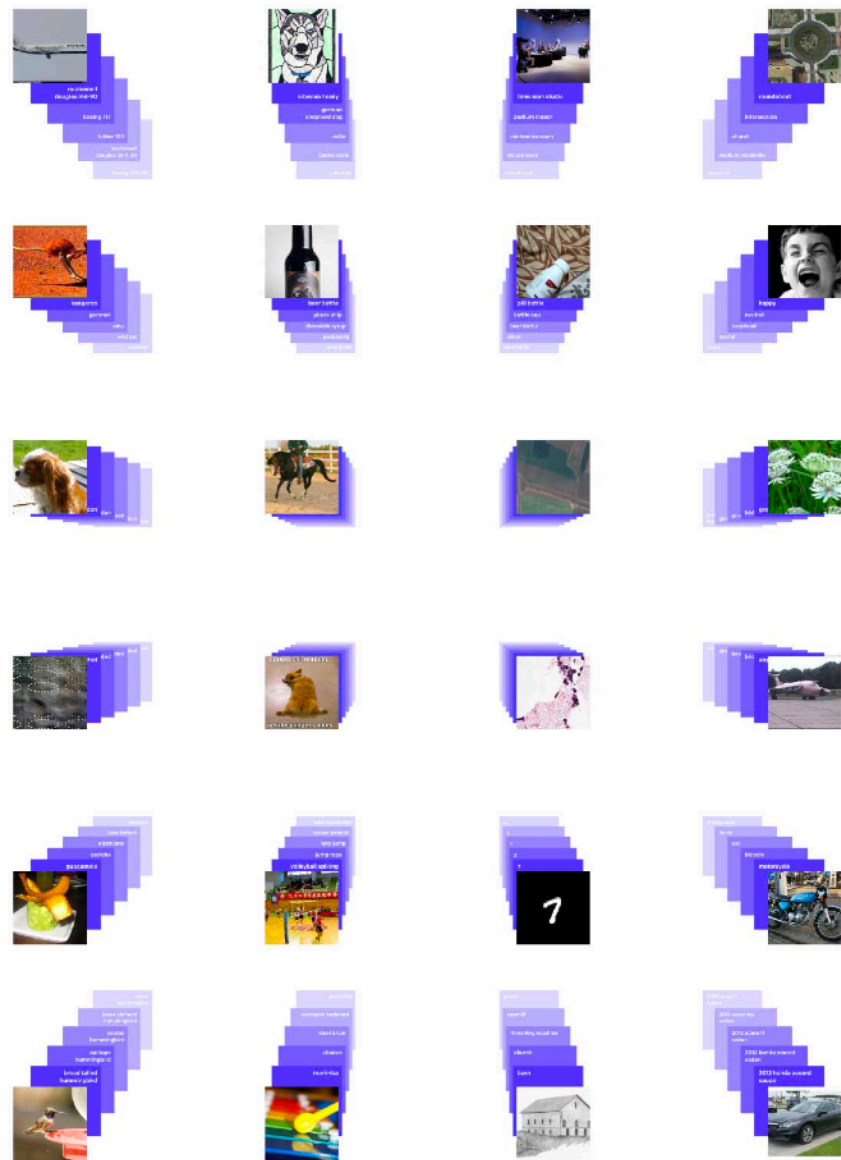
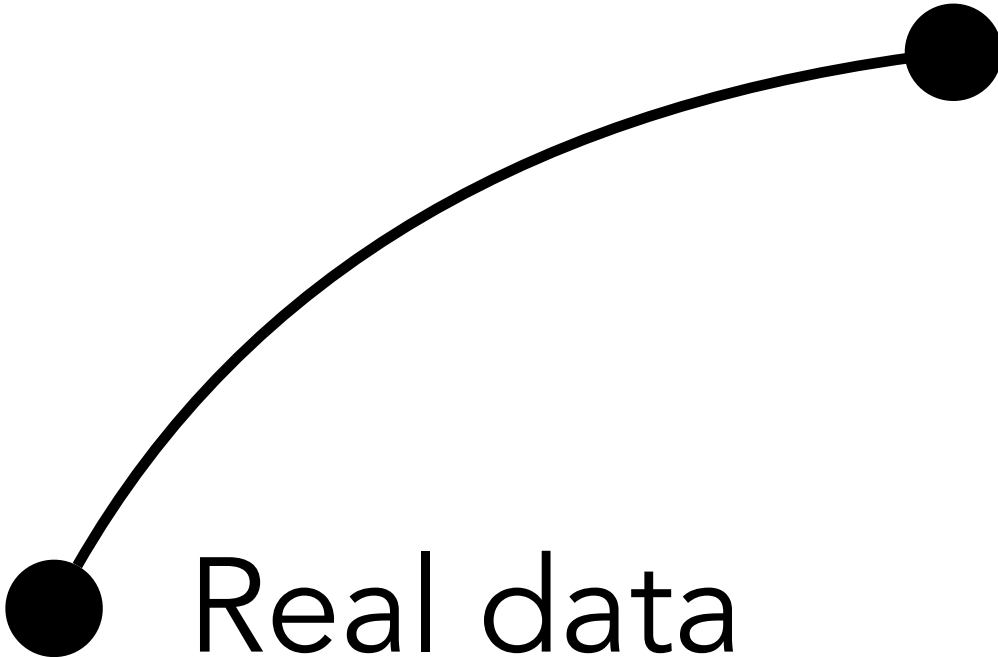
Data++

Machine learning is data-driven intelligence

Big data has been the major factor behind many of the recent advances in AI

Can we make *better* kinds of data?

Data sources for computer vision



**COVID-CT-MD: COVID-19 CT Scans**
Applicable in Machine Learning
[This collection is shared privately](#)




[Afshar et al., Nature 2021]

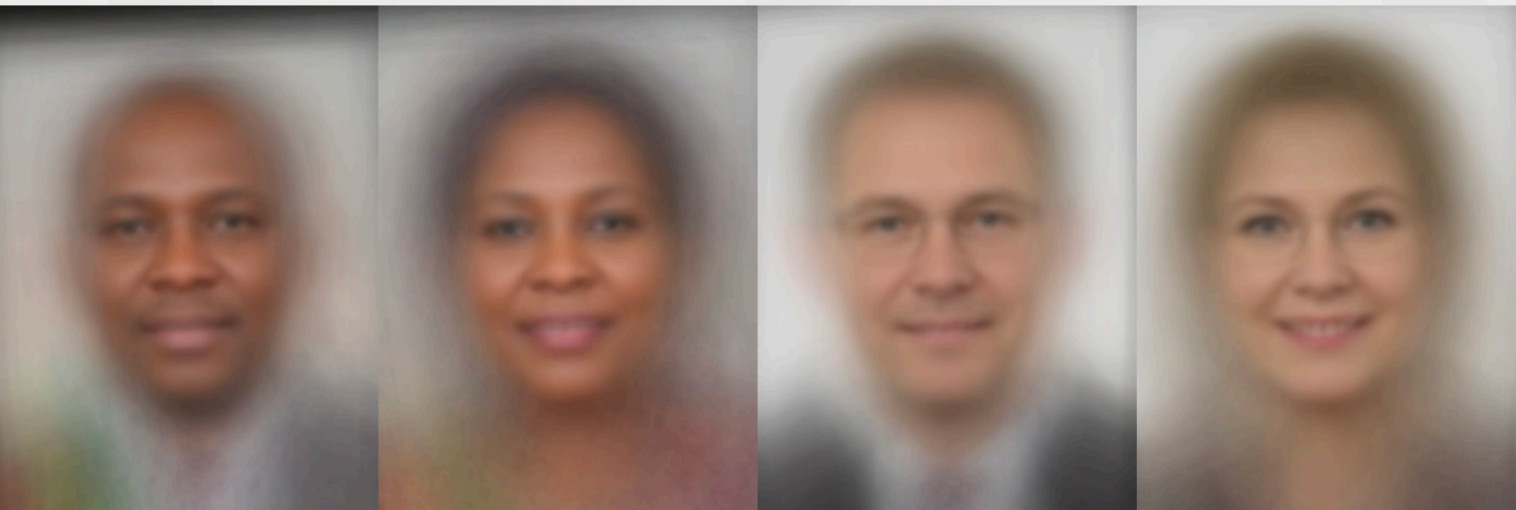


[Deng et al., CVPR 2009]



[Torralba & Efros, CVPR 2011]

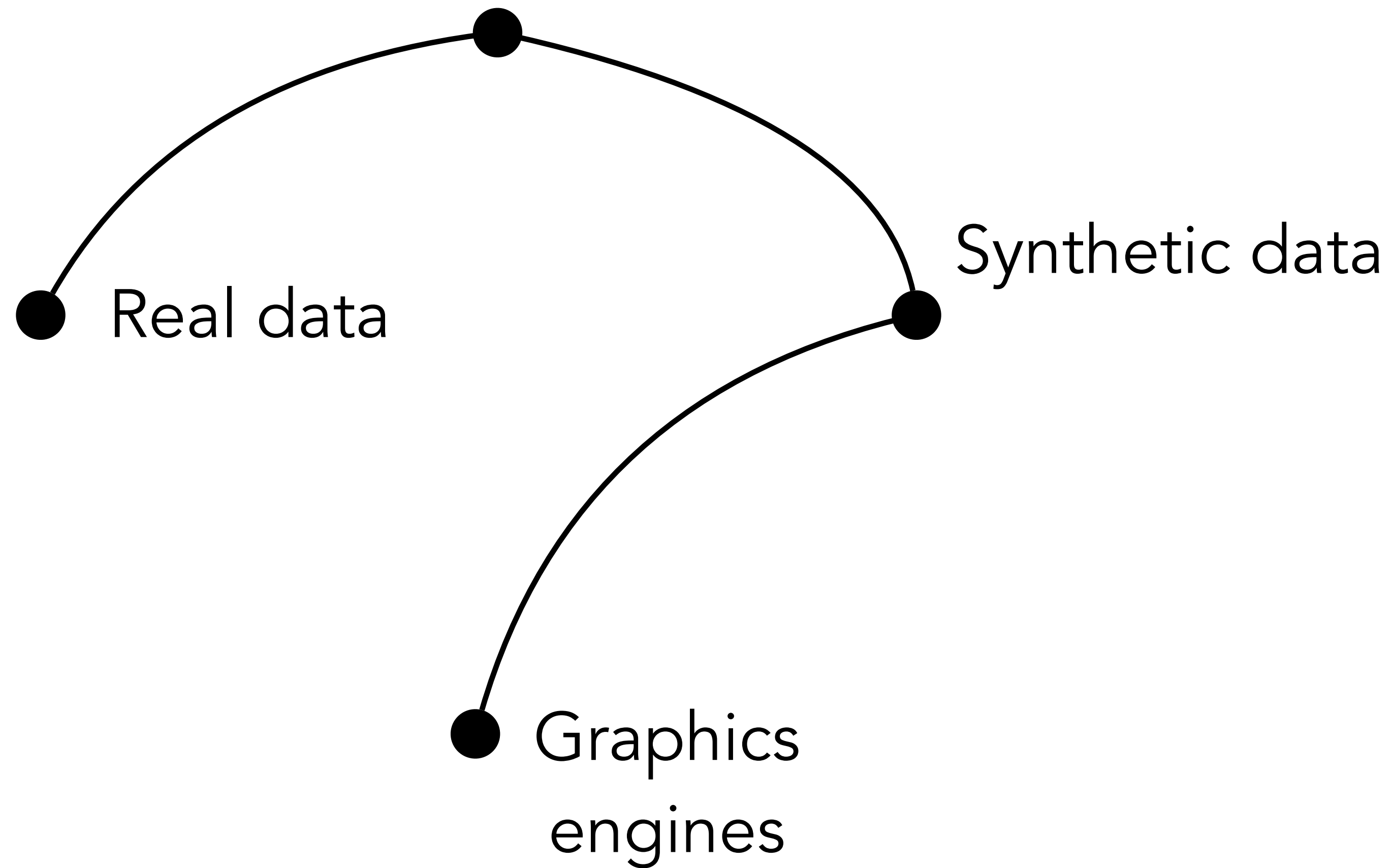
Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
 Microsoft	94.0%	79.2%	100%	98.3%	20.8%
 FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
 IBM	88.0%	65.3%	99.7%	92.9%	34.4%



[Gender Shades: Buolamwini & Gebru, FAccT 2018]

[CLIP: Radford*, Kim* et al. 2021]

Data sources for computer vision

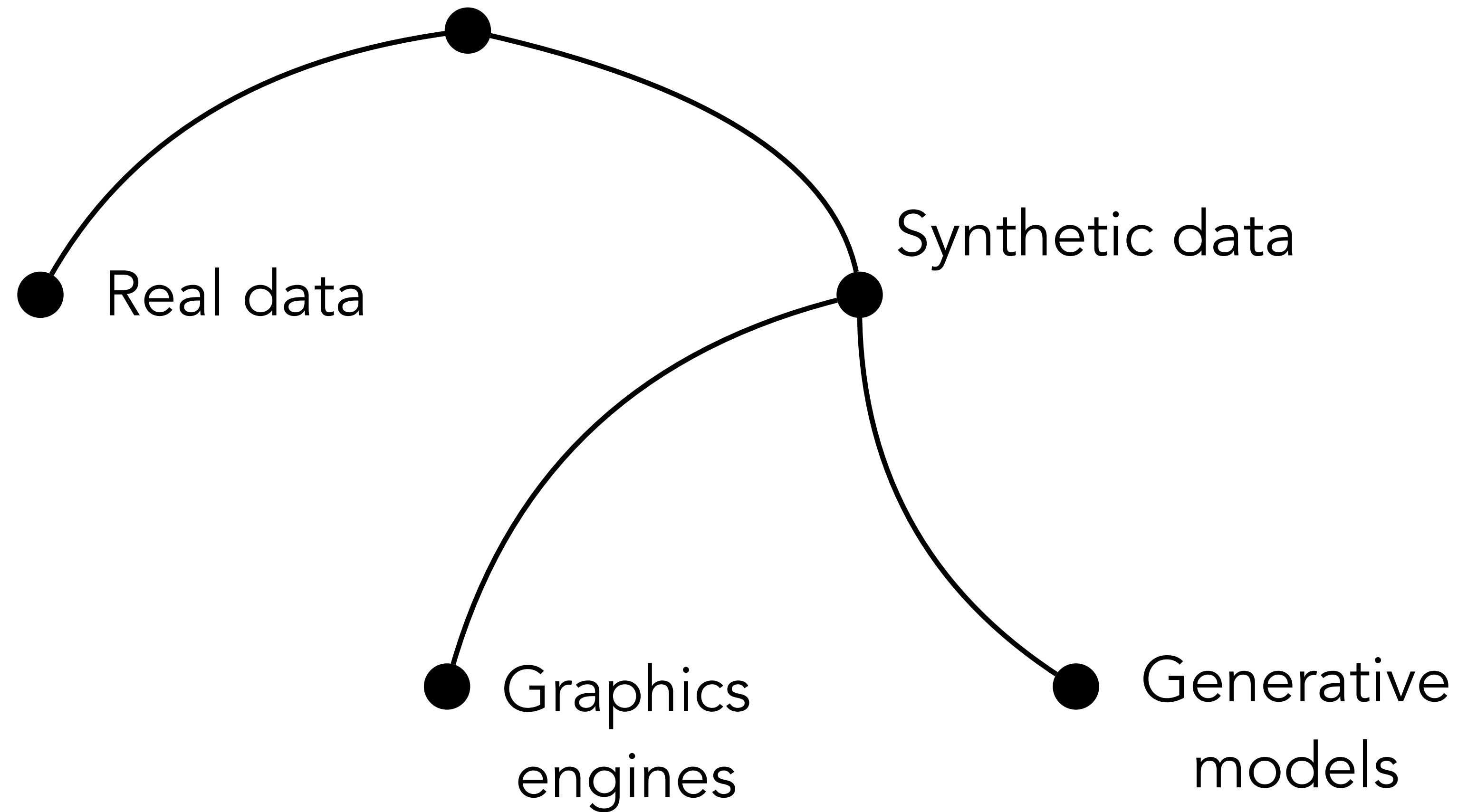


[Sintel: Butler, Wulff, Stanley, Black, ECCV 2012]



[OpenAI Dextrous Hand, 2018]

Data sources for computer vision

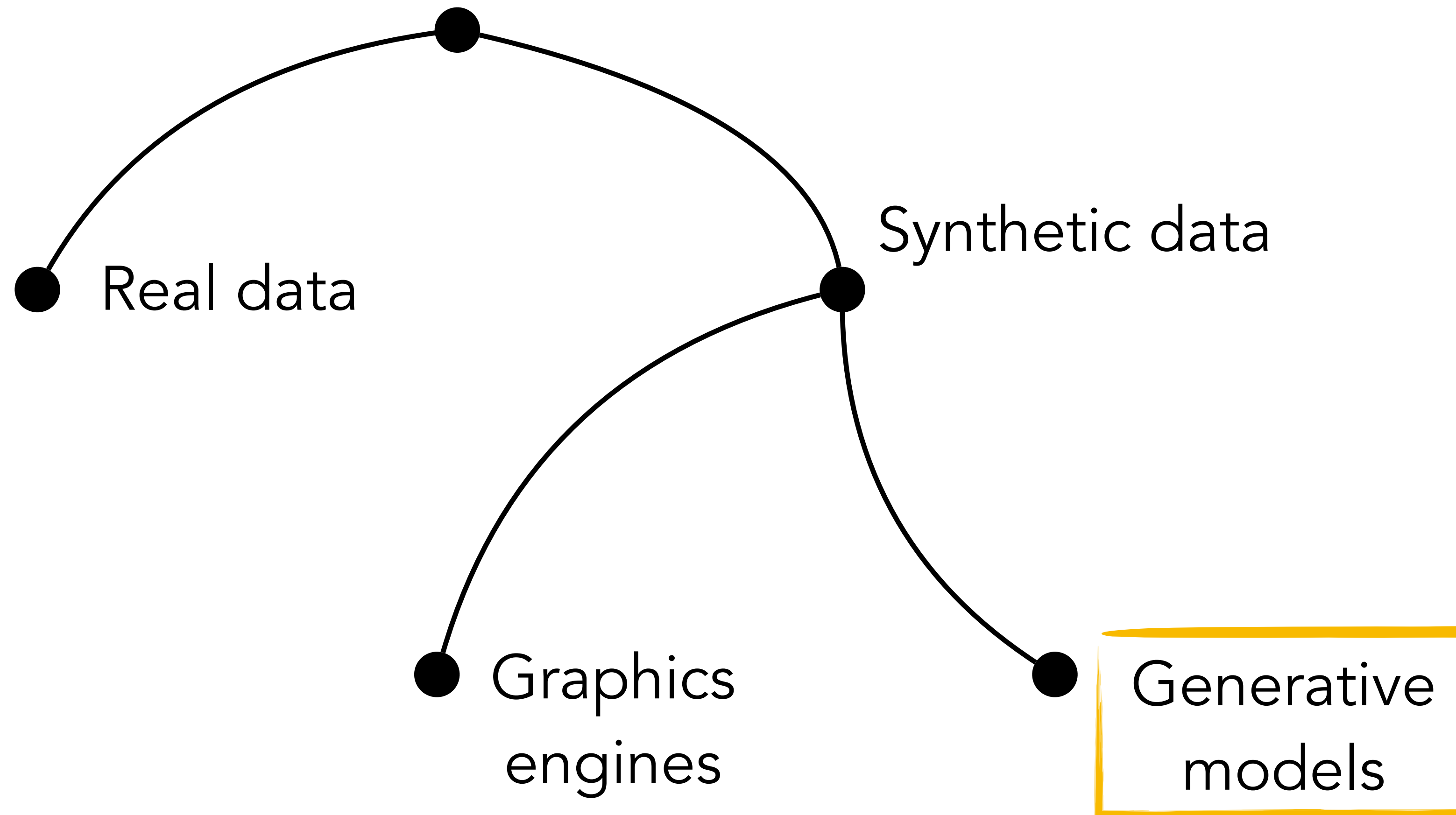


[BigGAN: Brock, Donahue, Simonyan, ICLR 2019]



[StyleGAN: Karras, Laine, Aila, CVPR 2019]

Data sources for computer vision



[BigGAN: Brock, Donahue, Simonyan, ICLR 2019]



[StyleGAN: Karras, Laine, Aila, CVPR 2019]

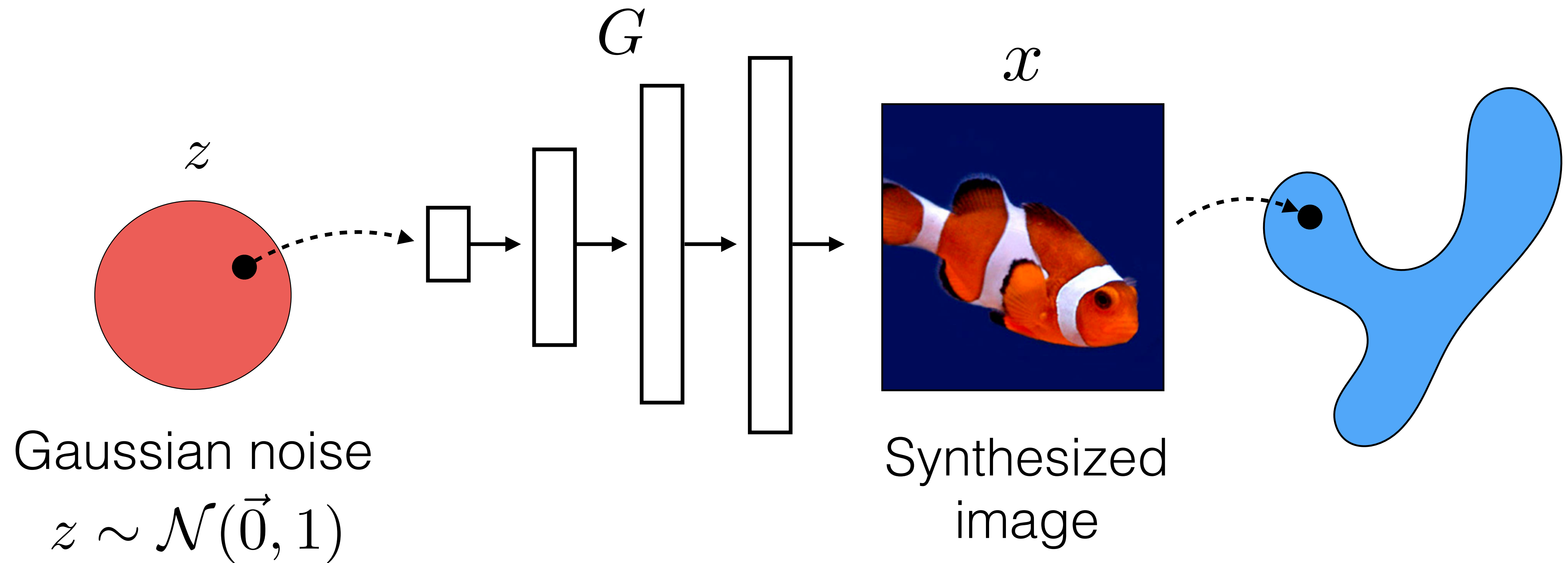
Generative models continuously approximate real images



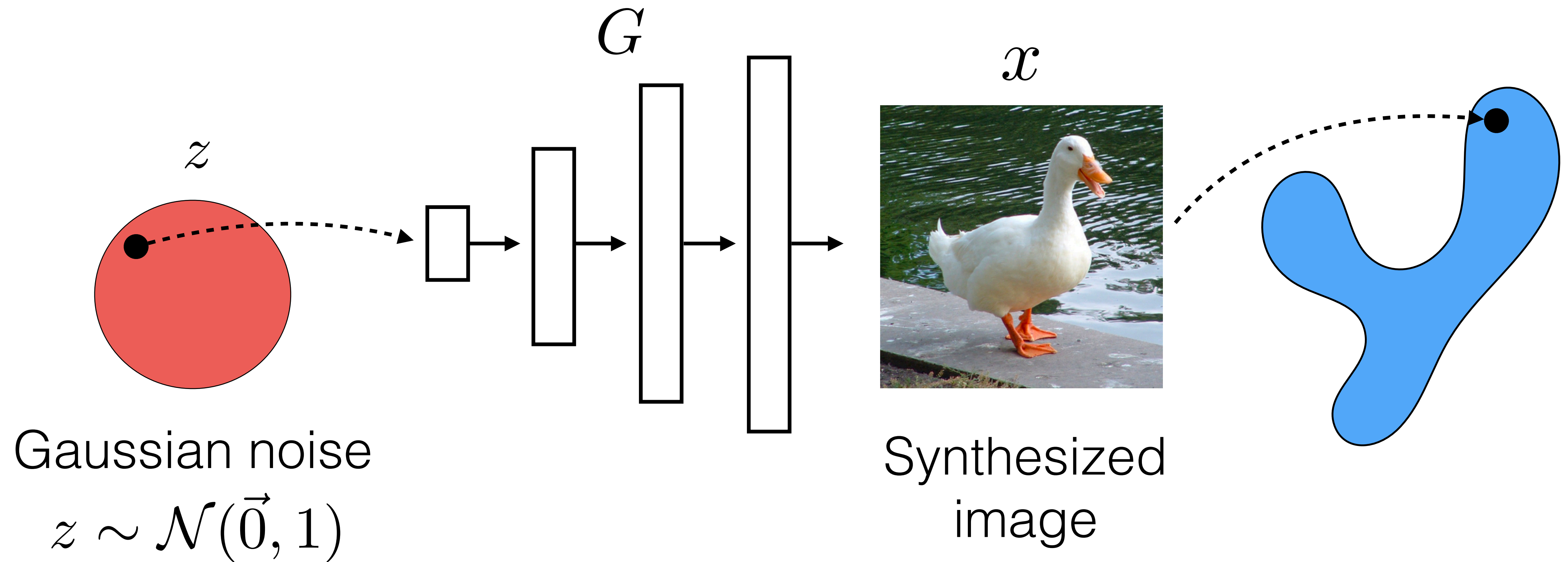
[Slide Credit: Lucy Chai]

[Goodfellow et al. 2014; StyleGAN2. Karras et al. 2020]

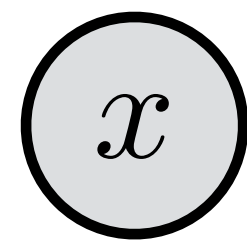
Generative Models



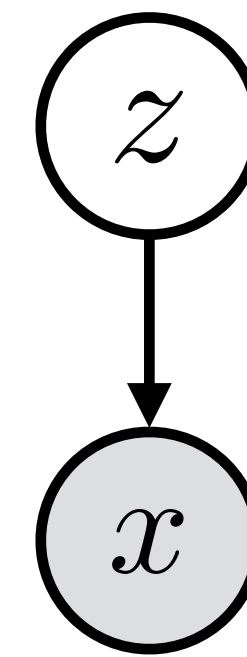
Generative Models



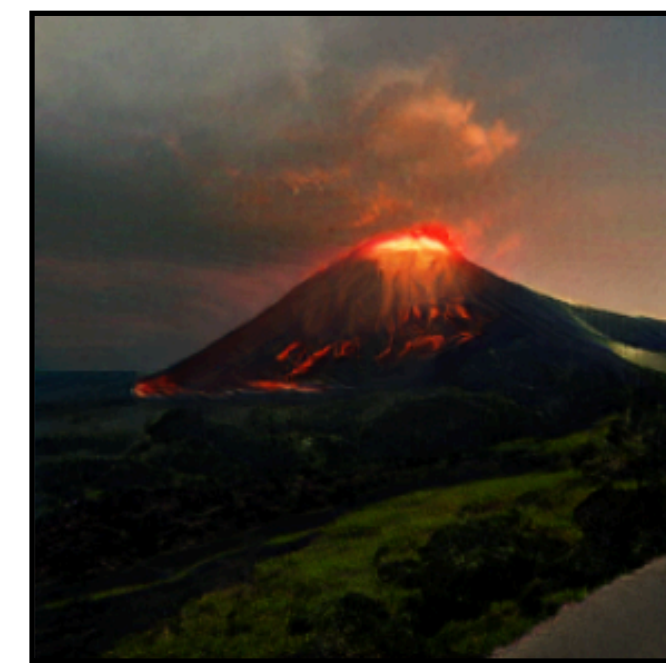
Generative Data



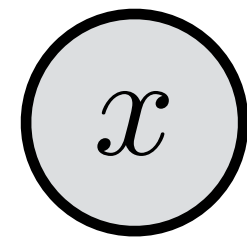
Data



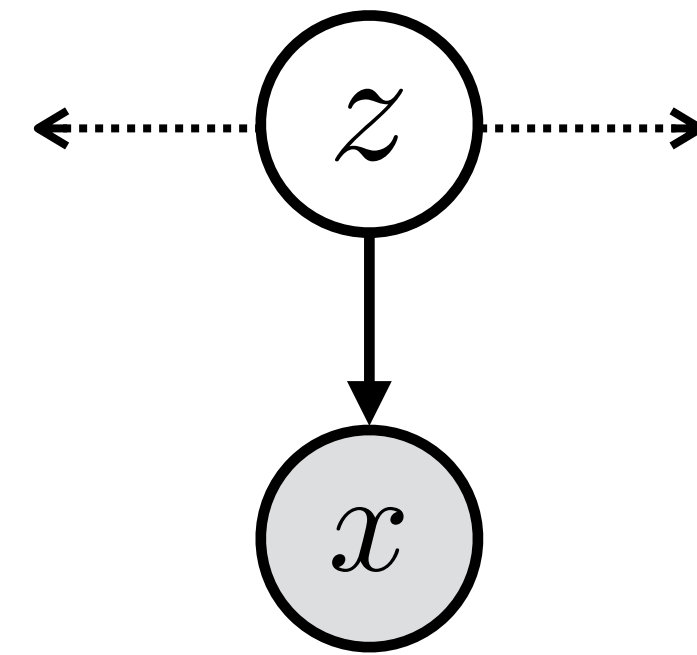
Data++



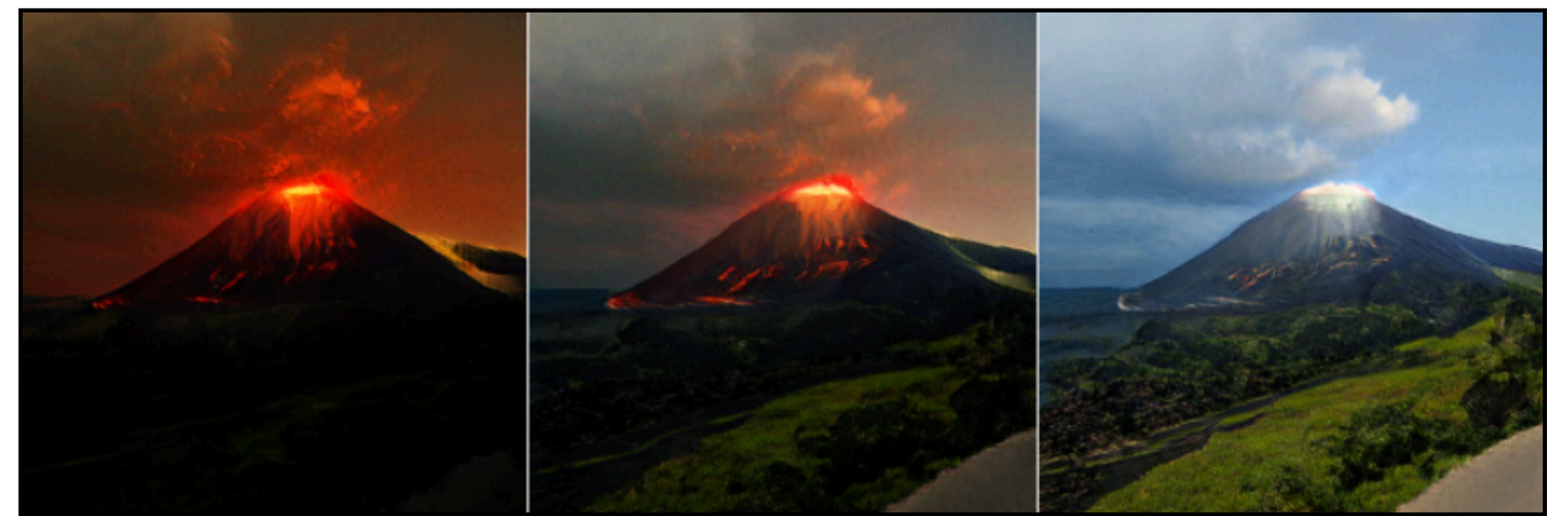
Generative Data



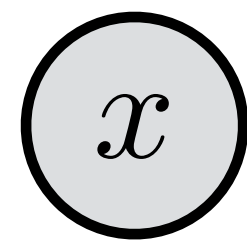
Data



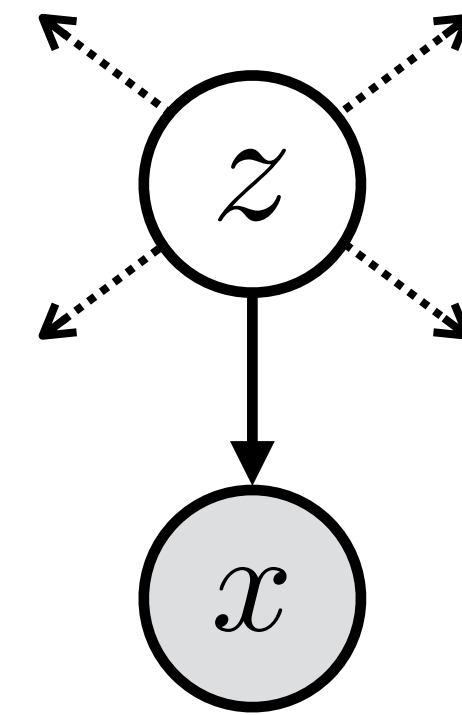
Data++



Generative Data



Data



Data++



Data++: making data a first class object

$$\mathbb{X} = \{x, z, G, G^{-1}\}$$

Interpolation: $\alpha\mathbb{X}_1 + (1 - \alpha)\mathbb{X}_2 \rightarrow \mathbb{X}_3$

Manipulation: $\mathbb{X}_1 + w \rightarrow \mathbb{X}_2$ [Goetschalckx*, Andonian*, Oliva, Isola, ICCV 2019]
[Jahanian*, Chai*, Isola, ICLR 2020]

Composition: $\mathbb{X}_1[m] + \mathbb{X}_2[1 - m] \rightarrow \mathbb{X}_3$ [Chai, Wulff, Isola, ICLR 2021]

Optimization: $\arg \min_{\mathbb{X}} f(\mathbb{X})$ [Lin, Florence, Barron, et al. , IROS 2021]

→ Graphics, visualization, data aug, counterfactual reasoning, ...

Interpolation in data space

Data space
(Natural image manifold)

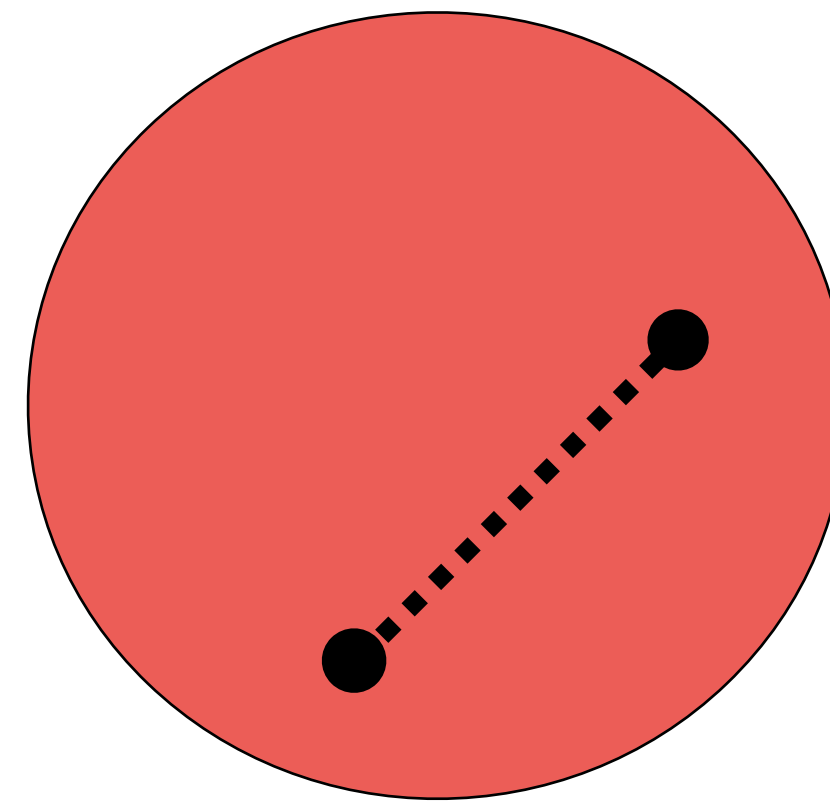
\mathbf{X}



Interpolation in latent space

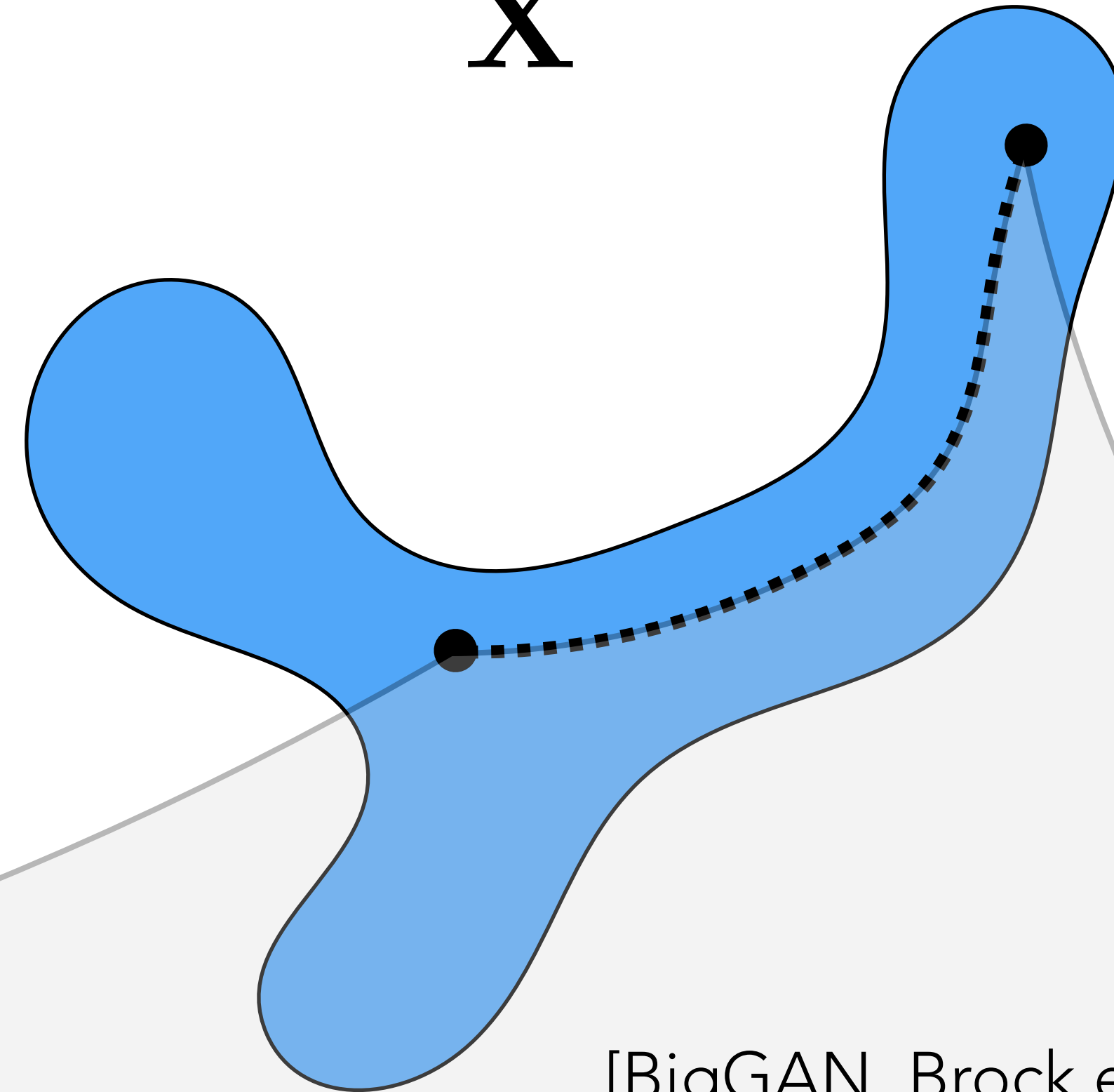
Latent space
(Gaussian)

\mathbf{Z}



Data space
(Natural image manifold)

\mathbf{X}



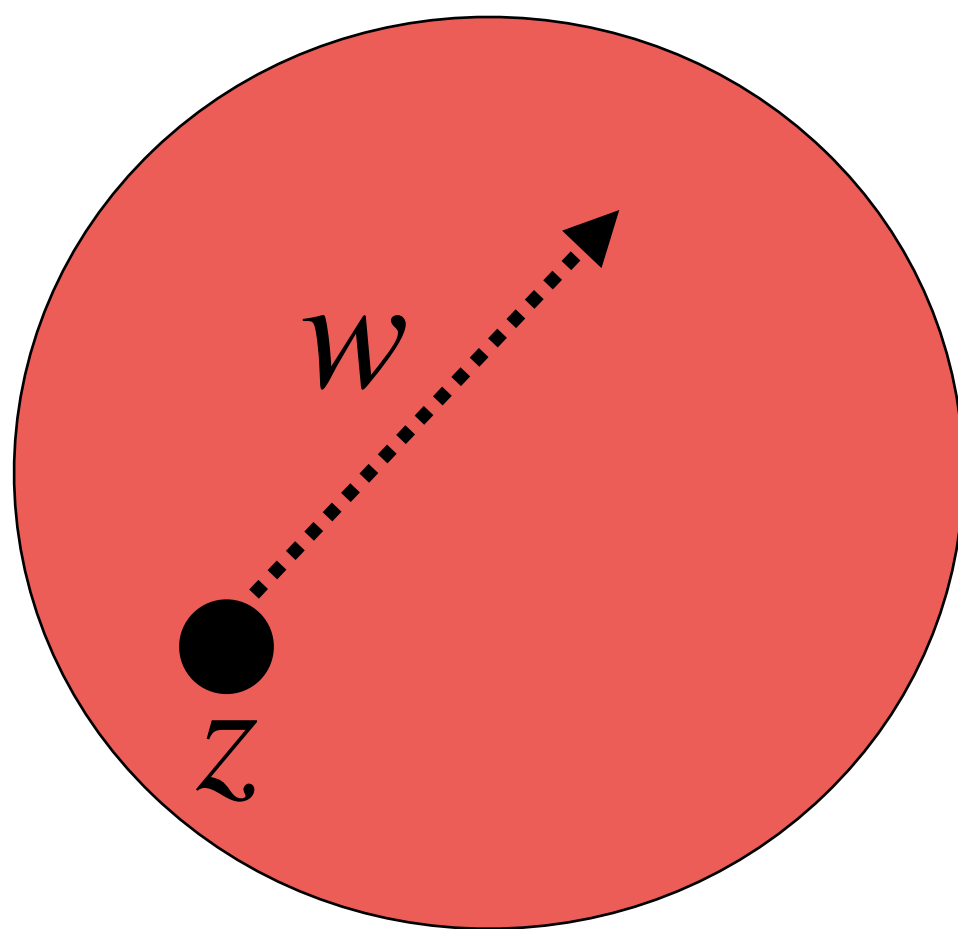
[BigGAN, Brock et al. 2018]



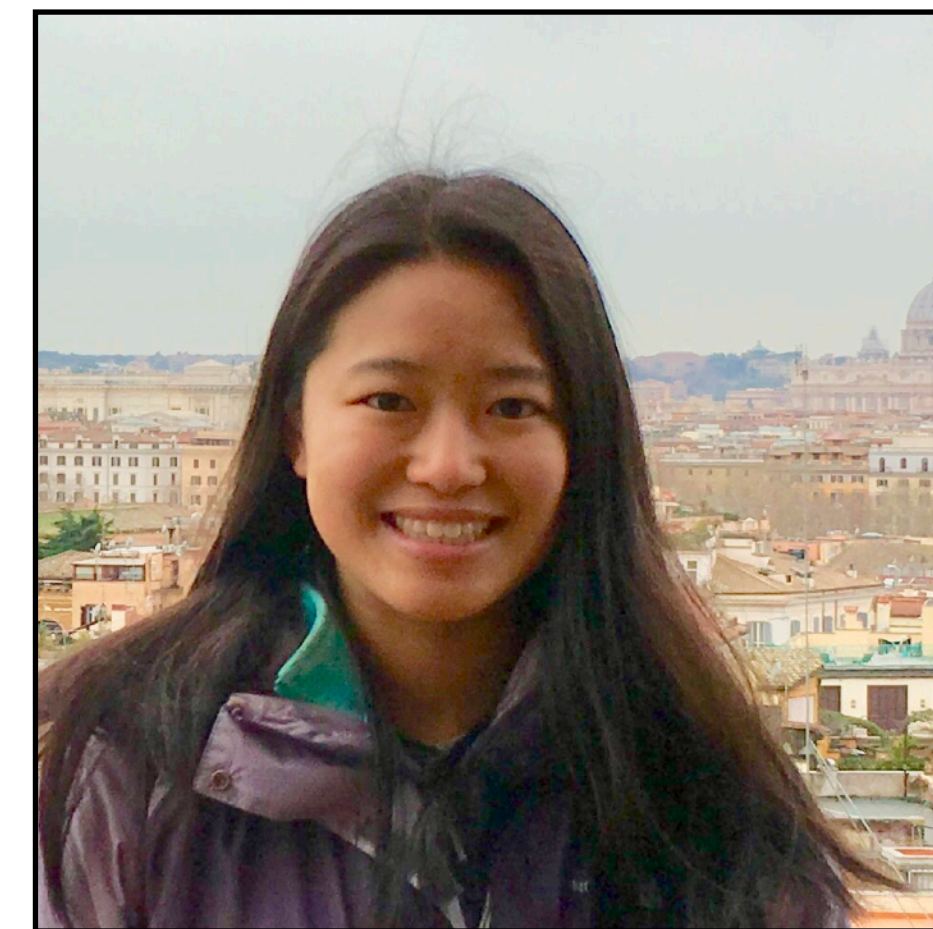
On the “steerability” of Generative Adversarial Networks

[Jahanian*, Chai*, Isola, ICLR 2020]

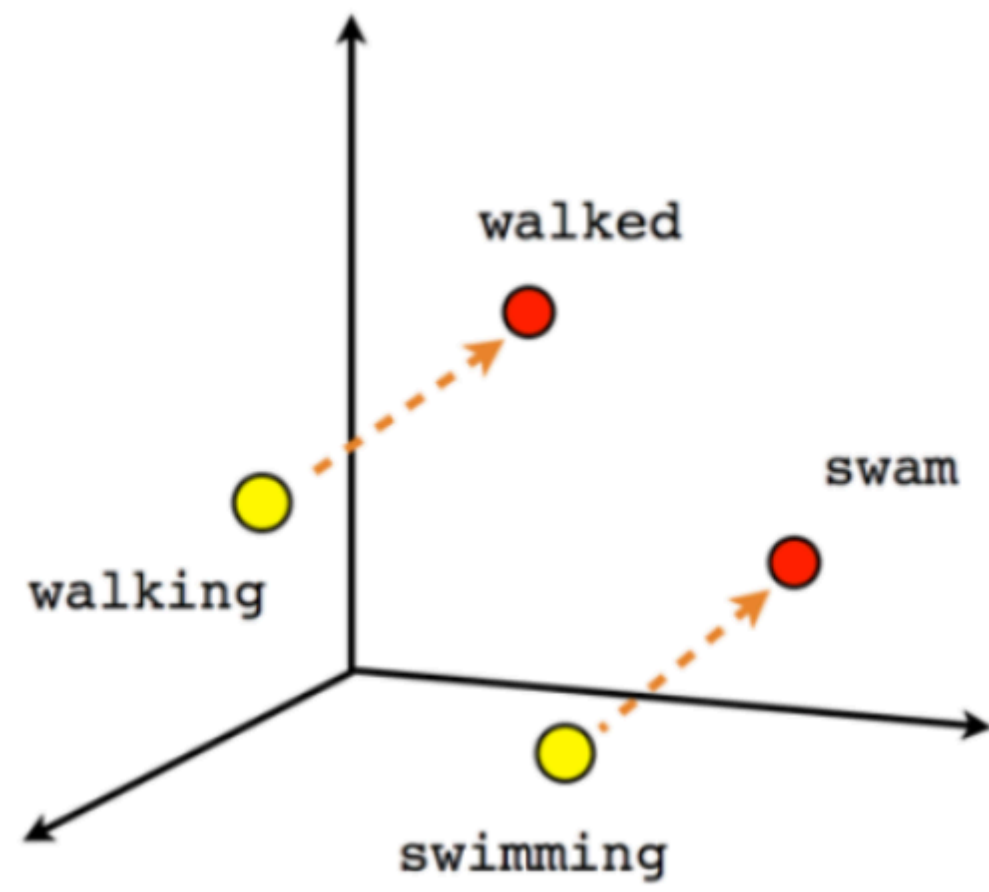
To what extent can we meaningfully **manipulate** data++ via vector arithmetic? $\mathbb{X}_1 + w \rightarrow \mathbb{X}_2$



Ali Jahanian

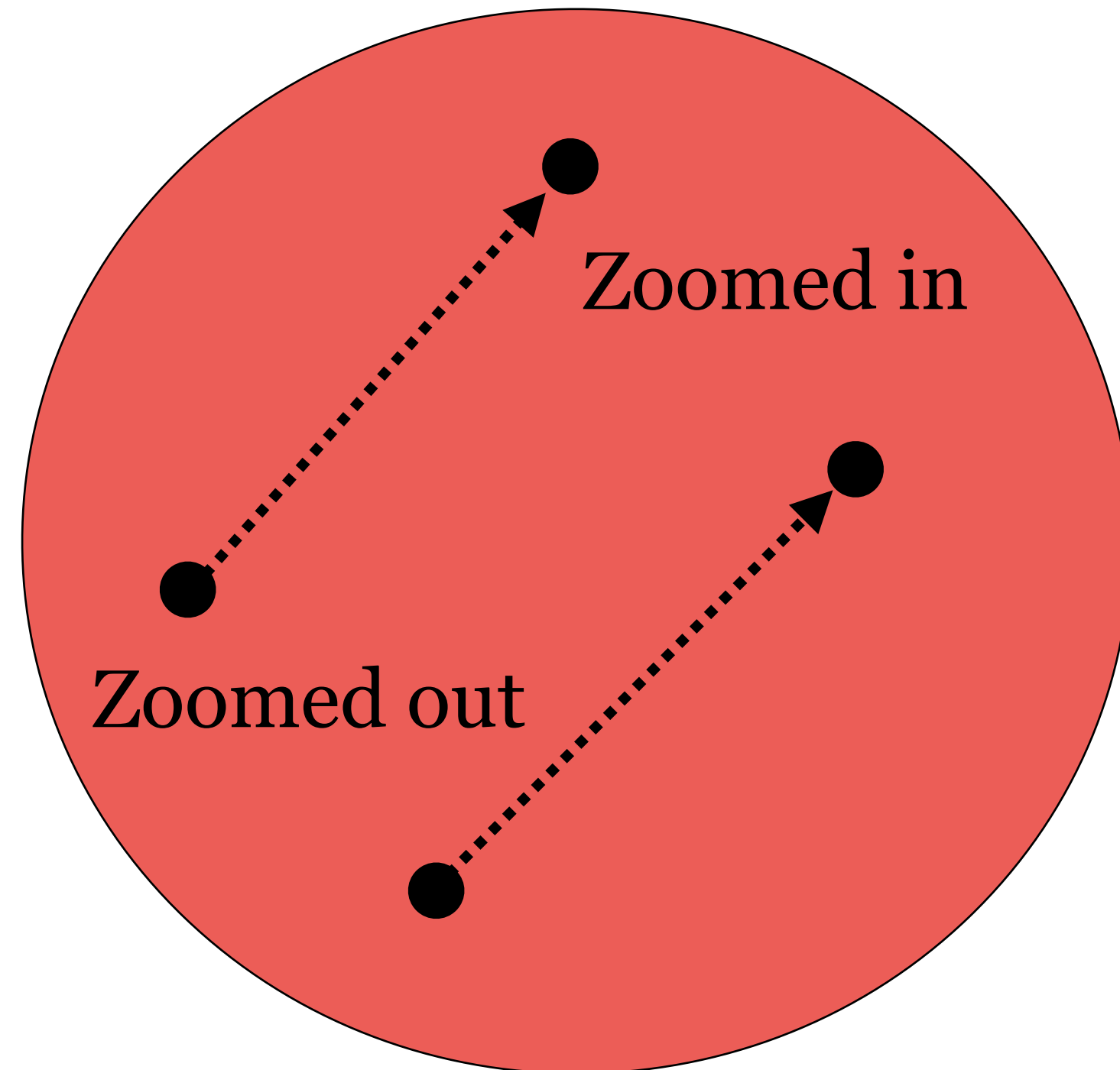


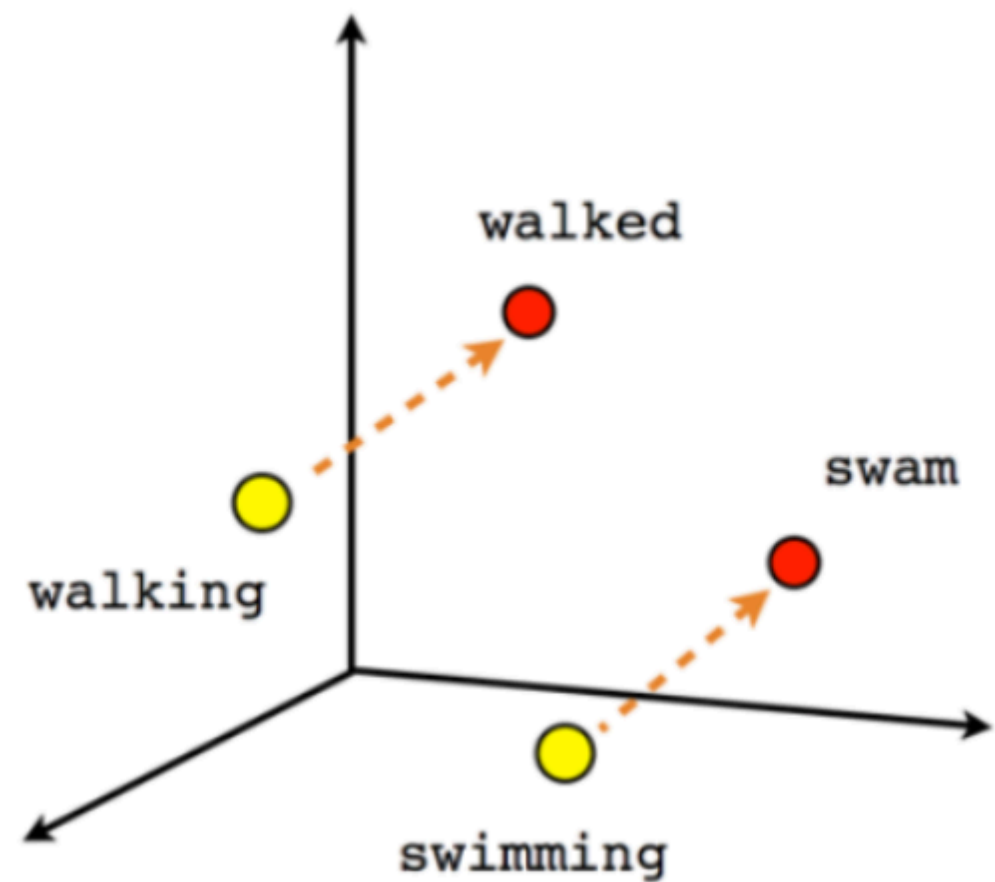
Lucy Chai



[word2vec, Mikolov et al., 2013]

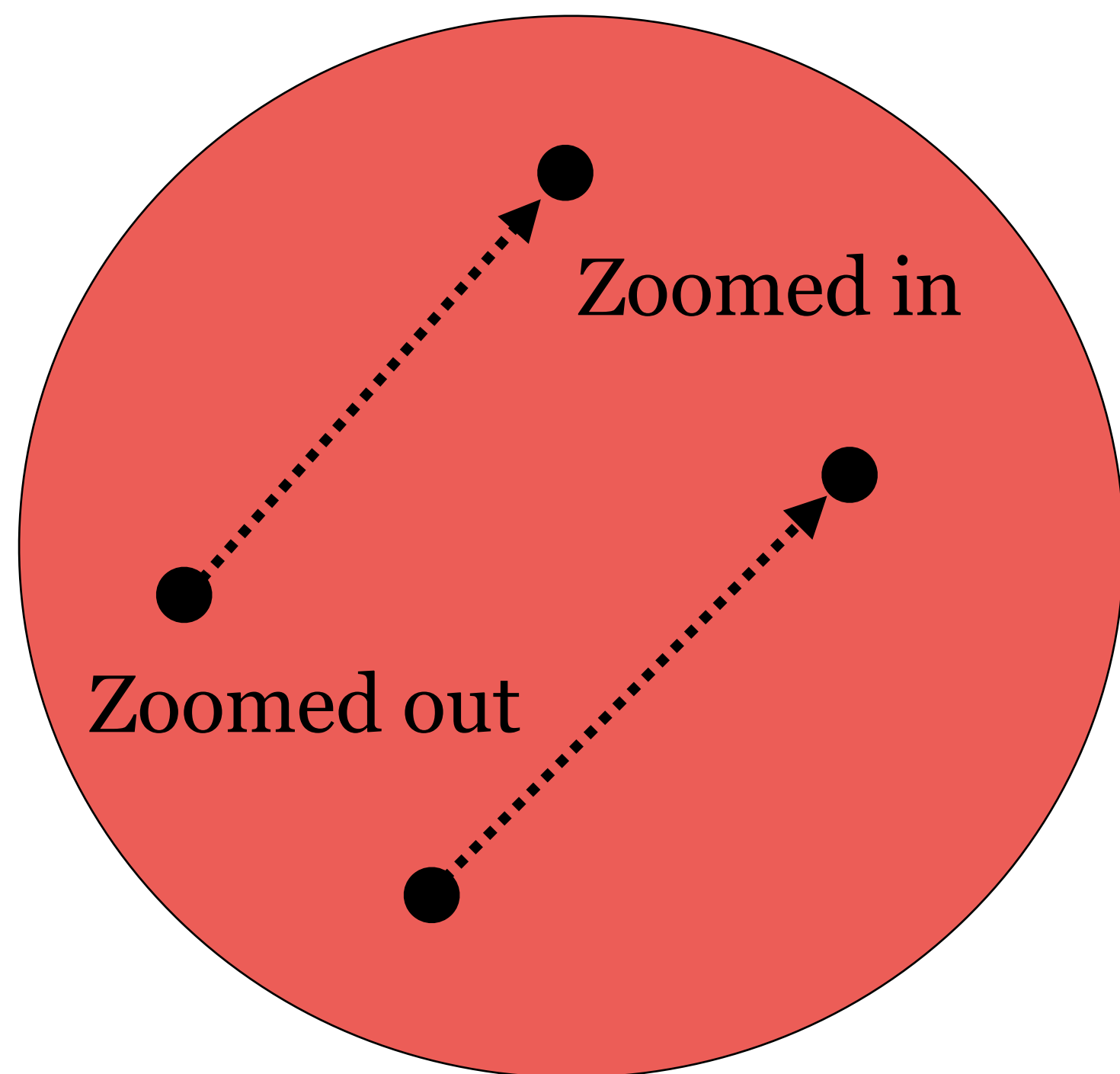
[DCGAN, Radford, Metz, Chintala, 2015]





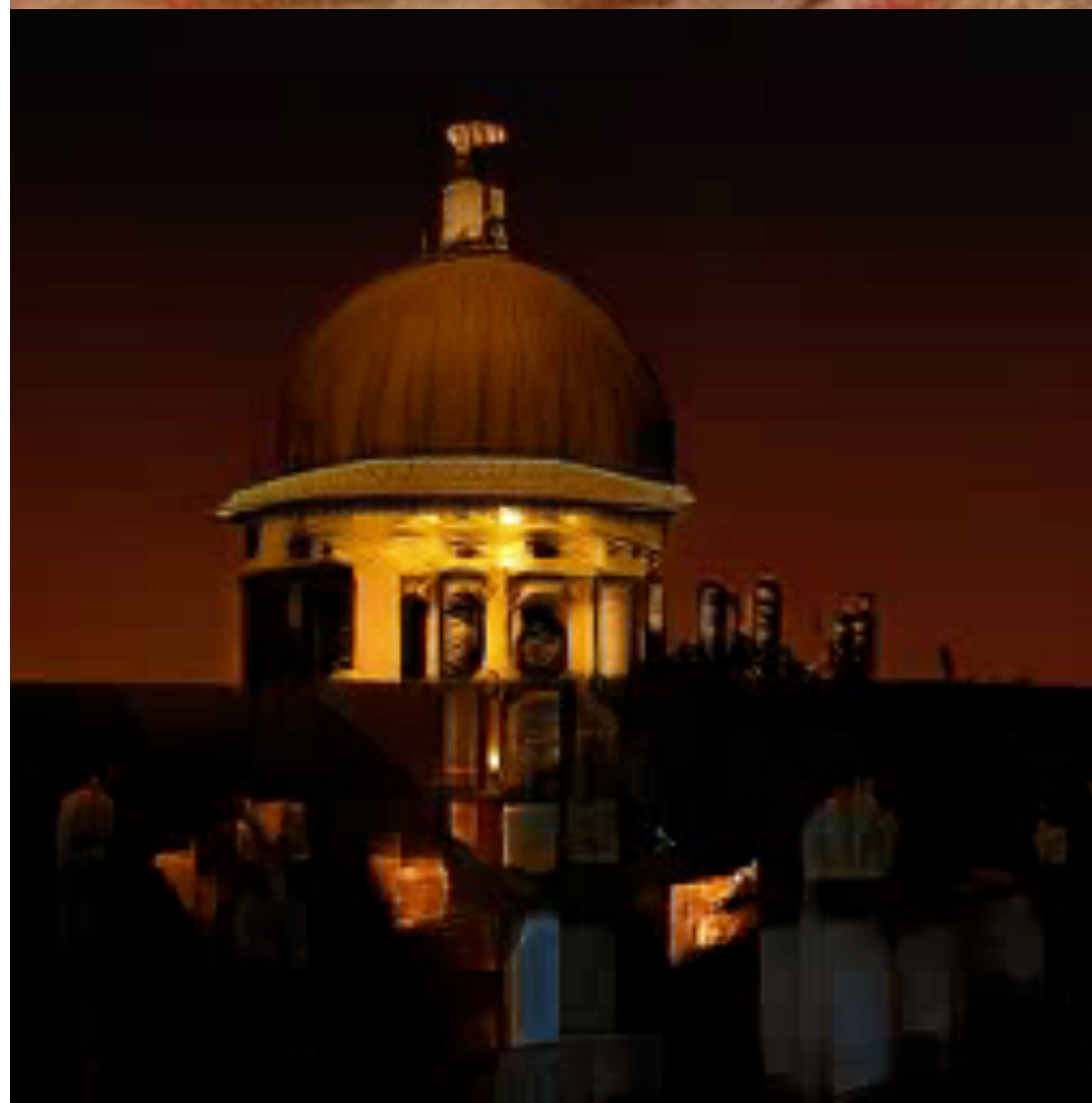
[word2vec, Mikolov et al., 2013]

[DCGAN, Radford, Metz, Chintala, 2015]



Zoom

Shift



Brighten

Darken

Data++ supports counterfactual reasoning

i.e. “What would it have looked like if ...?”

Observation



Counterfactual hallucinations



see also: [Mao, Cha, Gupta, Wang, Yang, Vondrick, 2020]

[Sauer & Geiger, 2021]

[Liu, Kailkhura, Loveland, Han, 2019]

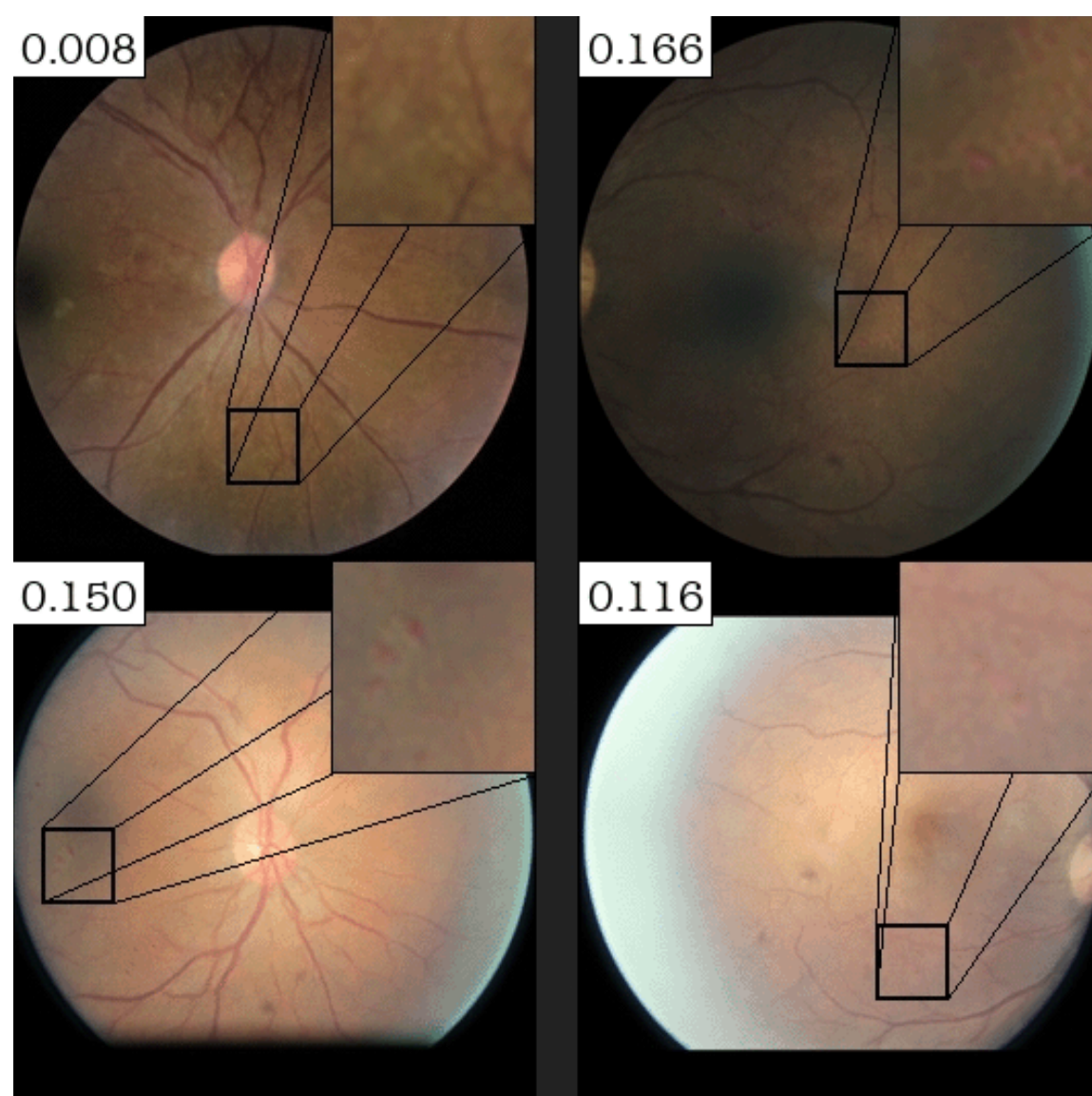
[Goetschalckx, Andonian, Oliva, Isola, 2019]

[Oktay, Vondrick, Torralba, 2018]

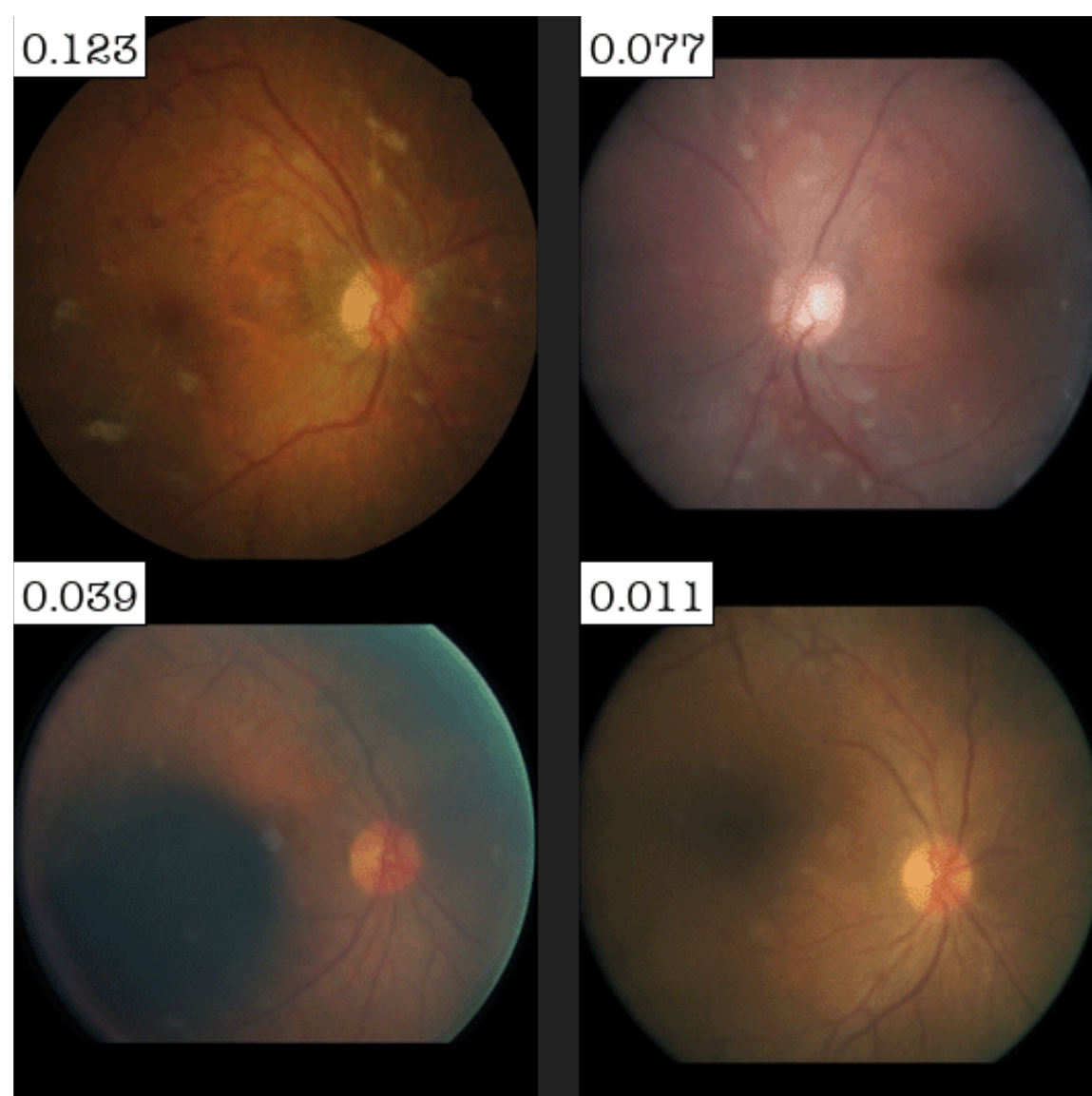
...

Explaining in Style: Training a GAN to Explain a Classifier

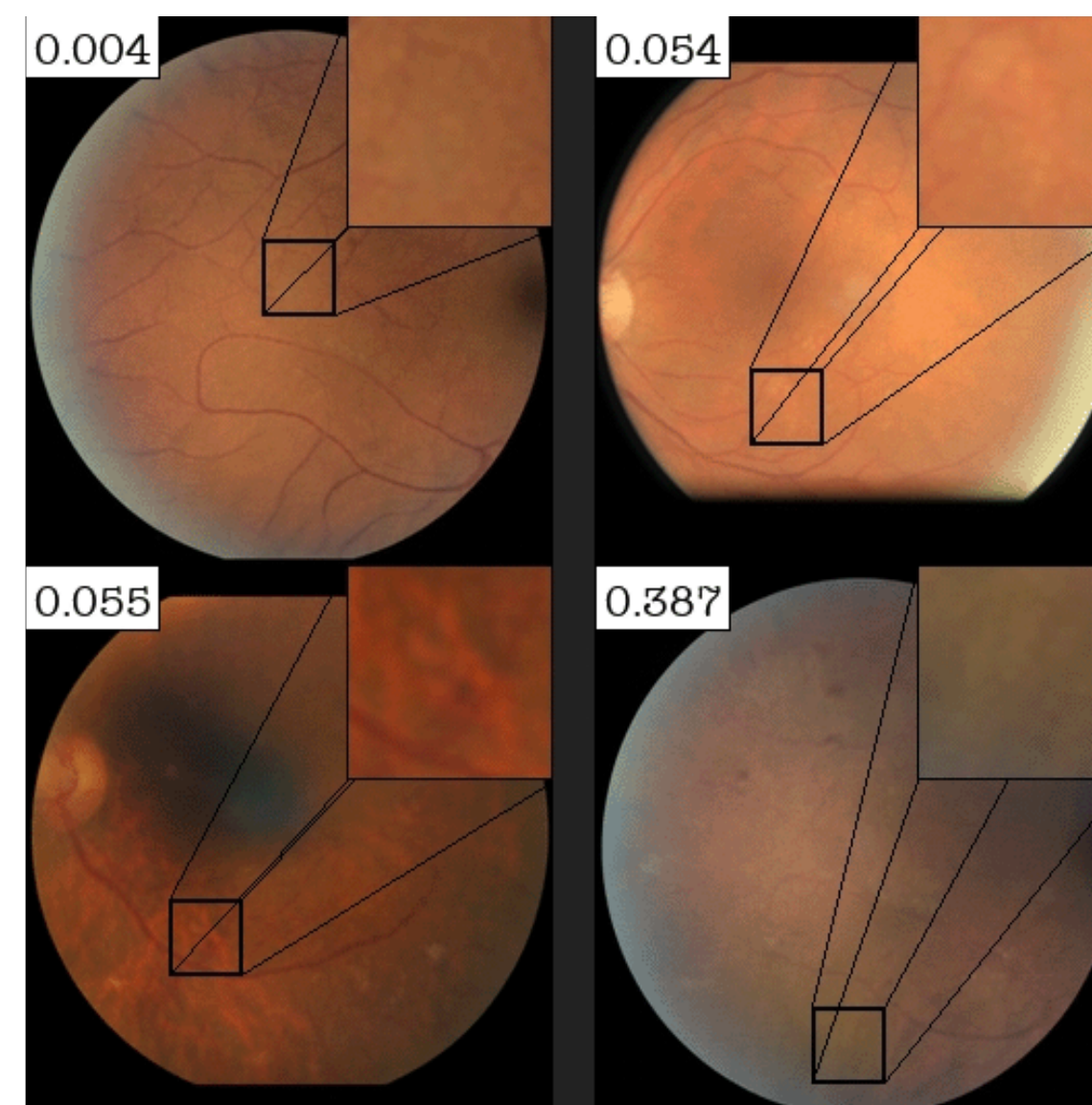
Oran Lang*, Yossi Gandelsman*, Michal Yarom*, Yoav Wald*, Gal Elidan, Avinatan Hassidim, William T. Freeman, Phillip Isola, Amir Globerson, Michal Irani, Inbar Mosseri
ICCV 2021



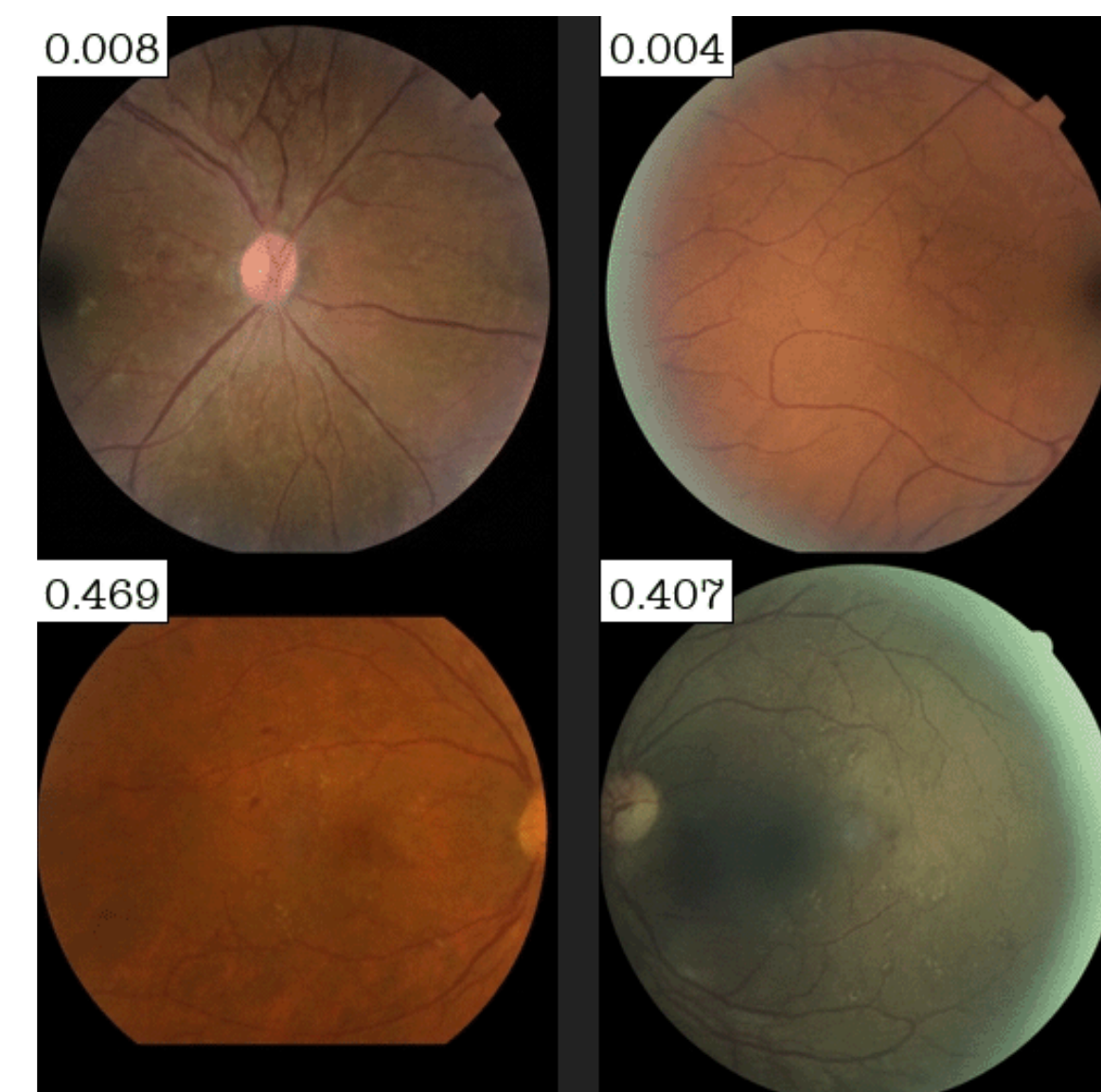
Attribute #1
“Exudates”



Attribute #2
“Cotton Wool”



Attribute #3
“Hemorrhages”



Attribute #4
“Clustered Exudates”

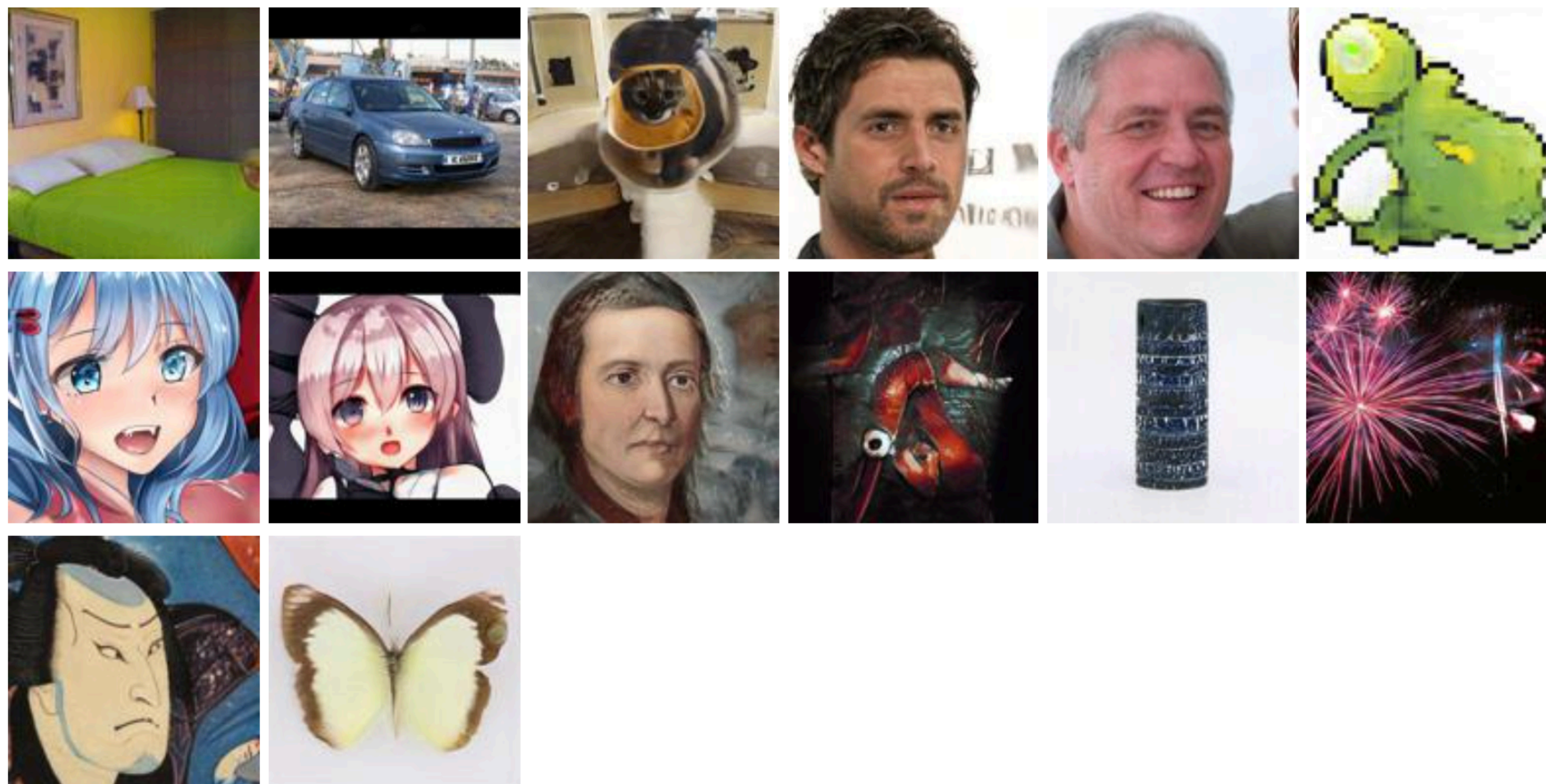
Retinal Fundus Classifier:

The Era of Big Data —> The Era of Big Models?

Awesome Pretrained StyleGAN

A collection of pre-trained [StyleGAN](#) models trained on different datasets at different resolution.

For the equivalent collection for StyleGAN 2, see [this repo](#)



The GPT-3 and DALL-E datasets are not public, but model samples are / likely will be.

In the future, will models be our primary interface to data?

Also happening in biotech, healthcare, robotics, etc.

[<https://github.com/justinpinkney/awesome-pretrained-stylegan>]

Dall-e 2