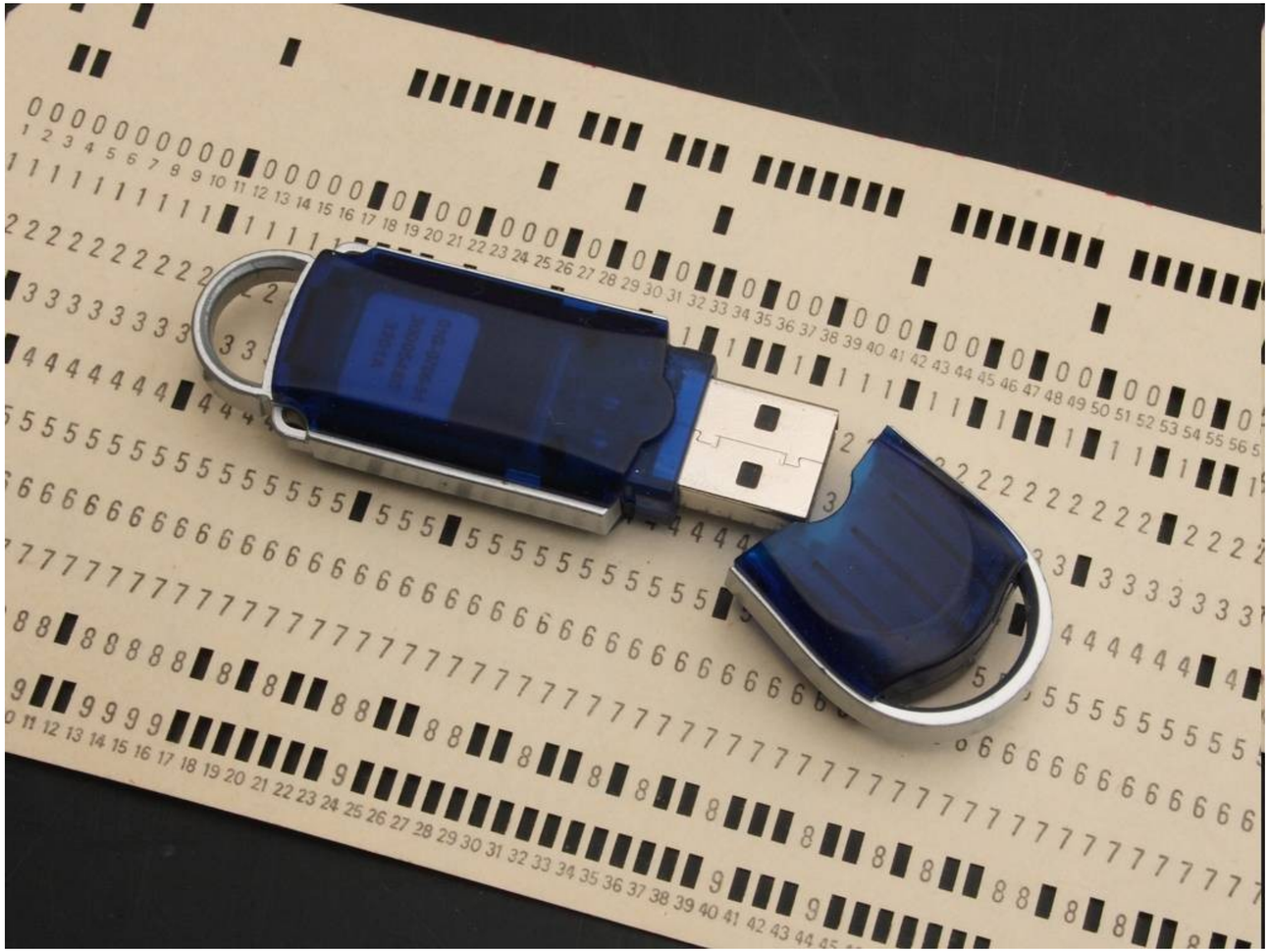
“Data, Data, Data… Watson, I need Data!”



CS180: Intro to Computer Vision & Comp. Photography
Efros & Ng, UC Berkeley, Fall 2026

My issues with “The Bitter Lesson”?

The Bitter Lesson

Rich Sutton

March 13, 2019

The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin. The ultimate reason for this is Moore's law, or rather its generalization of continued exponentially falling cost per unit of computation. Most AI research has been conducted as if the computation available to the agent were constant (in which case leveraging human knowledge would be one of the only ways to improve performance) but, over a slightly longer time than a typical research project, massively more computation inevitably becomes available. Seeking an improvement that makes a difference in the shorter term, researchers seek to leverage their human knowledge of the domain, but the only thing that matters in the long run is the leveraging of computation. These two need not run counter to each other, but in practice they tend to. Time spent on one is time not spent on the other. There are psychological commitments to investment in one approach or the other. And the human-knowledge approach tends to complicate methods in ways that make them less suited to taking advantage of general methods leveraging computation. There were many examples of AI researchers' belated learning of this bitter lesson, and it is instructive to review some of the most prominent.

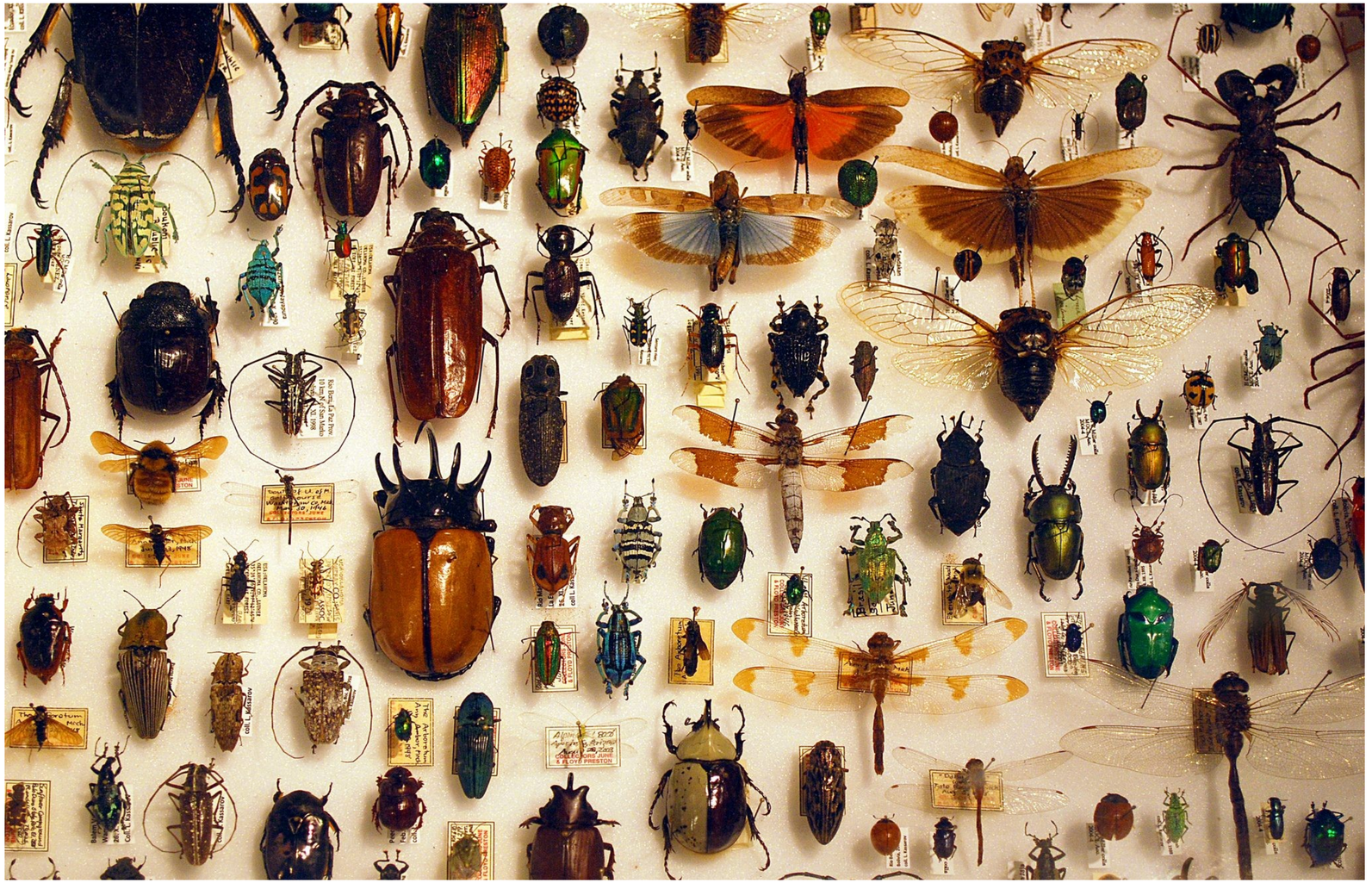


- Is it actually **bitter**?
- Is it even the **right** lesson?

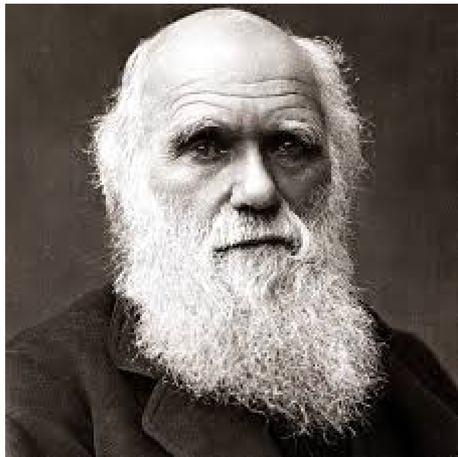


Infinite complexity

Biology’s Bitter Lesson?



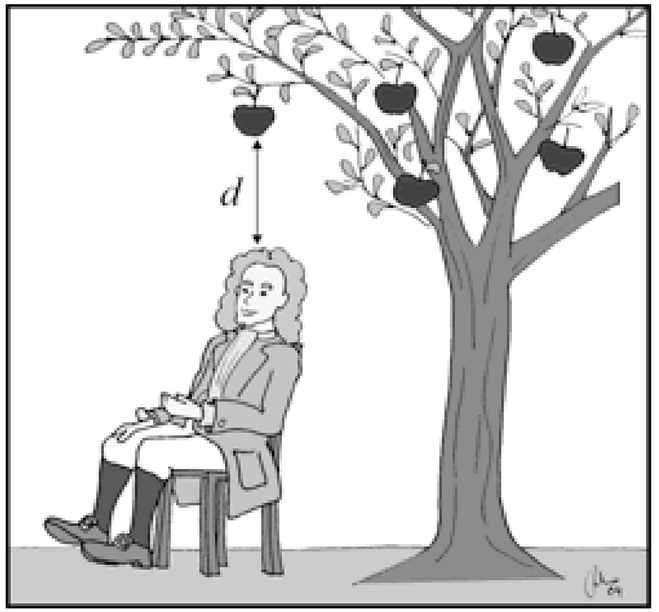
Everything is simple, once you go meta!

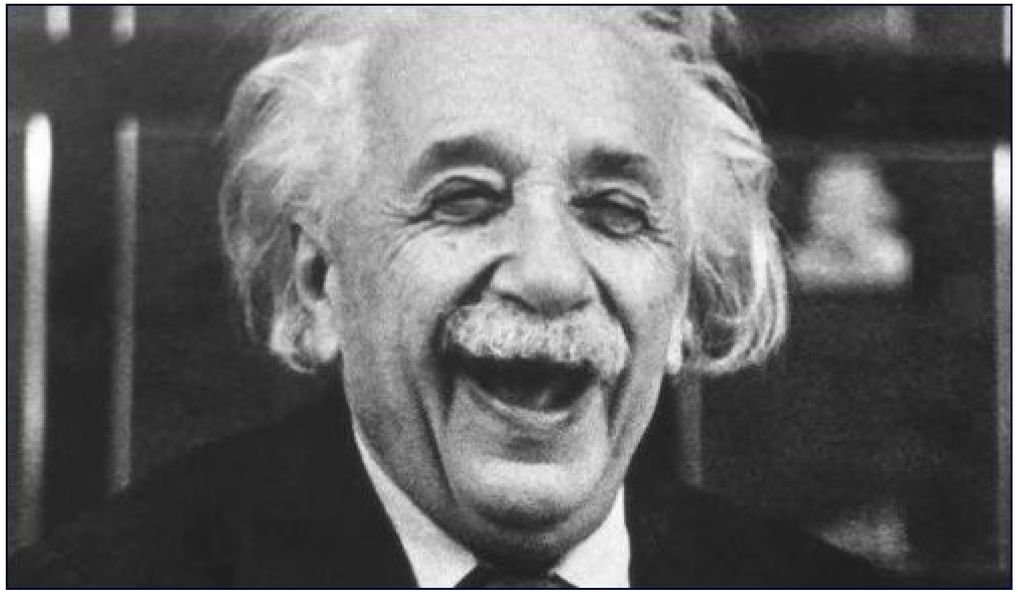


Physics’ Bitter Lesson?

Physics before Newton is ‘miscellaneous early efforts’

- Don Geman, circa 2007





But was Einstein embittered?

Is it even the **right** lesson?

The Bitter Lesson

Rich Sutton

March 13, 2019

The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin. The ultimate reason for this is Moore's law, or rather its generalization of continued exponentially falling cost per unit of computation. Most AI research has been conducted as if the computation available to the agent were constant (in which case leveraging human knowledge would be one of the only ways to improve performance) but, over a slightly longer time than a typical research project, massively more computation inevitably becomes available. Seeking an improvement that makes a difference in the shorter term, researchers seek to leverage their human knowledge of the domain, but the only thing that matters in the long run is the leveraging of computation. These two need not run counter to each other, but in practice they tend to. Time spent on one is time not spent on the other. There are psychological commitments to investment in one approach or the other. And the human-knowledge approach tends to complicate methods in ways that make them less suited to taking advantage of general methods leveraging computation. There were many examples of AI researchers' belated learning of this bitter lesson, and it is instructive to review some of the most prominent.



- *"the actual contents of minds are **tremendously, irredeemably complex**; we should stop trying to find simple ways to think about the contents of minds"*

A **Humbling** Lesson for Humans

- Humans turned out to be far simpler than we thought!
- 99% of the time, we are just nearest neighbor machines:
  - walking cliches
  - pastiches from high school reading lists
- We are in the next Copernican Revolution. Hold on tight…



Is it even the **right** lesson?

The Bitter Lesson

Rich Sutton

March 13, 2019

The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin. The ultimate reason for this is Moore's law, or rather its generalization of continued exponentially falling cost per unit of computation. Most AI research has been conducted as if the computation available to the agent were constant (in which case leveraging human knowledge would be one of the only ways to improve performance) but, over a slightly longer time than a typical research project, massively more computation inevitably becomes available. Seeking an improvement that makes a difference in the shorter term, researchers seek to leverage their human knowledge of the domain, but the only thing that matters in the long run is the leveraging of computation. These two need not run counter to each other, but in practice they tend to. Time spent on one is time not spent on the other. There are psychological commitments to investment in one approach or the other. And the human-knowledge approach tends to complicate methods in ways that make them less suited to taking advantage of general methods leveraging computation. There were many examples of AI researchers' belated learning of this bitter lesson, and it is instructive to review some of the most prominent.



- "*computation*" is mentioned **21** times
- "*data*" is mentioned **once**, in passing

Data / Compute trade-off?

Data













Compute

Data / Compute trade-off?



Simulation

Data

What if we want to simulate <u>our</u> world?

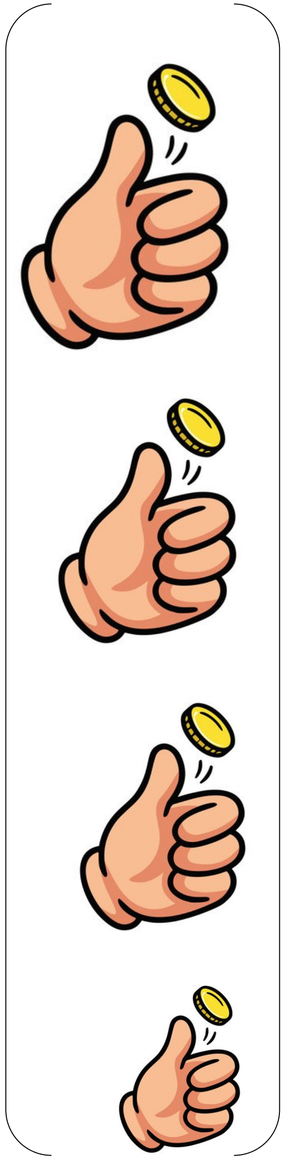










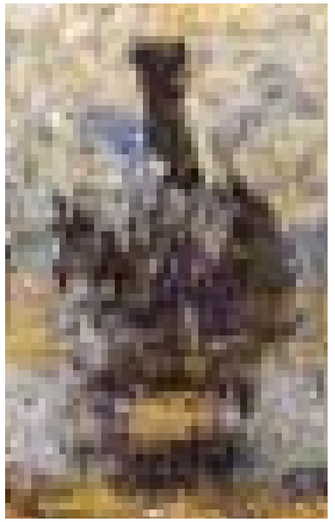
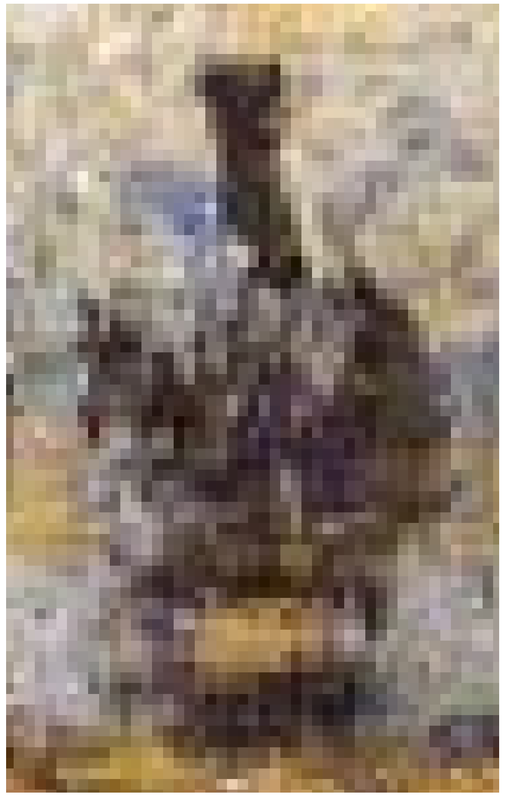


(much of) The Magic is in the Data

Importance of Past Experience



Claude Monet
Gare St.Lazare
Paris, 1877

There is almost nothing inside!

Seeing more than meets the eye



Video by Antonio Torralba (starring Rob Fergus)

But actually…



Video by Antonio Torralba (starring Rob Fergus)



“Our perception relies on memory as much as it does on incoming information, which blurs the border between perception and cognition.”

-- Moshe Bar



“Our perception relies on memory as much as it does on incoming information, which blurs the border between perception and cognition.”

-- Moshe Bar

“Mind" is largely an emergent property of "data."

-- Lance Williams

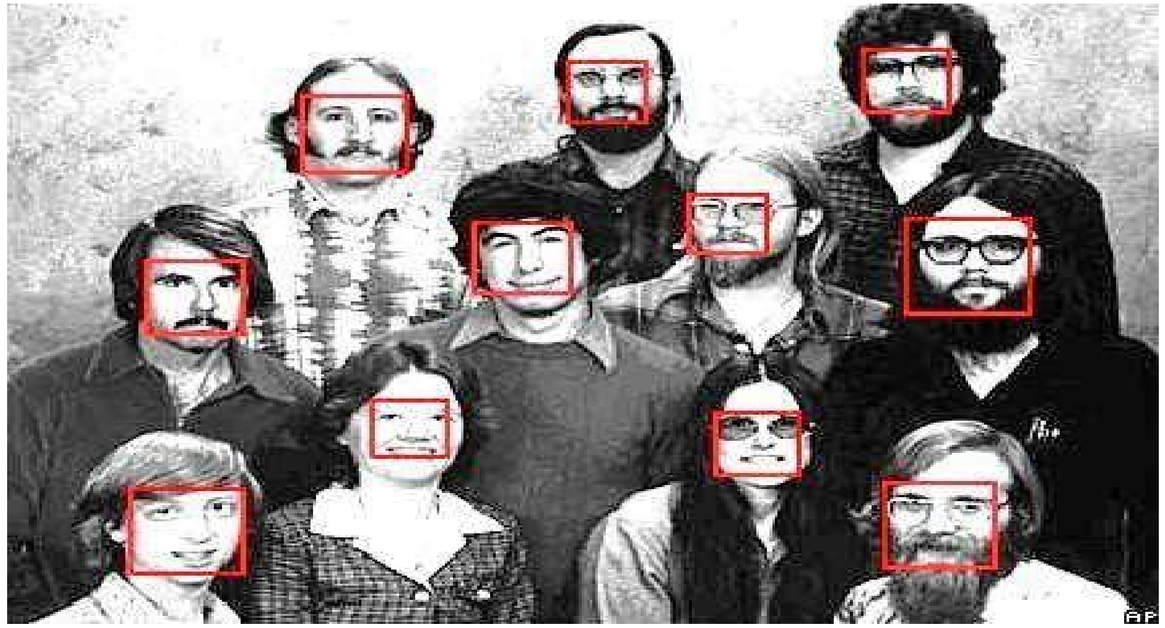
Still, data itself got no respect

Data

Features

Learning Algorithm

Face Detection: Early Success Story





- Rowley, Baluja, and Kanade, 1998
  - features: **pixels**, classifier**: neural network**
- Schniderman & Kanade, 1999
  - features: **pairs of wavelet coeff**., classifier**: naïve Bayes**
- Viola & Jones, 2001
  - features: **haar**, classifier: **boosted cascade**

Our Scientific Narcissism

All things being equal, we prefer to credit our own cleverness

“Unreasonable Effectiveness of Data”

[Halevy, Norvig, Pereira 2009]

- Parts of our world can be explained by elegant mathematics:
  - physics, chemistry, astronomy, etc.
- But some cannot:
  - psychology, genetics, economics, AI, etc.

Decade before “The Bitter Lesson”

- Enter: The **Magic of Data**

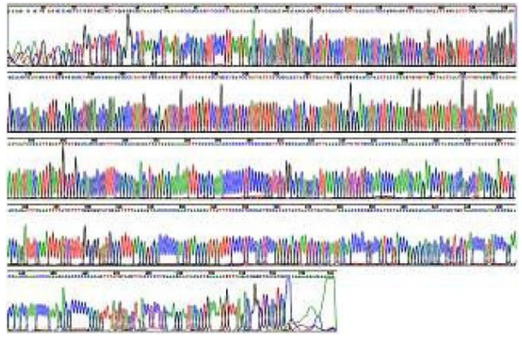
Some things are just messy





Navier-Stokes Equation

$$\frac{\partial \mathbf{u}}{\partial t} = -\left(\mathbf{u}\cdot\nabla\right)\mathbf{u} + v\nabla^2\mathbf{u} - \frac{1}{d}\nabla p + \mathbf{f}$$



+ weather
+ location
+ …

Let’s Define a Tree



Brown trunk moving upward and branching with leaves

Are these trees?

Defining is Hard; Recognizing is Easy





- Hard to give a good definition of tree
- But any 3-year-old can tell a tree from a non-tree
- The 3-year-old has **learned from data!**

Slide from Makdon Ismail

The Economist

Obama the warrior

Misgoverning Argentina

The economic shift from West to East

Genetically modified crops blossom

The right to eat cats and dogs

The data deluge

AND HOW TO HANDLE IT: A 14-PAGE SPECIAL REPORT



With enough data, brain-dead lookup (aka Nearest Neighbor) works surprisingly well

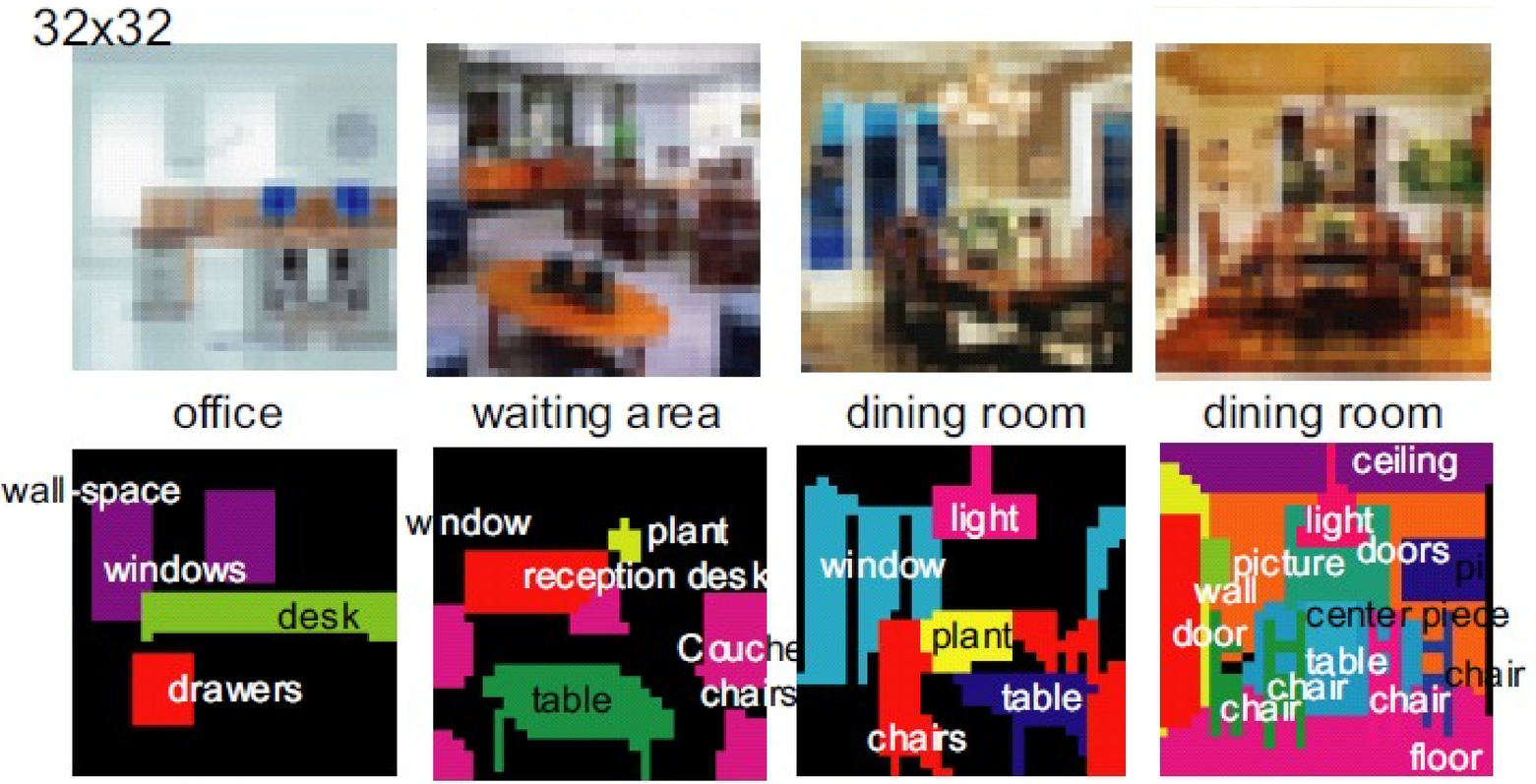
How many images does a 5-year-old see?



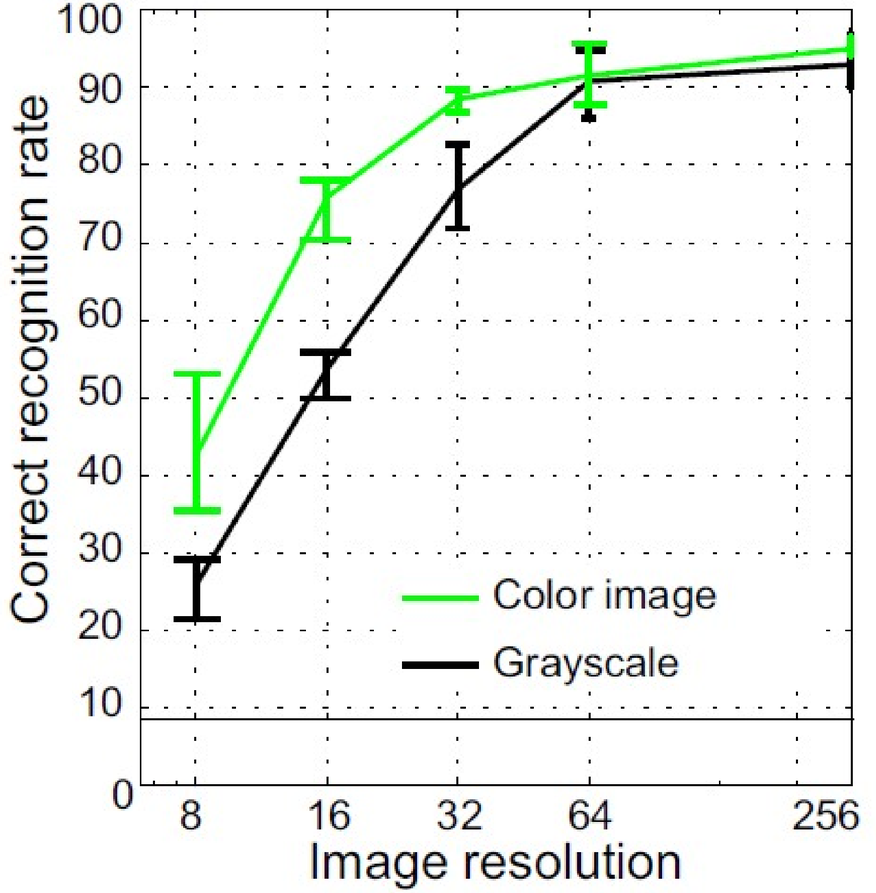
Where CIFAR-10/100 came from!

- [“80 million tiny images: a large dataset for non-parametric object and scene recognition”, Antonio Torralba, Rob Fergus and William T. Freeman. PAMI 2008.]

32x32 turns out to be enough!



c) Segmentation of 32x32 images

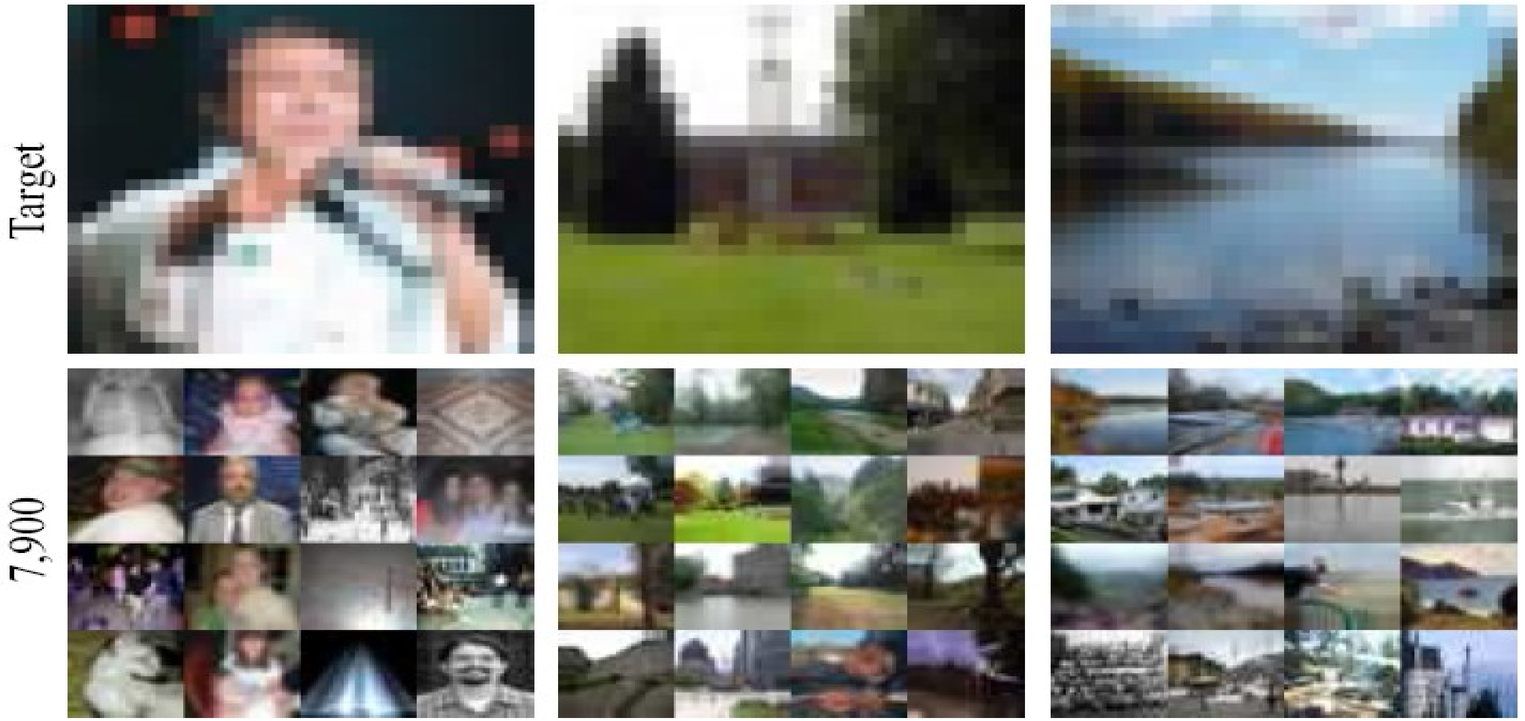


a) Scene recognition

Lots

Of

Images

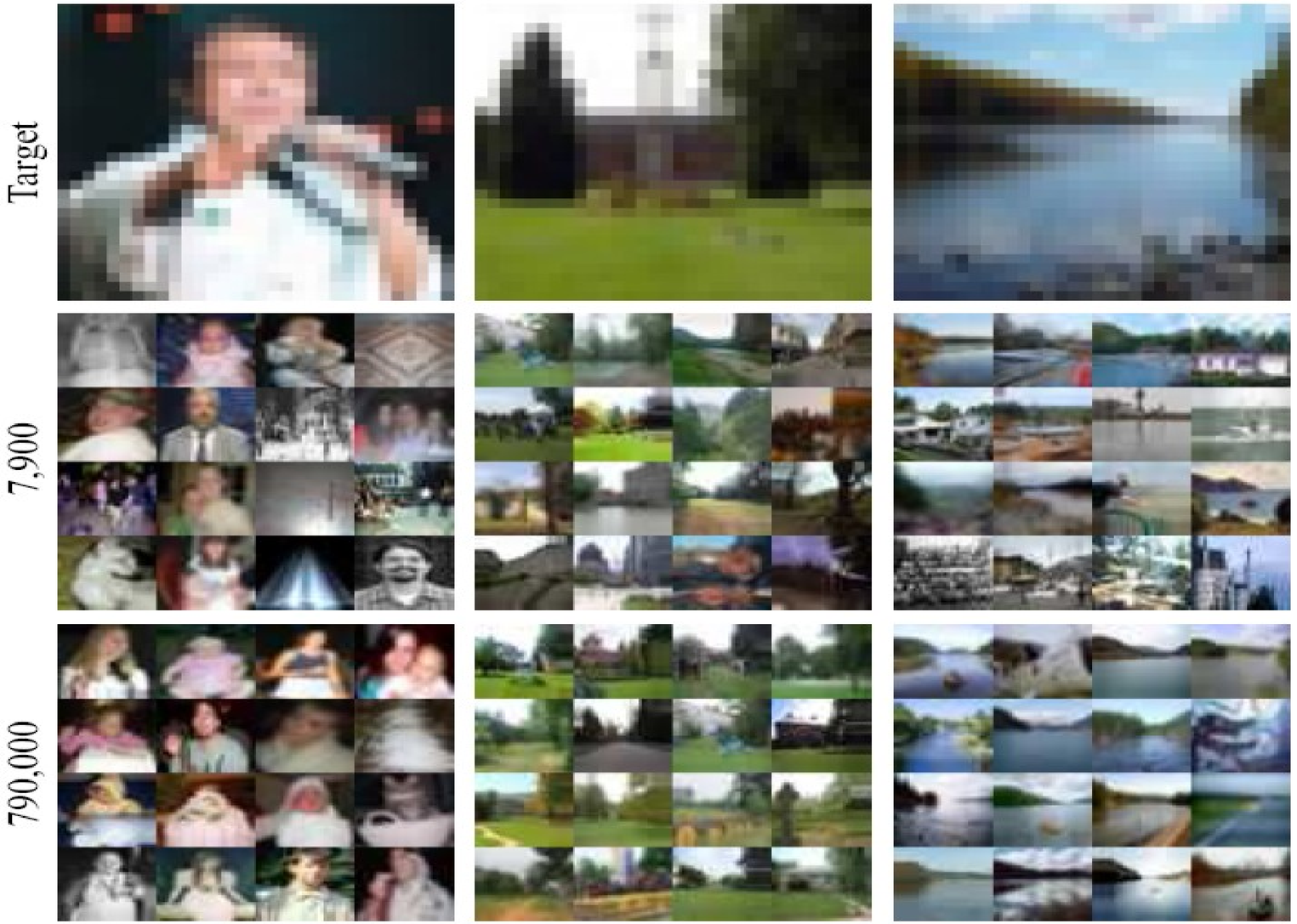


A. Torralba, R. Fergus, W.T.Freeman. PAMI 2008

Lots

Of

Images

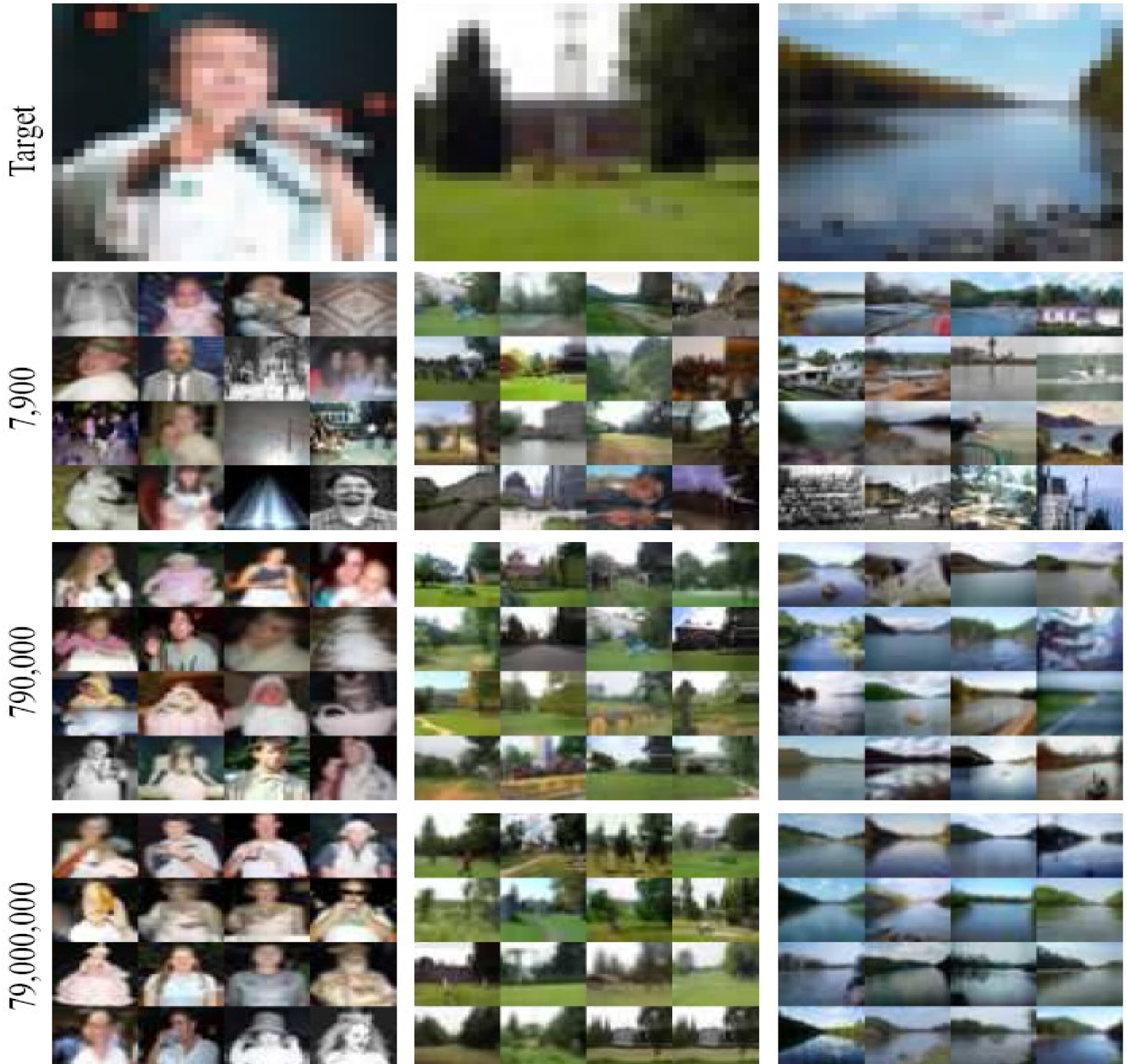


A. Torralba, R. Fergus, W.T.Freeman. PAMI 2008

Lots

Of

Images

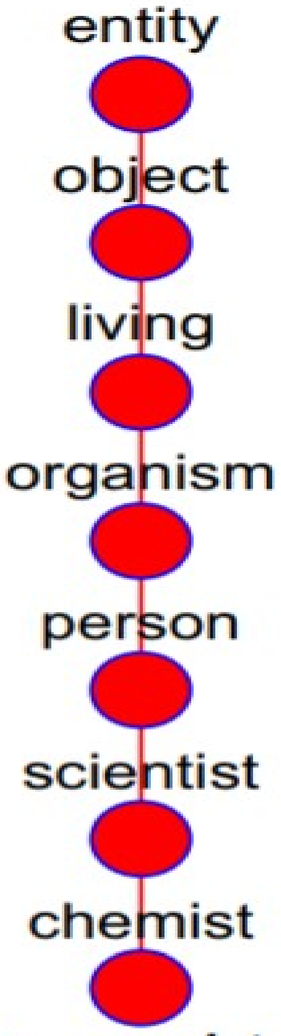




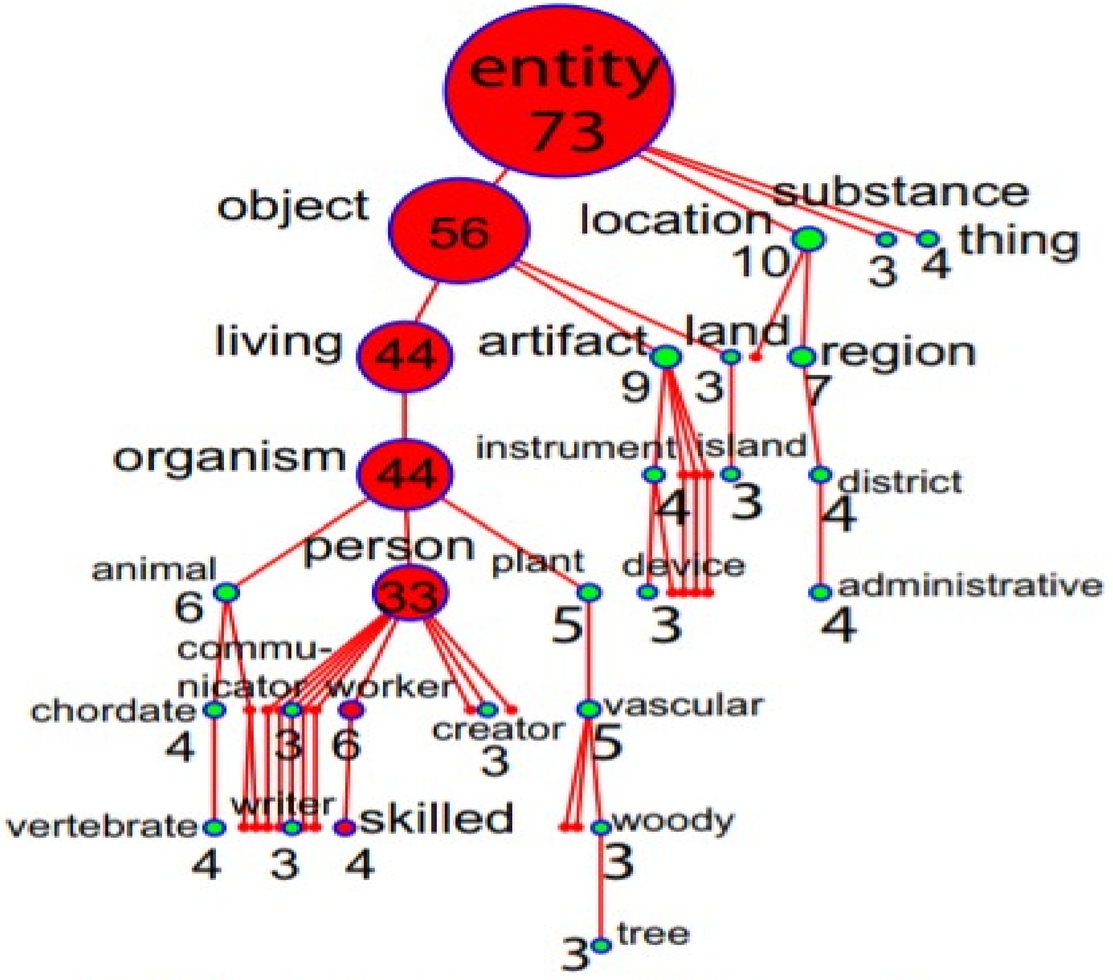
a) Input image



b) Neighbors



c) Ground truth



d) Wordnet voted branches

Automatic Colorization

Grayscale input High resolution



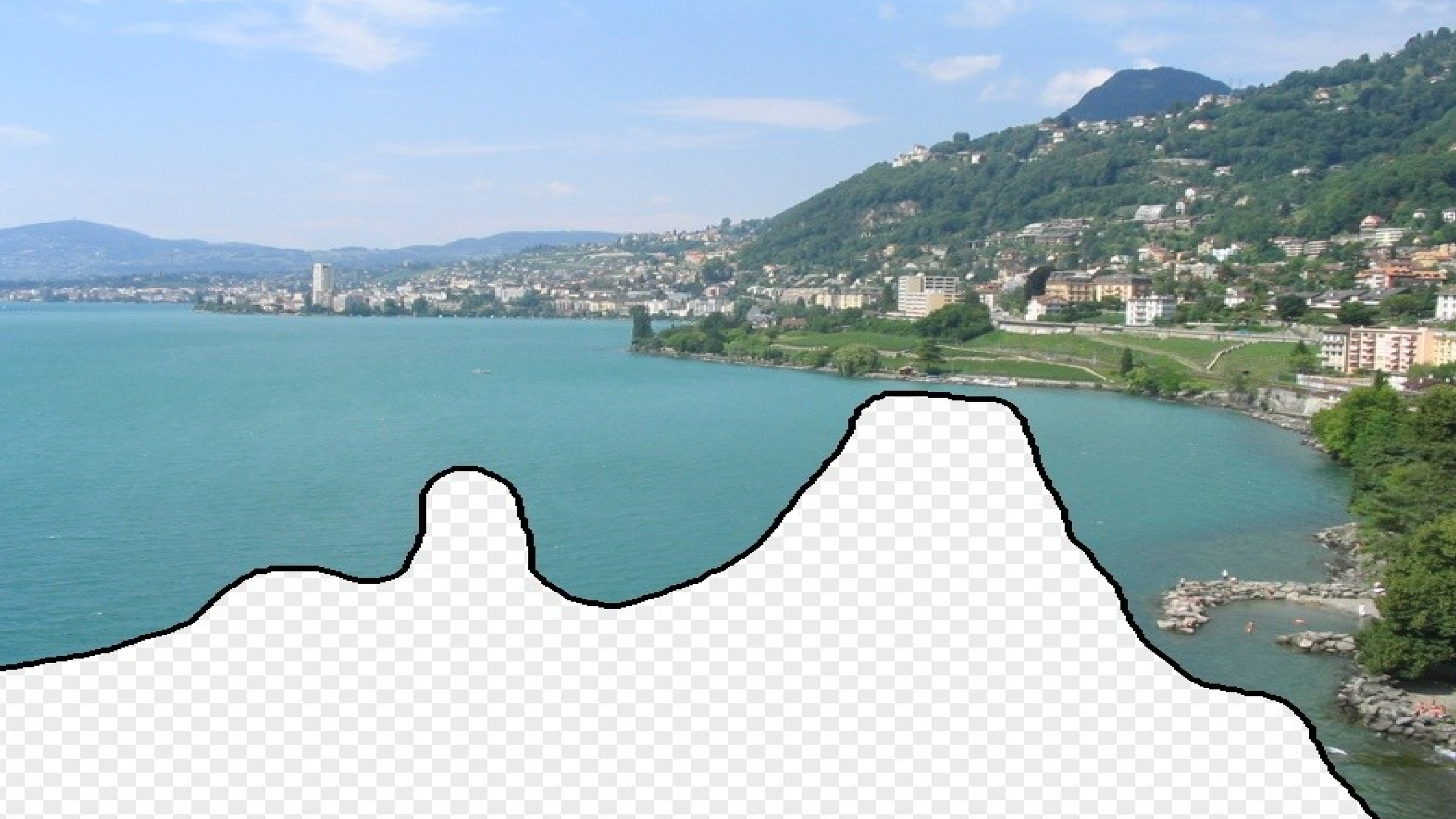
Colorization of input using average



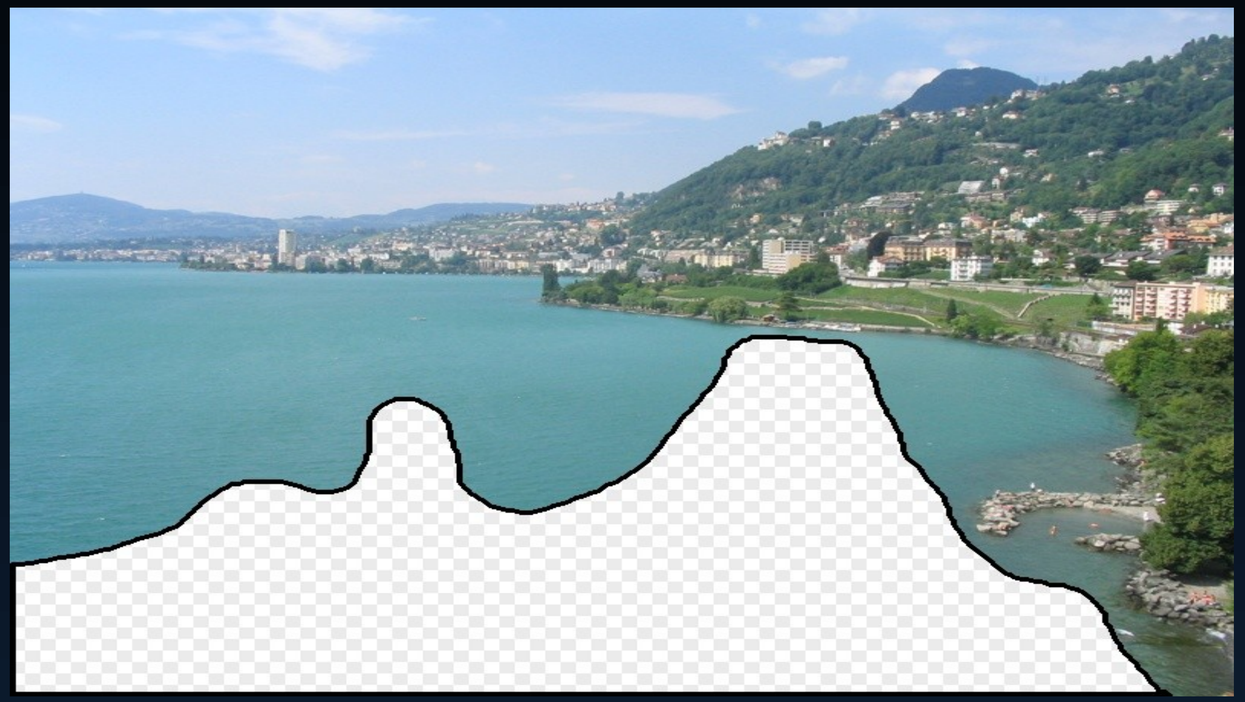
A. Torralba, R. Fergus, W.T.Freeman. 2008

[Hays & Efros, SIGGRAPH’07]






2 Million Flickr Images





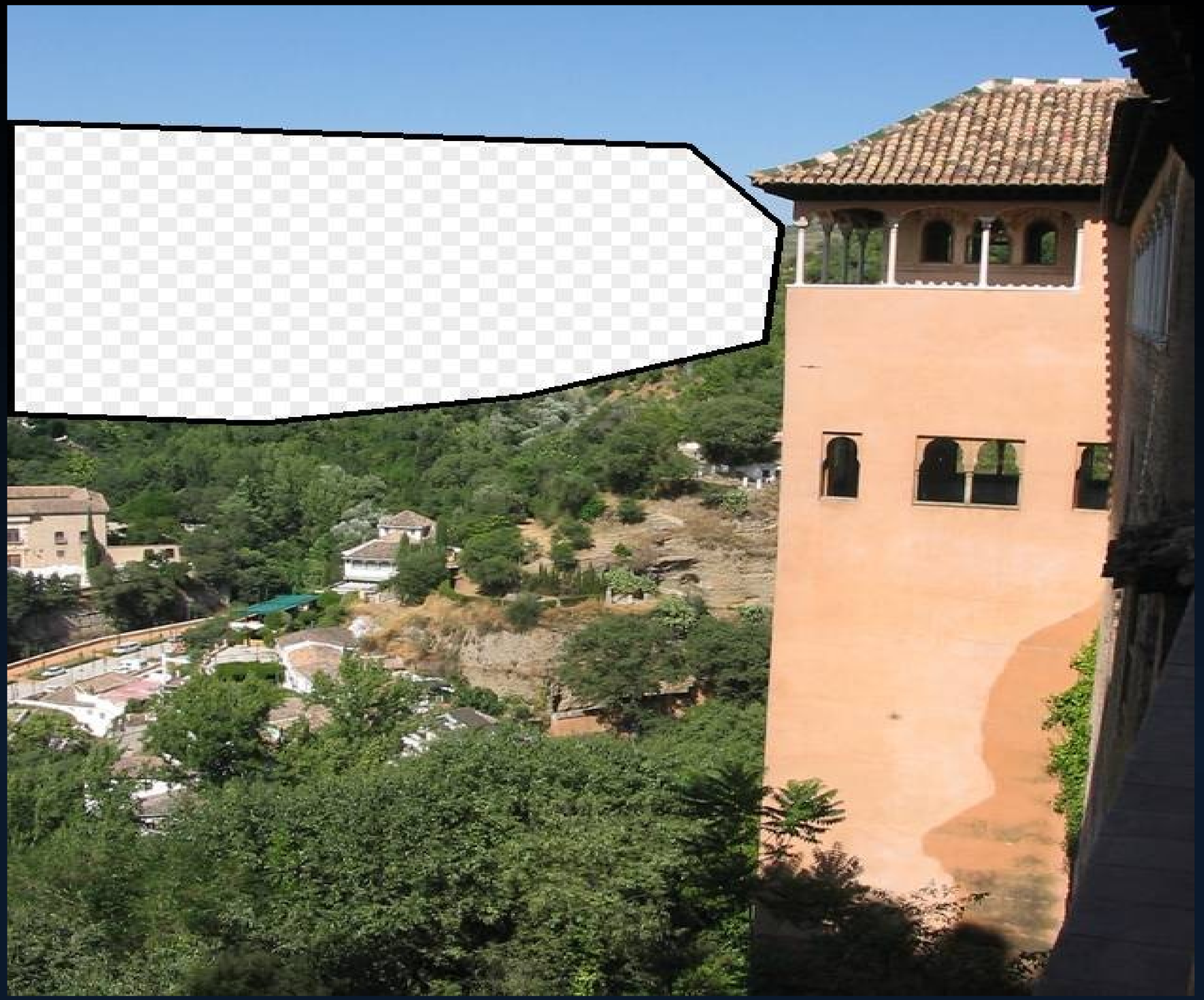








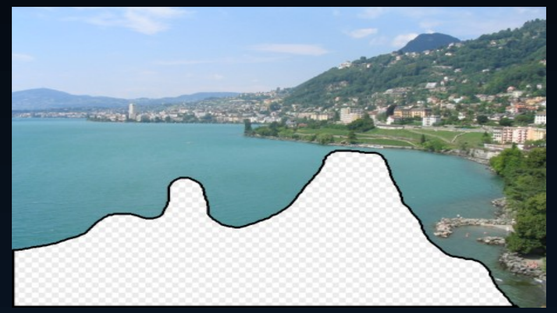






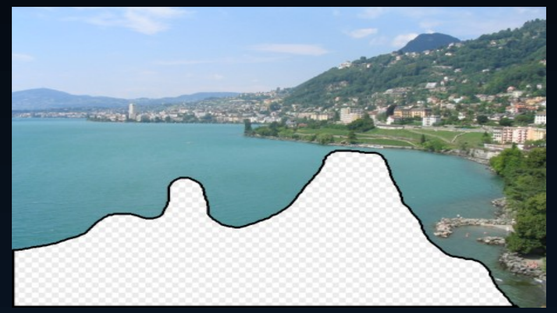


Why does it work?





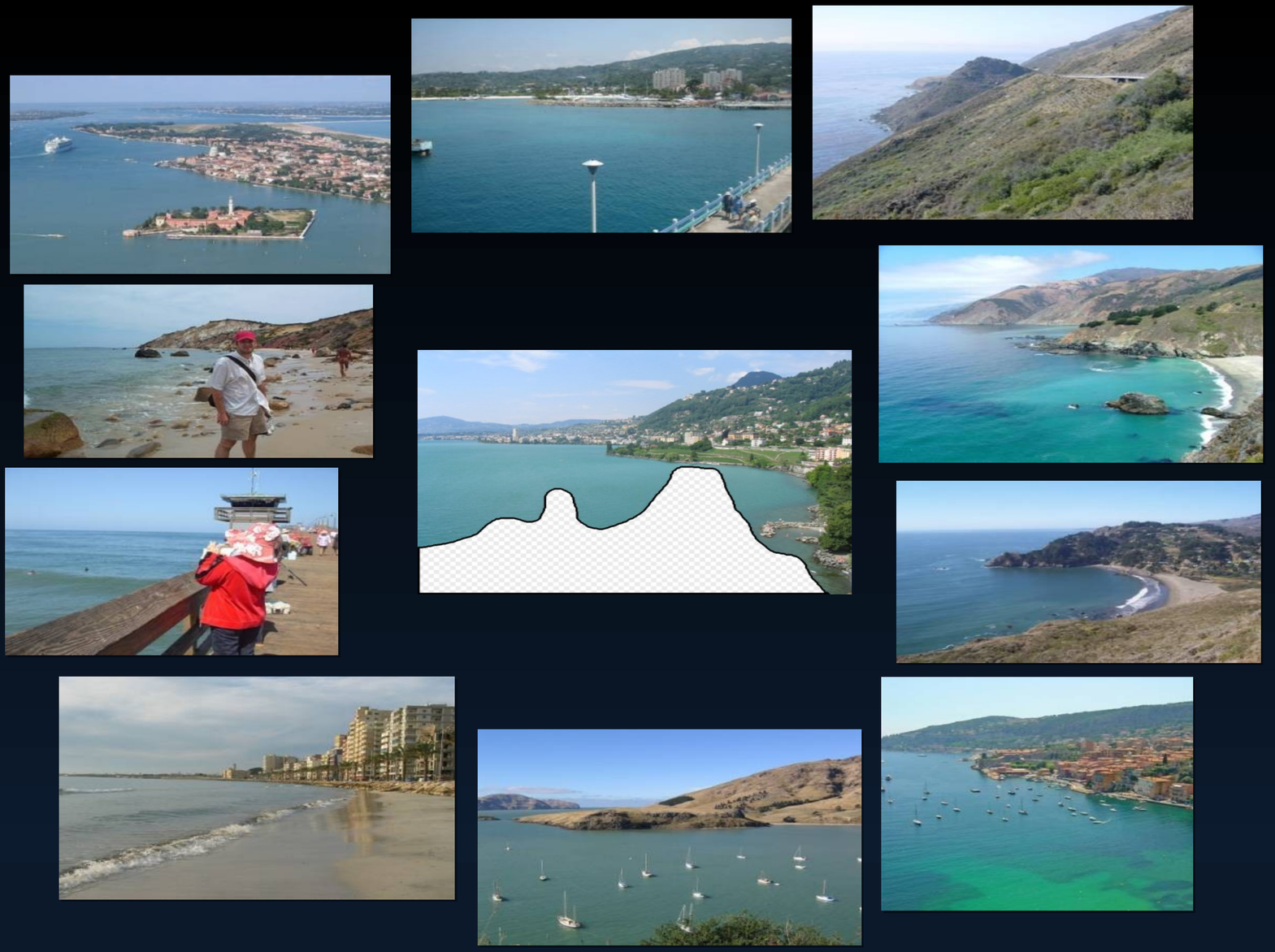







Nearest neighbors from a
collection of **20 thousand images**



Nearest neighbors from a
collection of **2 million images**

im2GPS
(using 6 million GPS-tagged Flickr images)



Query Photograph

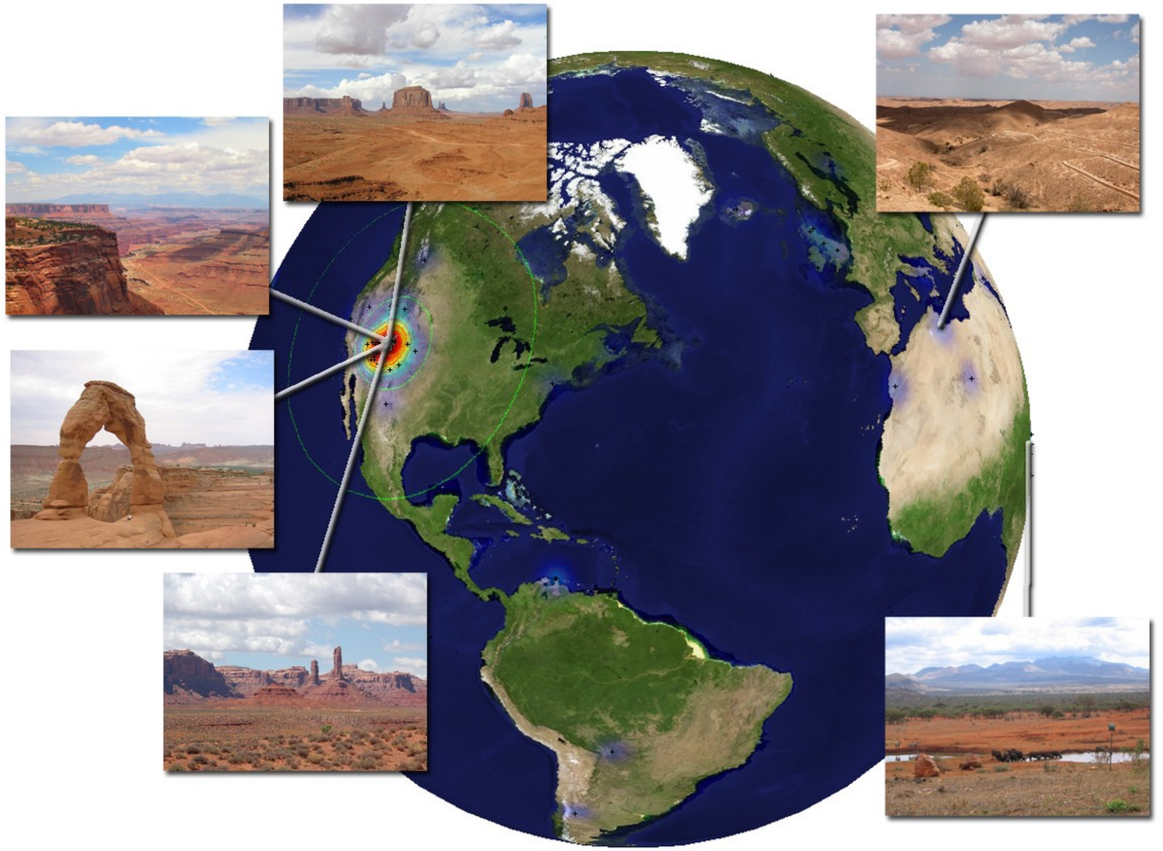
Im2gps [Hays & Efros, CVPR’08]


6 Million Flickr Images

im2GPS
(using 6 million GPS-tagged Flickr images)



Query Photograph



Visually Similar Scenes

Im2gps [Hays & Efros, CVPR’08]

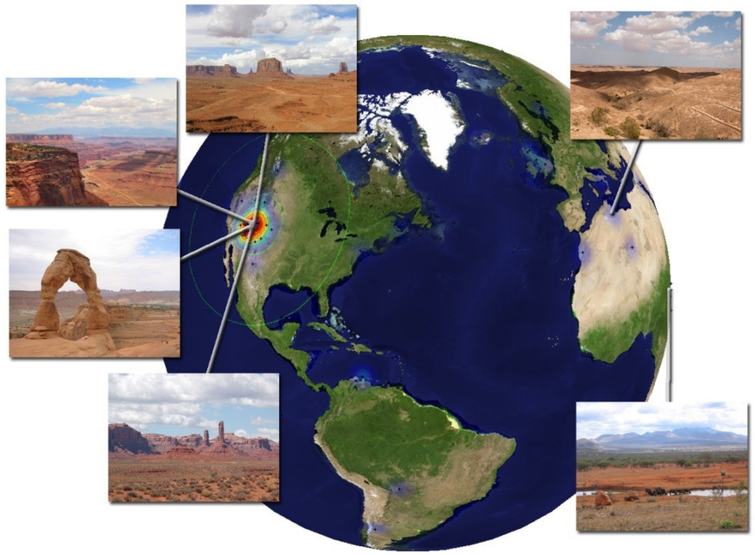
15 years later…

Algorithm vs. Data

im2gps, 2008



Query Photograph



Visually Similar Scenes

- Nearest Neighbors
- 6 million images

PlaNet, 2016



CC-BY-NC by stevekc



CC-BY-NC by edwin.11



CC-BY-NC by jonathanfh



(a) (b) (c)

- Deep Net
- 91 million images

Algorithm vs. Data

Method	Street 1 km	City 25 km	Region 200 km	Country 750 km	Continent 2500 km
Im2GPS (orig) [19]		12.0%	15.0%	23.0%	47.0%
Im2GPS (new) [20]	2.5%	21.9%	32.1%	35.4%	51.9%
PlaNet (900k)	0.4%	3.8%	7.6%	21.6%	43.5%
PlaNet (6.2M)	6.3%	18.1%	30.0%	45.6%	65.8%
PlaNet (91M)	**8.4%**	**24.5%**	**37.6%**	**53.6%**	**71.3%**

Weyand et al, PlaNet, ECCV 2016

But can humans ever remember so much?

What's the Capacity of Visual Long Term Memory?

PNAS

Visual long-term memory has a massive storage capacity for object details

Timothy F. Brady*, Talia Konkle, George A. Alvarez, and Aude Oliva*

Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139

Edited by Dale Purves, Duke University Medical Center, Durham, NC, and approved August 1, 2008 (received for review April 8, 2008)

One of the major lessons of memory research has been that human memory is fallible, imprecise, and subject to interference. Thus, although observers can remember thousands of images, it is widely assumed that these memories lack detail. Contrary to this

studied images, making it impossible to conclude whether the memories for each item in these previous experiments consisted of only the "gist" or category of the image, or whether they contained specific details about the images. Therefore, it re-



Aude Oliva, MIT

http://olivalab.mit.edu/MM/

What's the Capacity of Visual Long Term Memory?

What was known...

Standing (1973)

10,000 images

83% Recognition

... people can remember thousands of images

What was not known...

... what people are remembering for each item?



According to Standing

"Basically, my recollection is that we just separated the pictures into distinct thematic categories: e.g. cars, animals, single-person, 2-people, plants, etc.) Only a few slides were selected which fell into each category, and they were visually distinct."

Dogs
Playing Cards

"Gist" Only



Sparse Details



Highly Detailed

Slide by Aude Oliva

Massive Memory Experiment I

A stream of objects will be presented on the screen for
~ 3 second each.

Your primary task:

Remember them ALL!

afterwards you will be tested with…

Completely different objects…





Different exemplars of the same kind of object…





Different states of the same object…

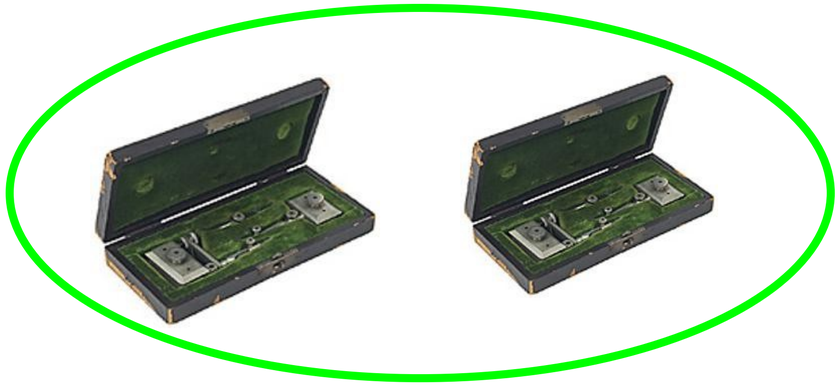


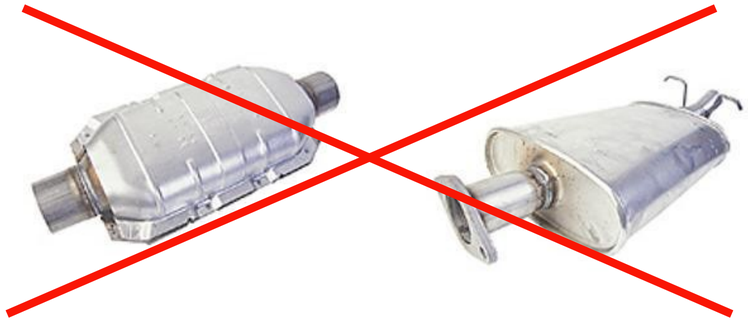


Massive Memory Experiment I

Your other task:

Detect exact repeats
anywhere in the stream





Ready?

(Seriously, get ready to clap. The images go by fast…)





<clap!>













<clap!>







10 Minutes Later...









<clap!>





<clap!>





30 Minutes Later...





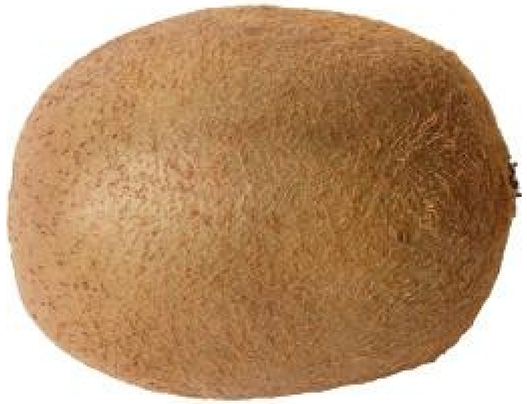








1 Hour Later...






POLICE





<clap!>





2 Hours Later...





<clap!>



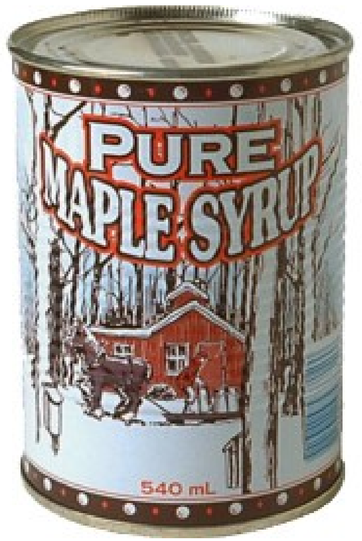








4 Hours Later...


PURE
MAPLE SYRUP
540 mL




STANDARD





<clap!>





5:30 Hours Later...



Which one did you see?

(go ahead and shout out your answer)



-A-



-B-



-A-



-B-

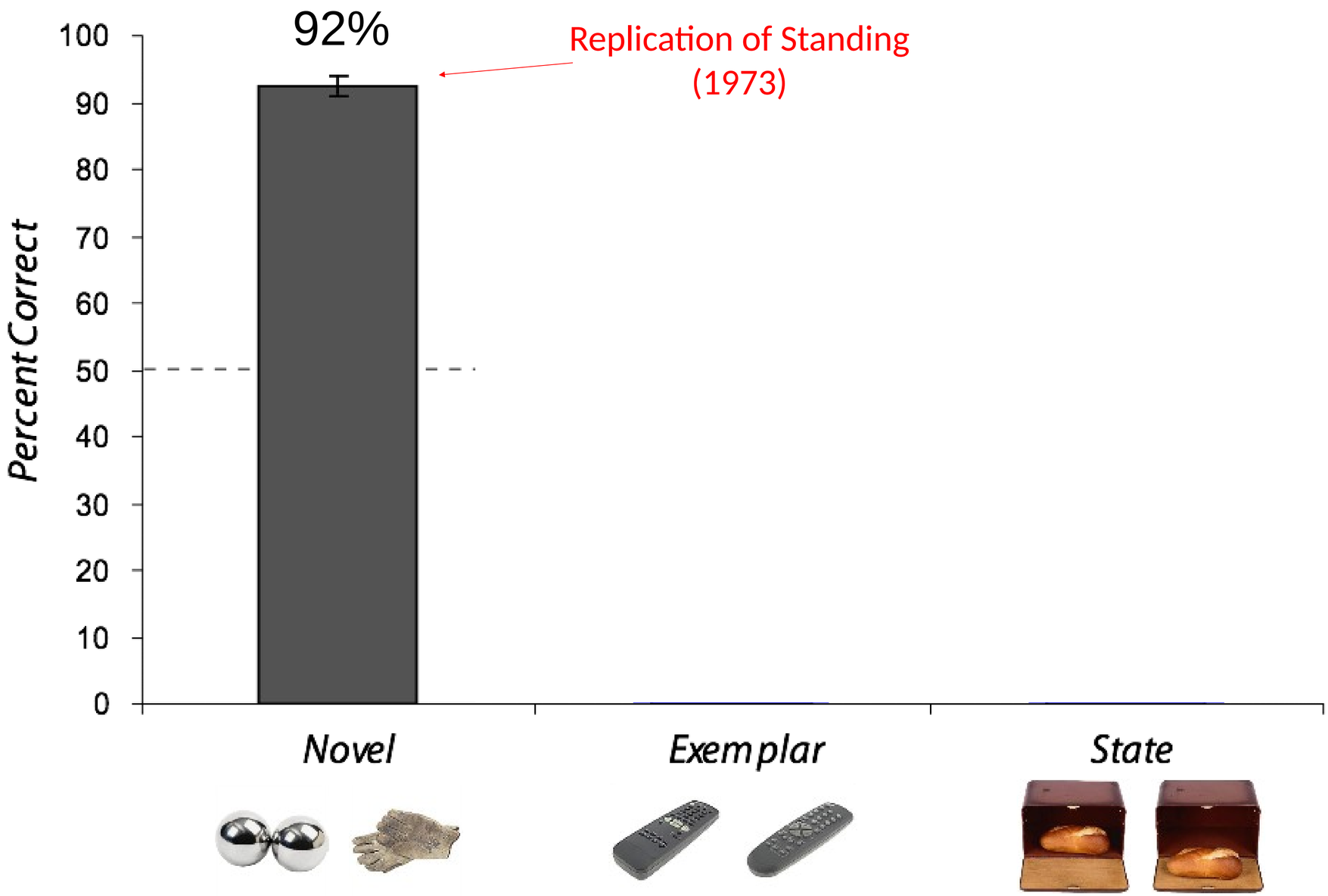


-A-

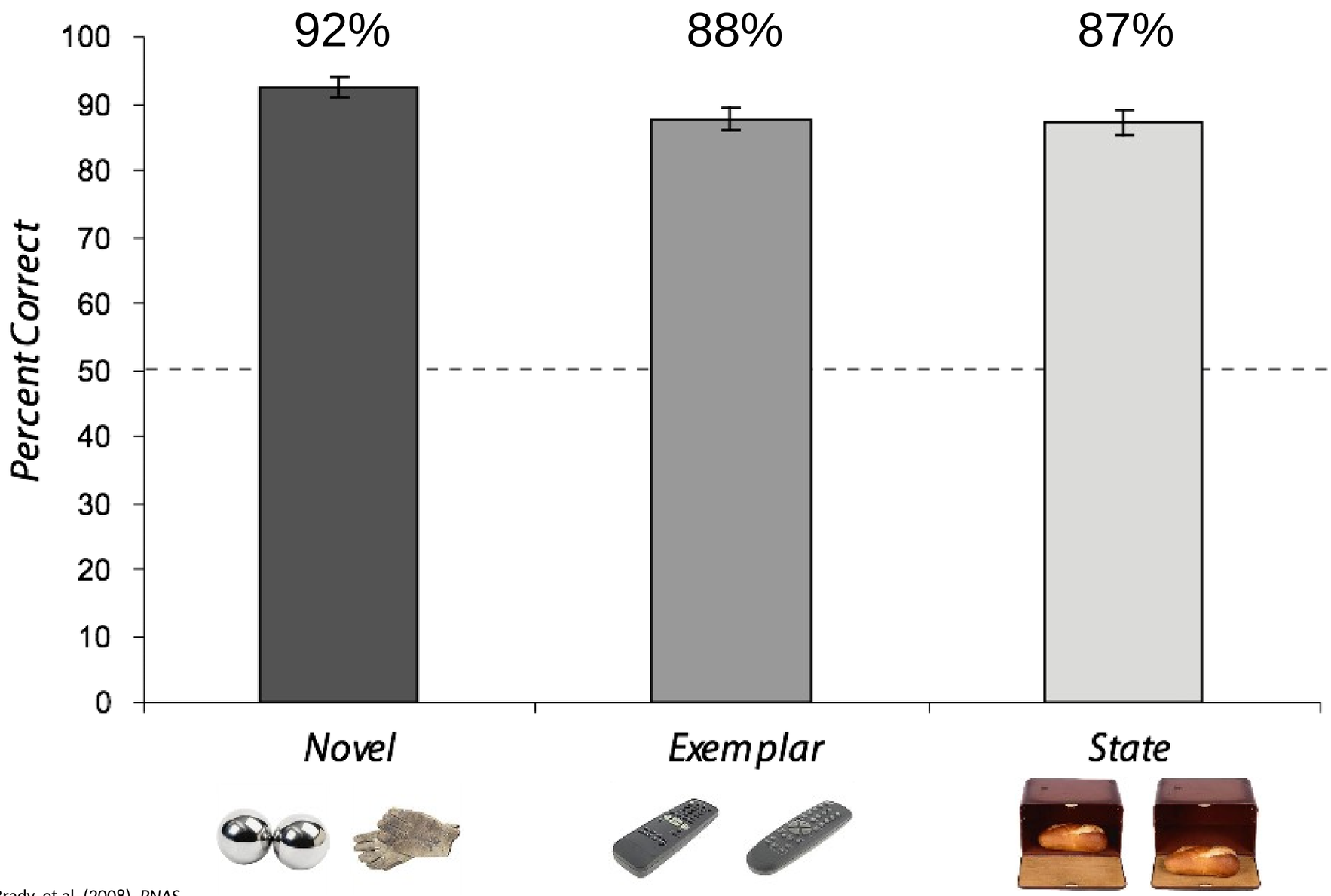


-B-

Recognition Memory Results

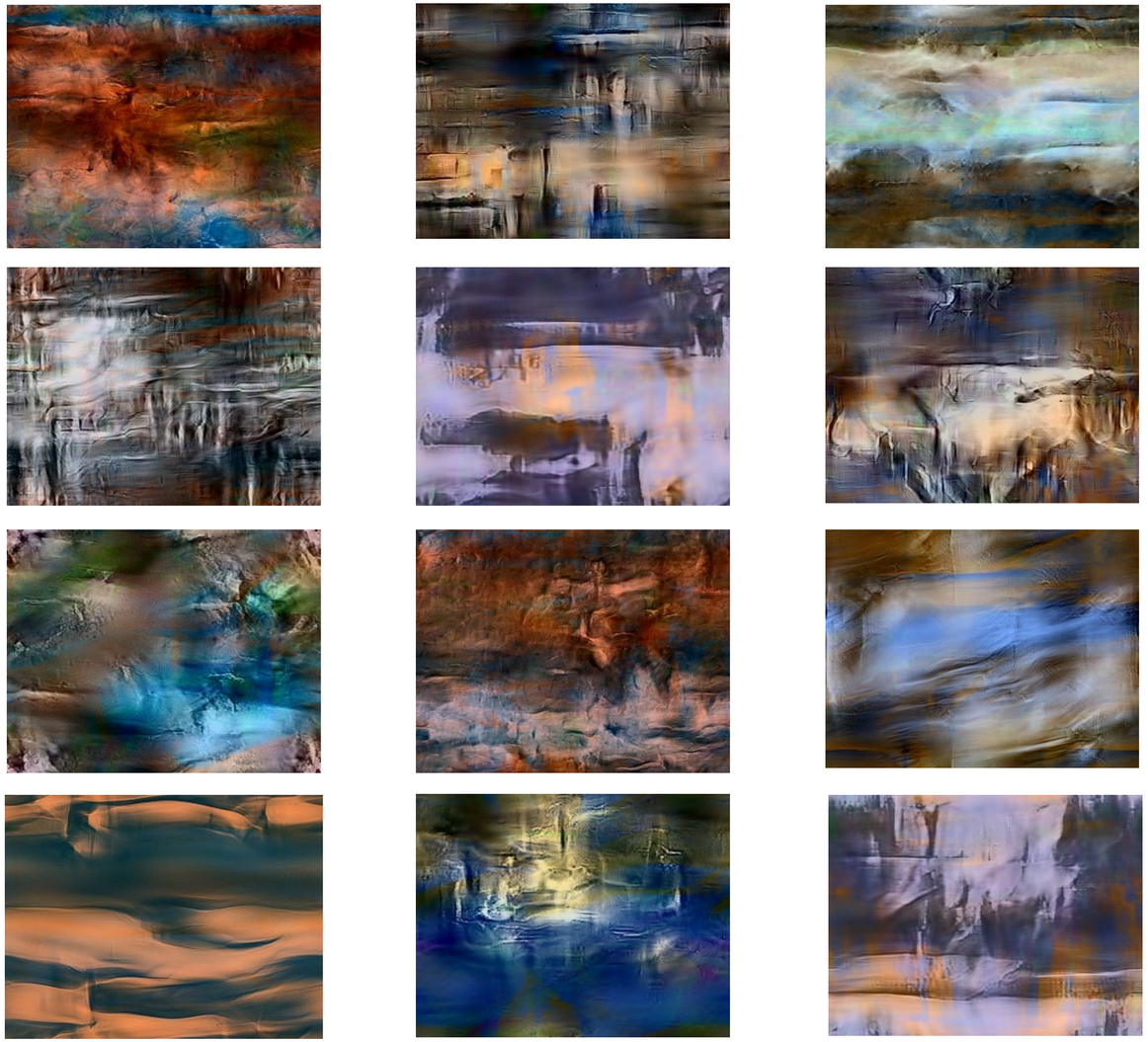


Recognition Memory Results



Brady, et al. (2008), *PNAS*

So, do humans have “photographic memory”?



“No meaning” textures matching 1st & 2nd order statistics of real scenes (by Portila & Simmonceli)

Isola & Oliva, unpublished

Clap your hands when
you see an image repeat

Ready?







<clap!>













<clap!>




BLUE SQUARE
BET NOW AT BLUESQUARE.COM
VODAFONE



<clap!>


PUMA
SPORTSTAFF




PUMA
SPORTSTAFF

<clap!>





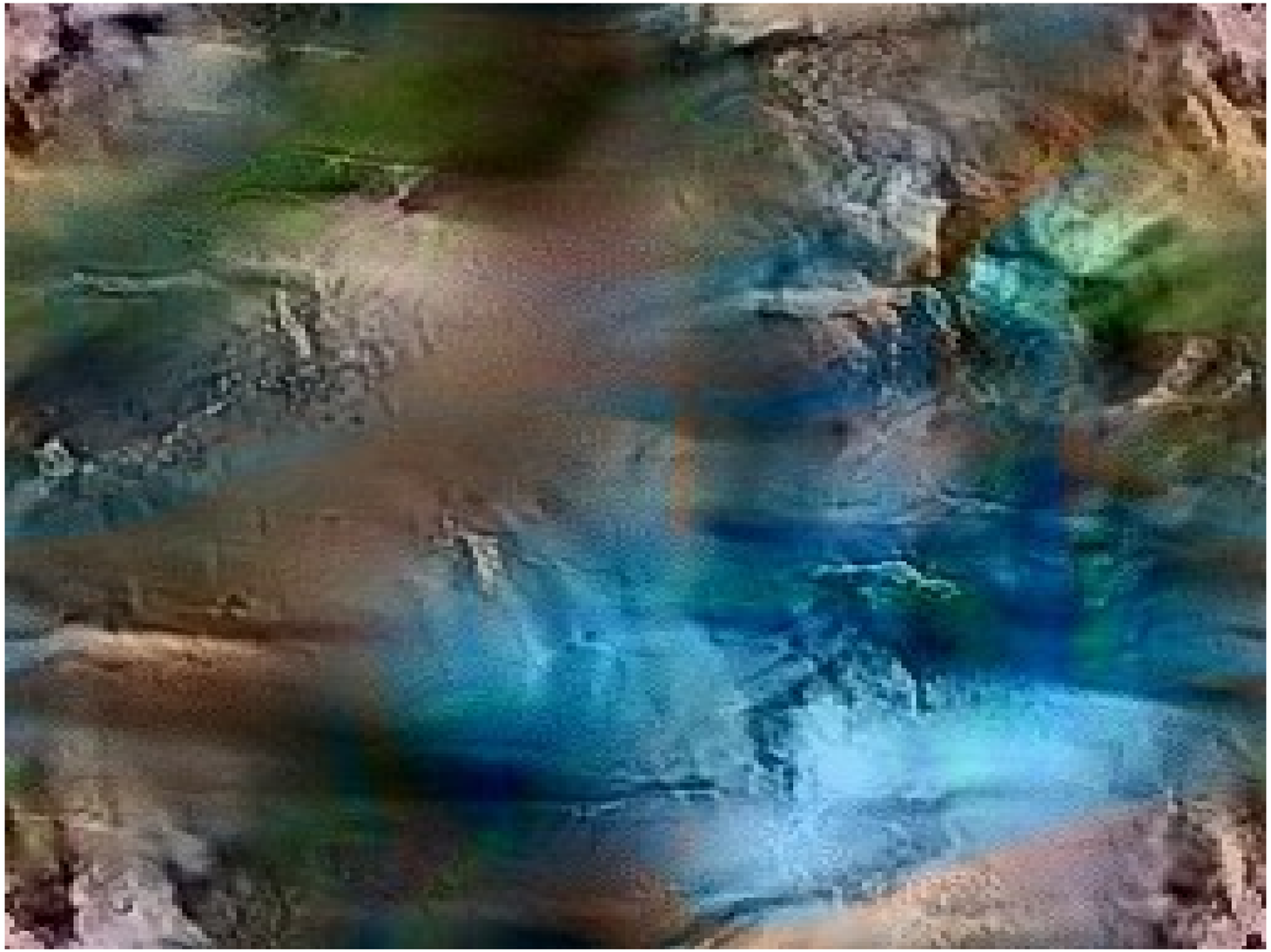






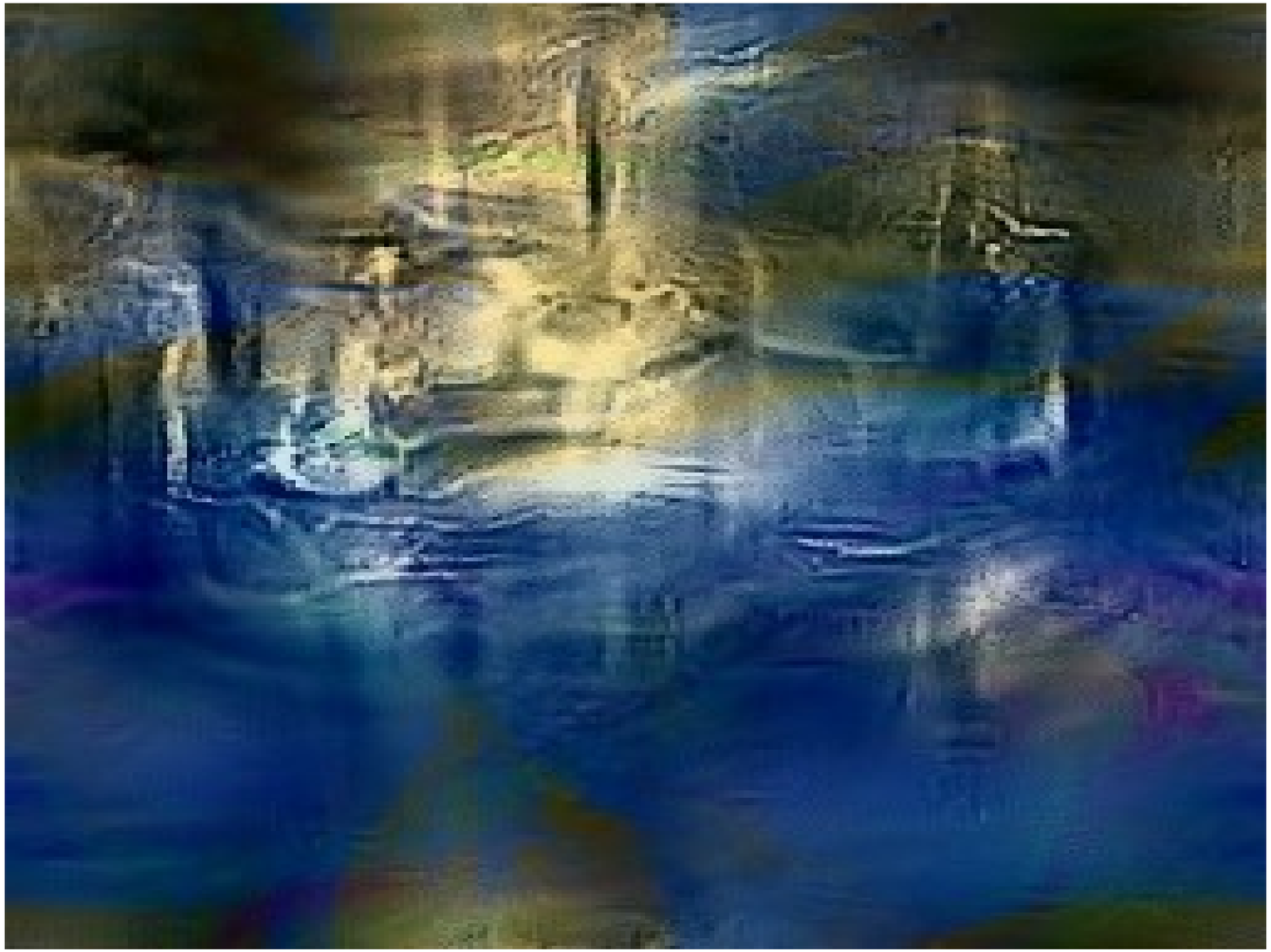
<clap!>







<clap!>

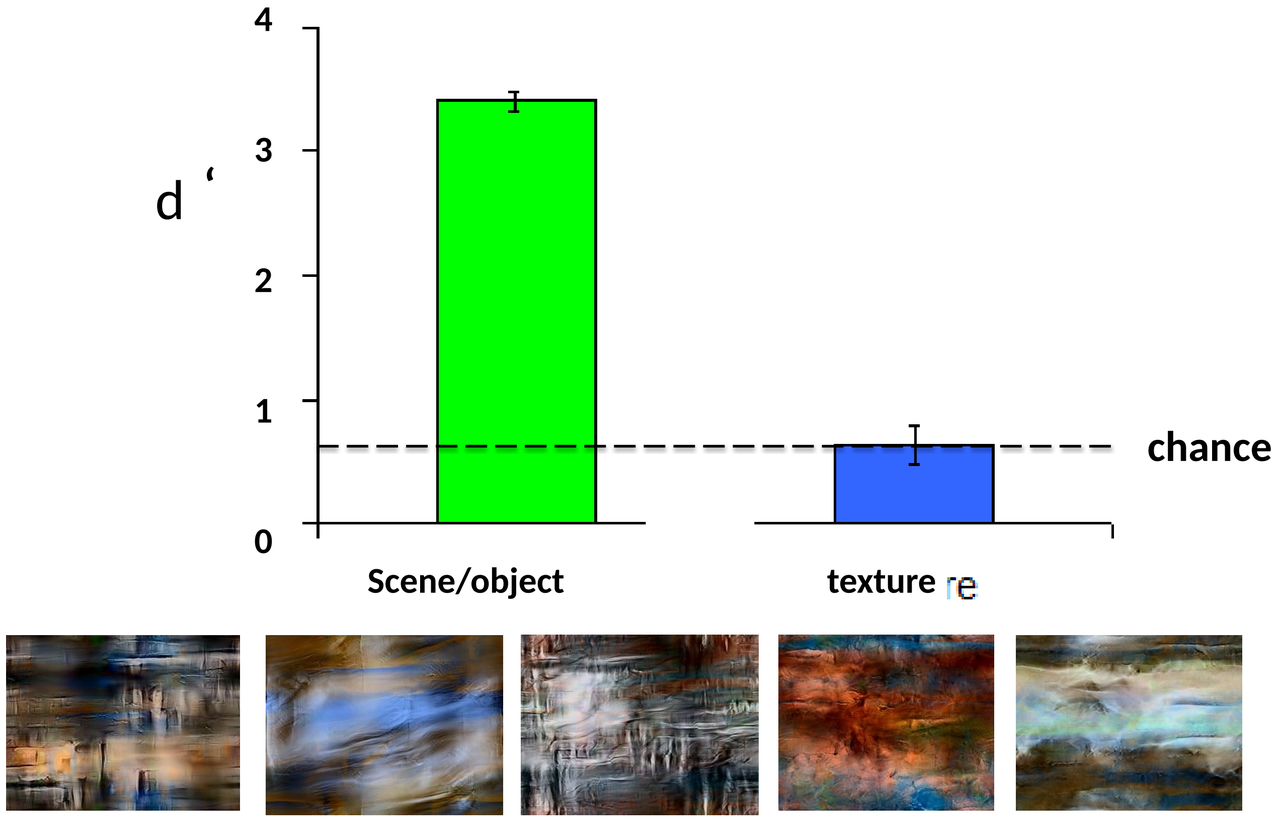



BLUE SQUARE
BET NOW AT BLUESQUARE.COM
VODAFONE

<clap!>



Humans do more than just memorize



Isola & Oliva, unpublished







Memorable
Hit rate: 67/70
False alarm rate: 4/80









Average
Hit rate: 59/81
False alarm rate: 7/92







Forgettable
Hit rate: 21/68
False alarm rate: 3/82

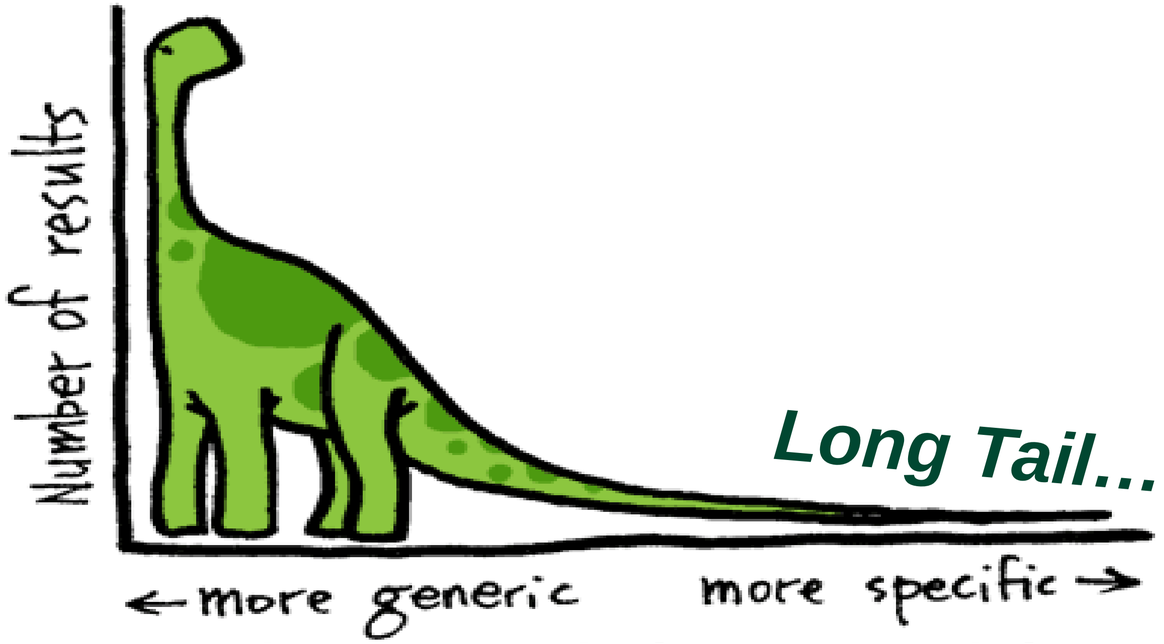
Memorability (Isola et al)

The Good News

Really stupid algorithms + Lots of Data

= “Unreasonable Effectiveness”

The Bad News: Raw Memorization (nearest-neighbor methods) have trouble with rare events:



- Modern generative models are able to **interpolate** and **compose** data from different sources:
  - e.g. DALL-E, Stable Diffusion, Imagen, Parti, Sora, etc.

Impressive compositionality:



a dolphin in an astronaut suit on saturn, artstation



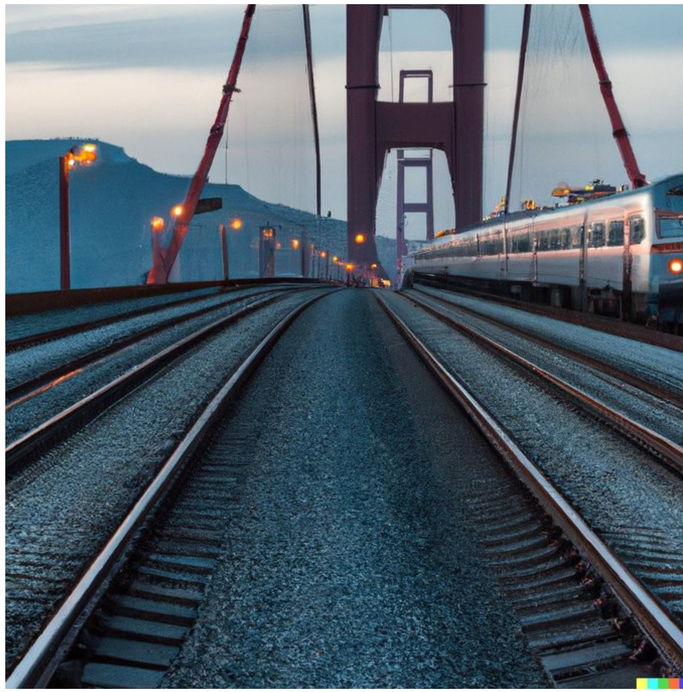
a propaganda poster depicting a cat dressed as french emperor napoleon holding a piece of cheese

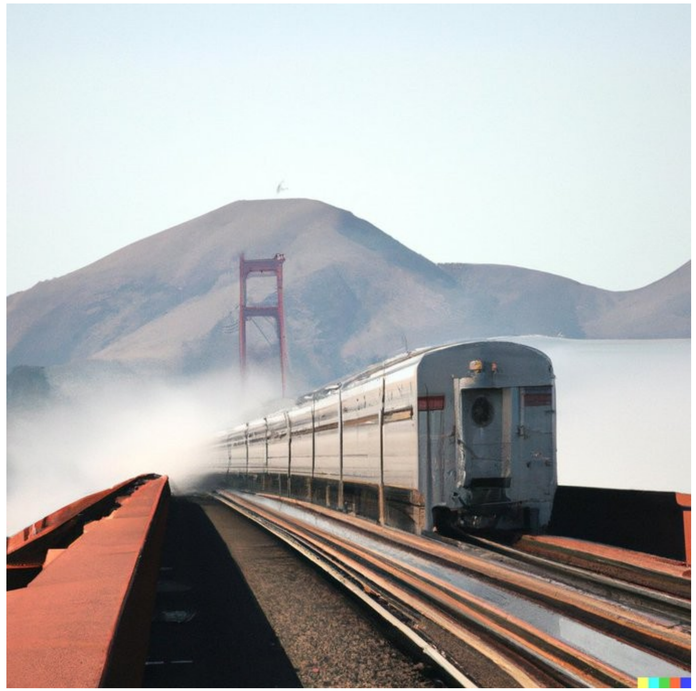
DALL-E 2

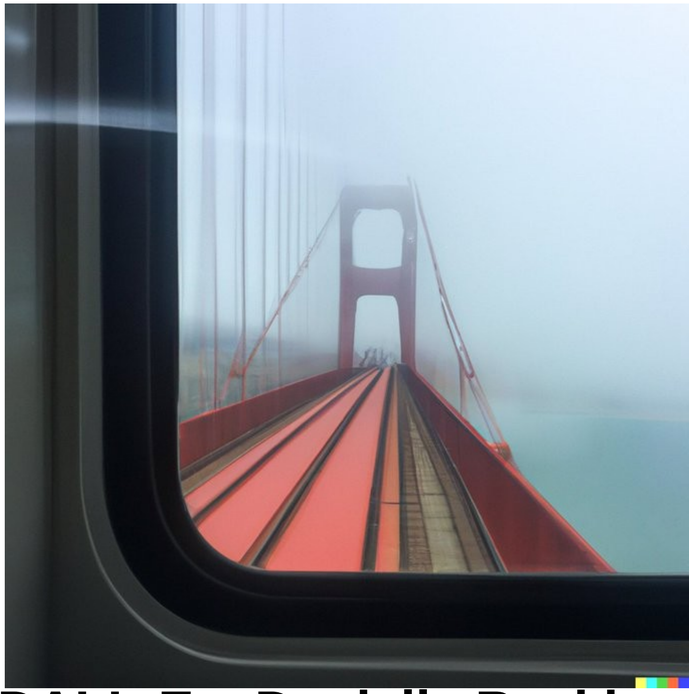
Impressive compositionality:











DALL-E + Danielle Baskin

Here we go again?

Barely Mentioned Data

Fancy Language Features

Amazing Algorithm!

Prompt: “squirrel reaching for a nut”



DALL-E 2

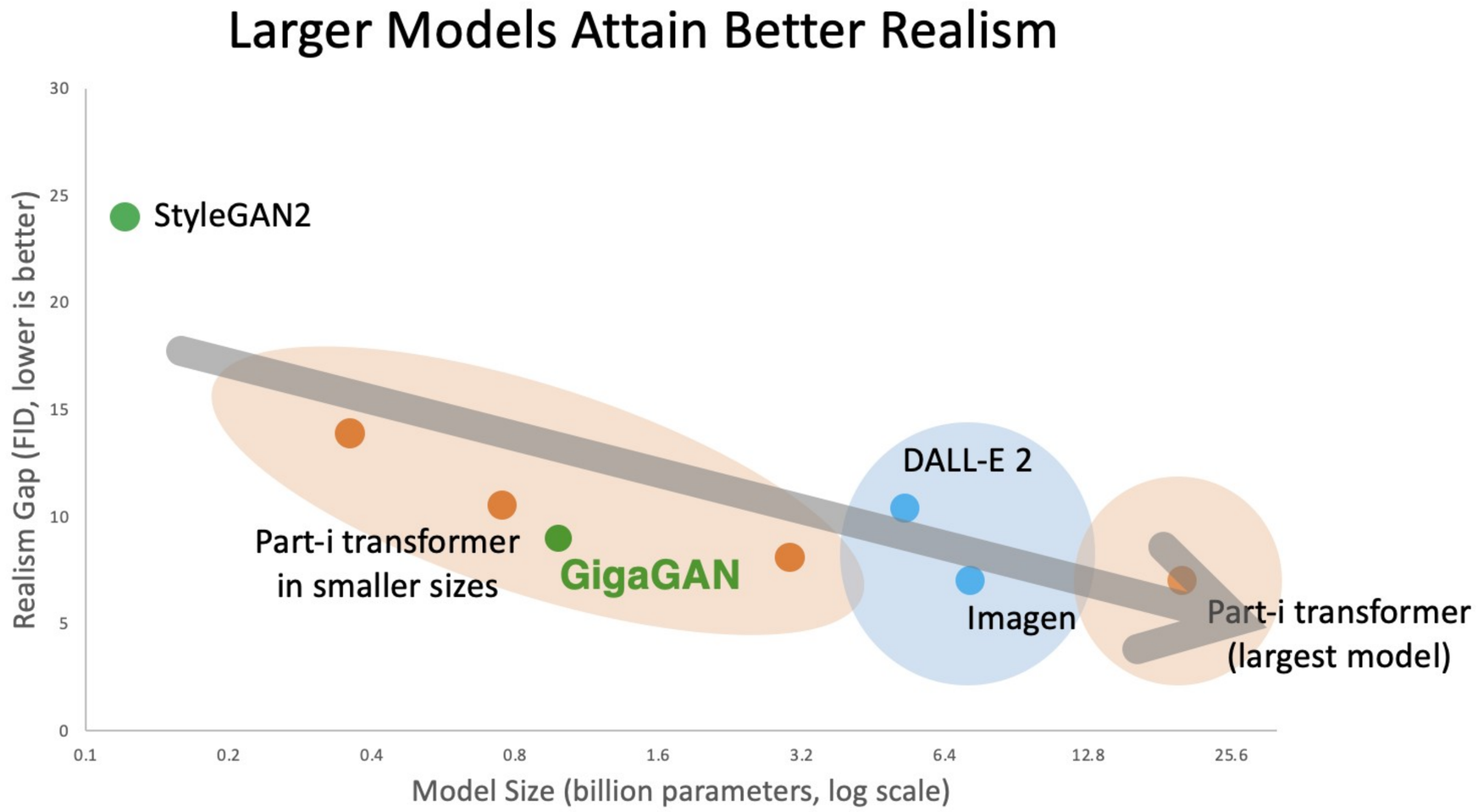


Parti



GigaGAN

data capacity vs. image quality



Slide by Taesung Park

Generated Image

Prompt:

“a green creature made of leaves and vines bursting out of the ground ready to attack; detailed, best on artstation, raymond swanland, magic the gathering, epic, stunning, masterpiece”



“influences” from the training data



generated image













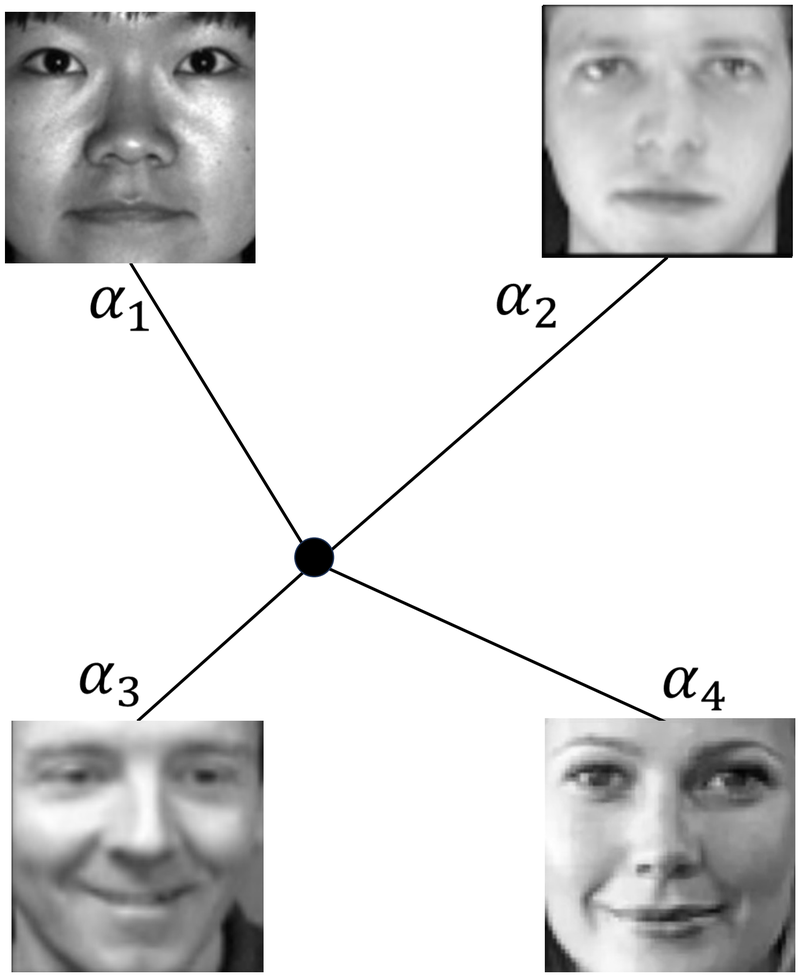
Wang et al, Evaluating Data Attribution for Text-to-Image Models, ICCV 2023

to paraphrase Arthur C. Clarke:

Interpolation in sufficiently high-dimensional space is indistinguishable from magic

But perhaps visual space isn’t even so huge?

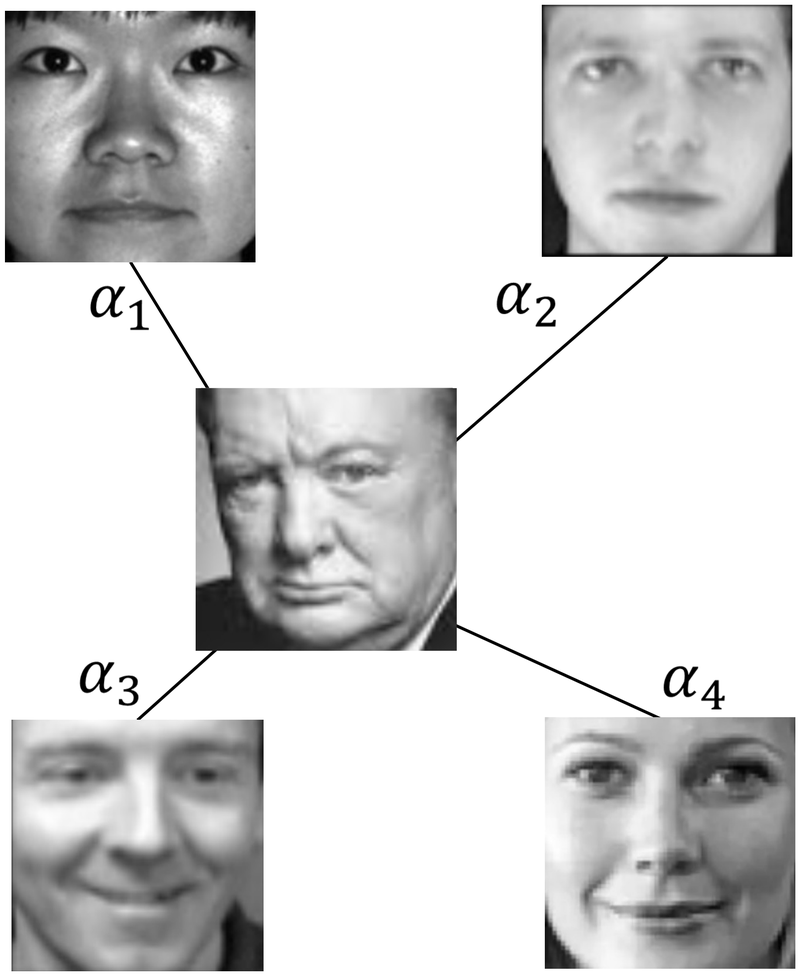
E.g. since 1990s we know that any novel face can be represented as a linear combination of a ~200 training faces



Interpreting Face Images using Active Appearance Models, Edwards, Taylor, Cootes, 1998
A Morphable Model for the Synthesis of 3D Faces, Blanz and Vetter, 1999

But perhaps visual space isn’t even so huge?

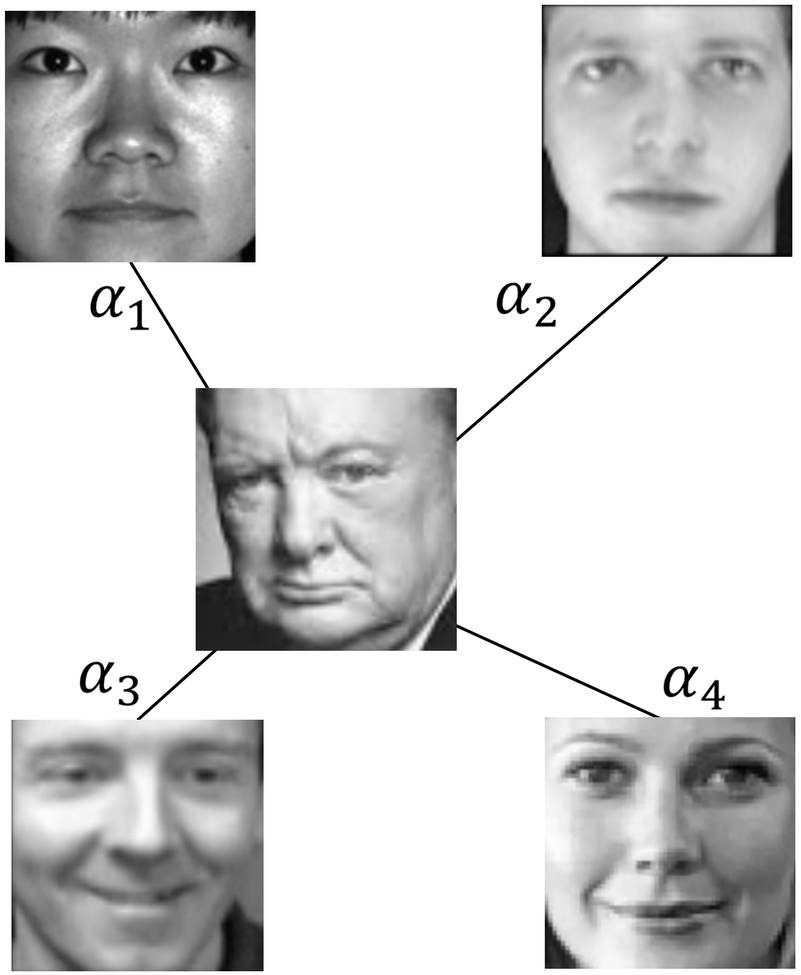
E.g. since 1990s we know that any novel face can be represented as a linear combination of a ~200 training faces



Interpreting Face Images using Active Appearance Models, Edwards, Taylor, Cootes, 1998
A Morphable Model for the Synthesis of 3D Faces, Blanz and Vetter, 1999

But perhaps visual space isn't even so huge?

E.g. since 1990s we know that any novel face can be represented as a linear combination of a ~200 training faces





There are more people on Earth than faces!

Interpreting Face Images using Active Appearance Models, Edwards, Taylor, Cootes, 1998
A Morphable Model for the Synthesis of 3D Faces, Blanz and Vetter, 1999

Interpreting the Weight Space of Customized Diffusion Models

Amil Dravid*
UC Berkeley, Snap

Yossi Gandelsman*
UC Berkeley

Kuan-Chieh Wang
Snap

Rameen Abdal
Stanford

Gordon Wetzstein
Stanford

Alyosha Efros
UC Berkeley

Kfir Aberman
Snap









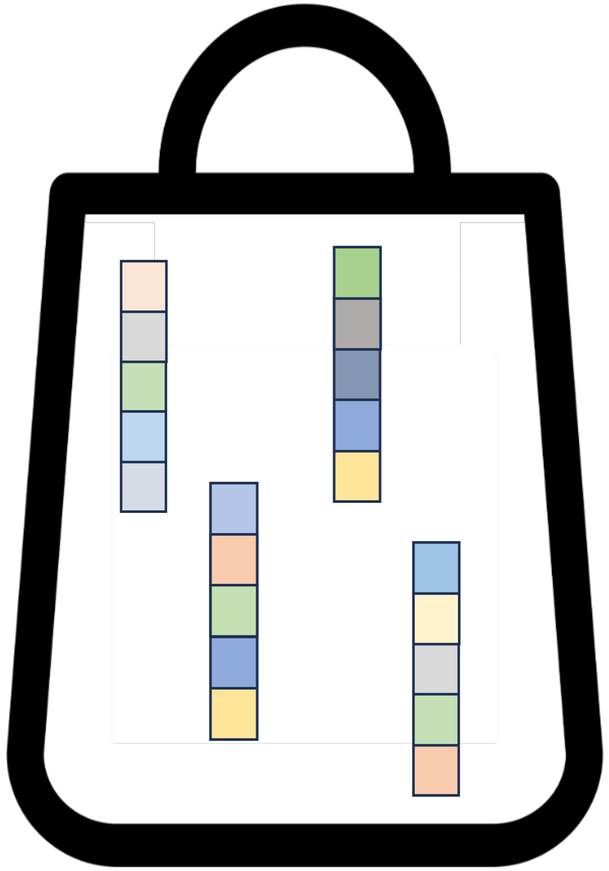
Neurips 2024

Linear Latent Spaces

Dataset of Images



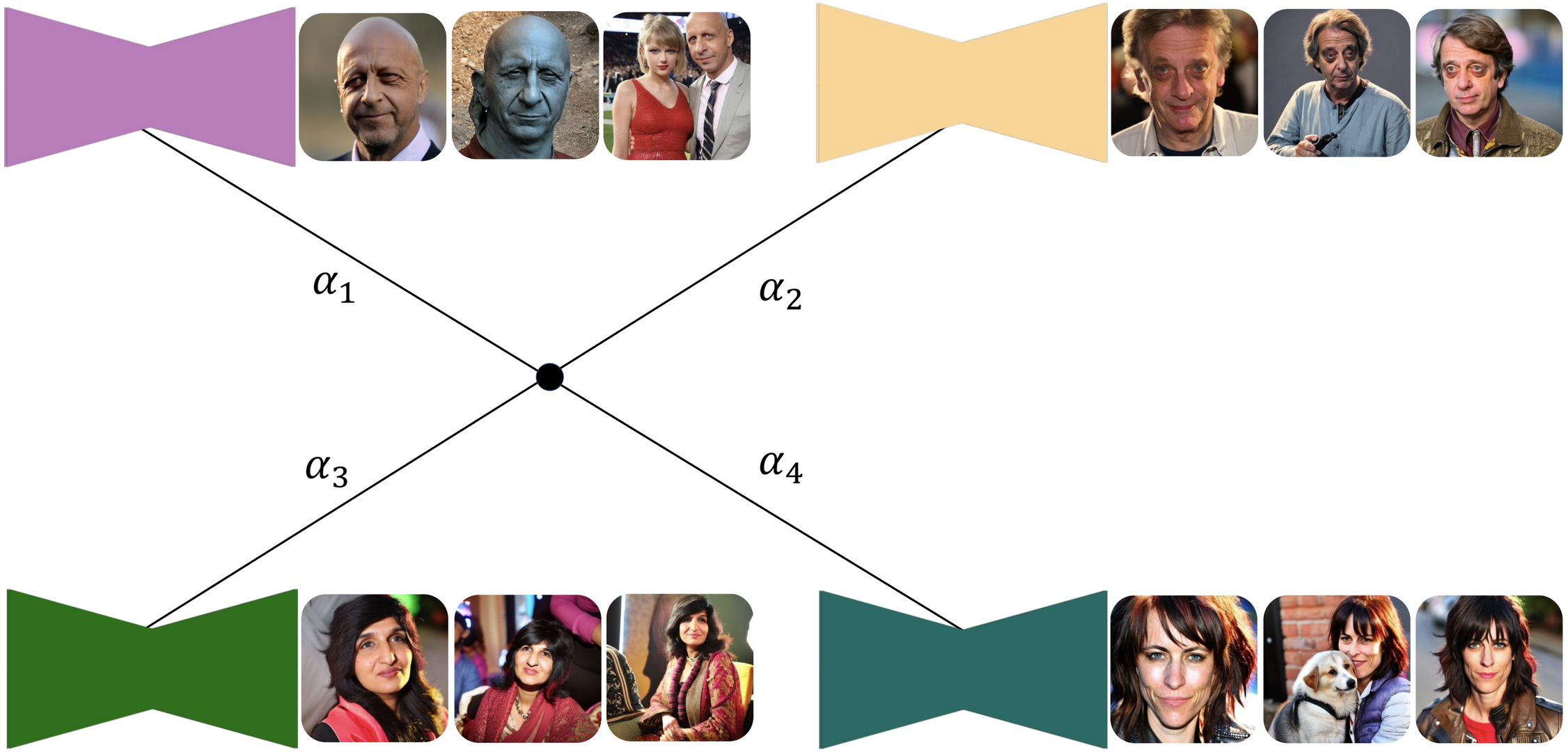
Dataset of Latent Vectors



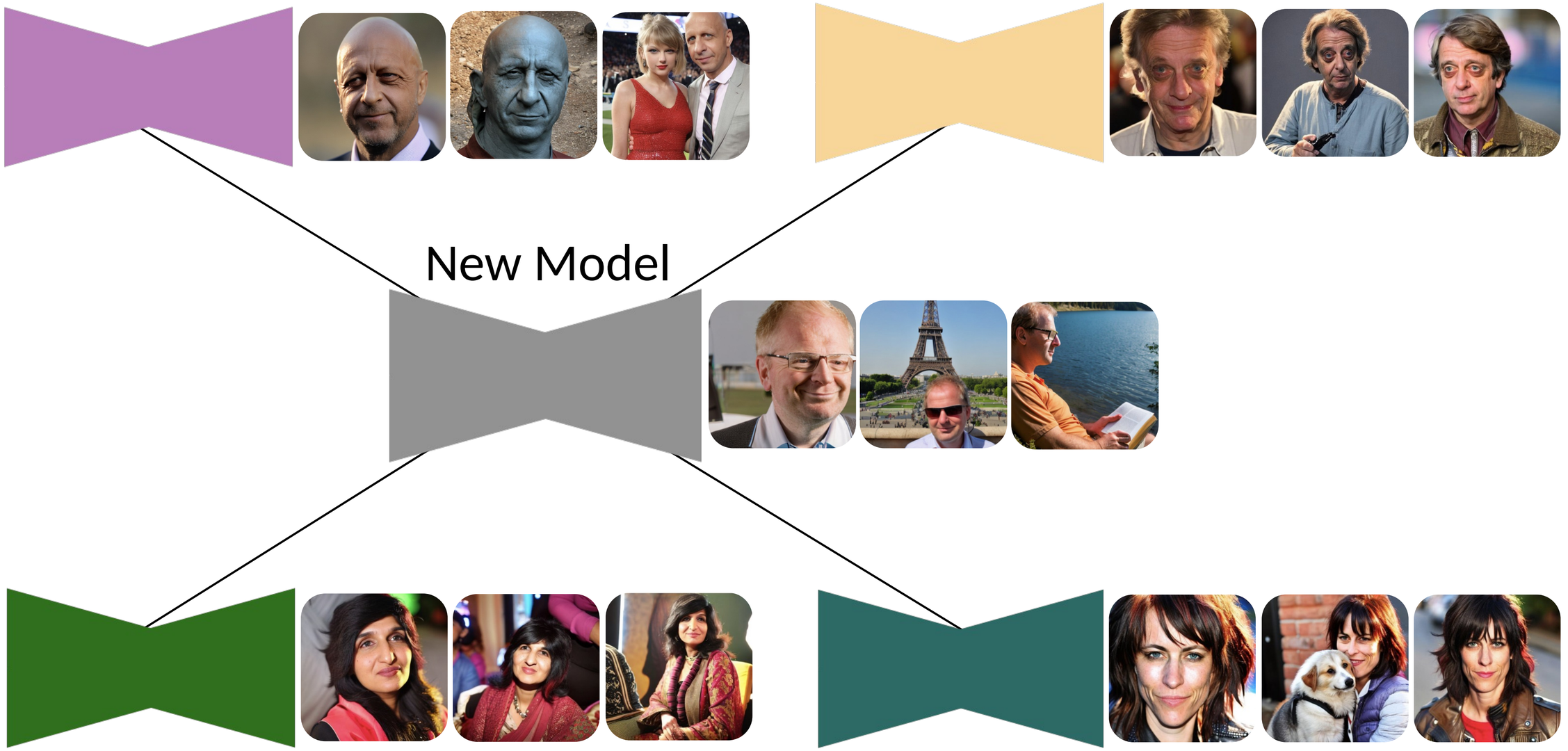
Dataset of Model Weights



Interpolation in Weight Space



Interpolation in Weight Space

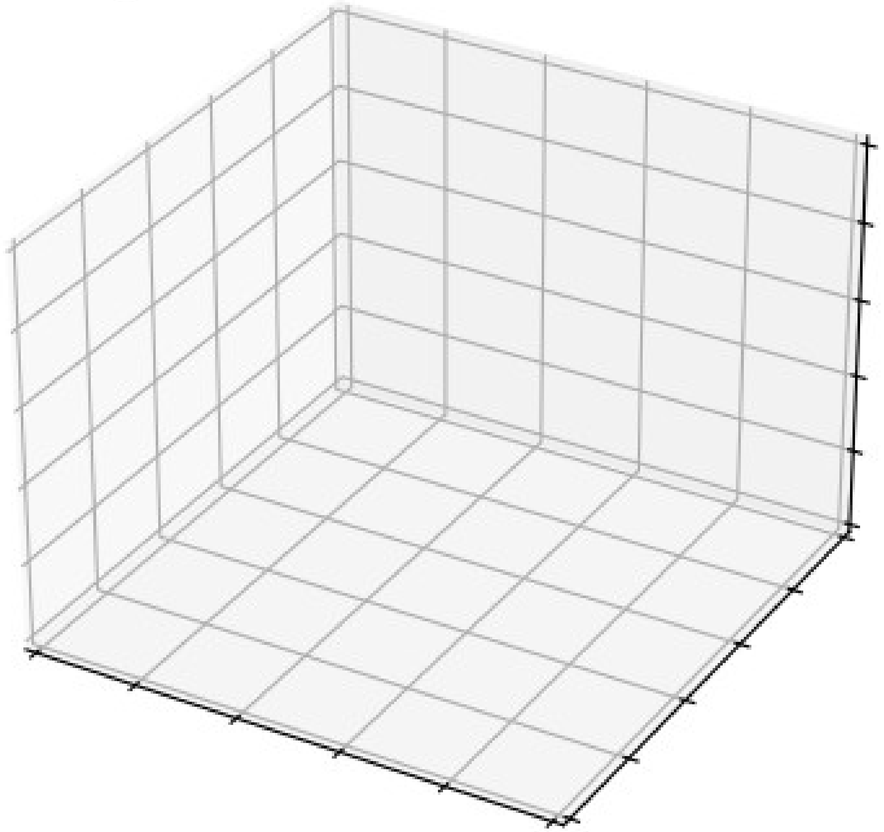


Creating a Dataset of Models



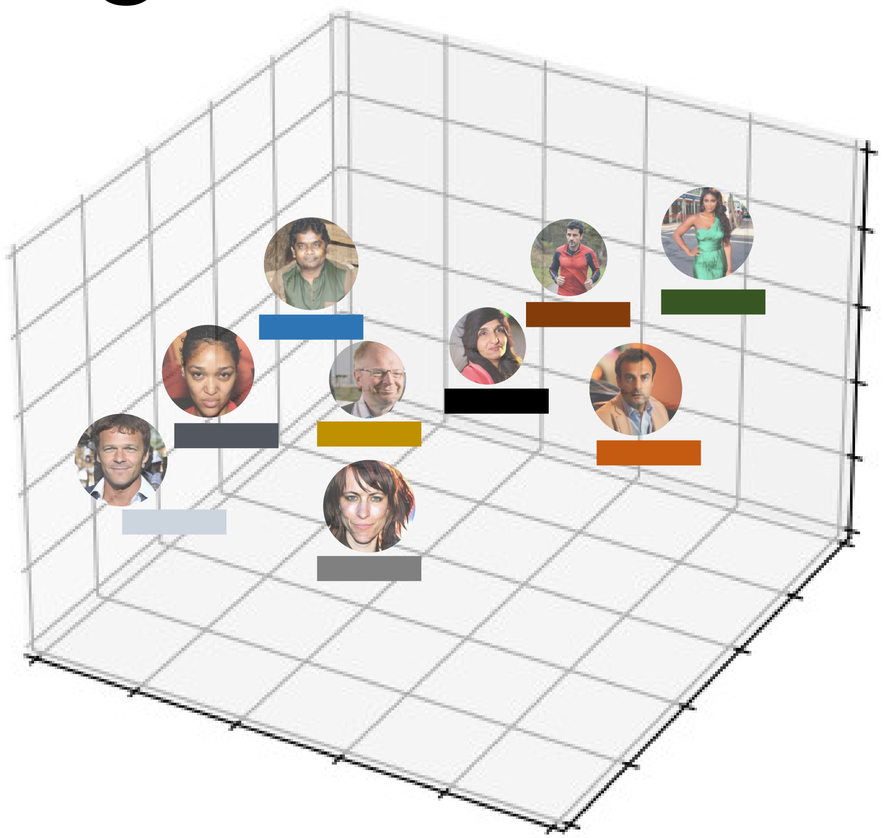
Creating a Dataset of Models

Model Weight Space



Creating a Dataset of Models

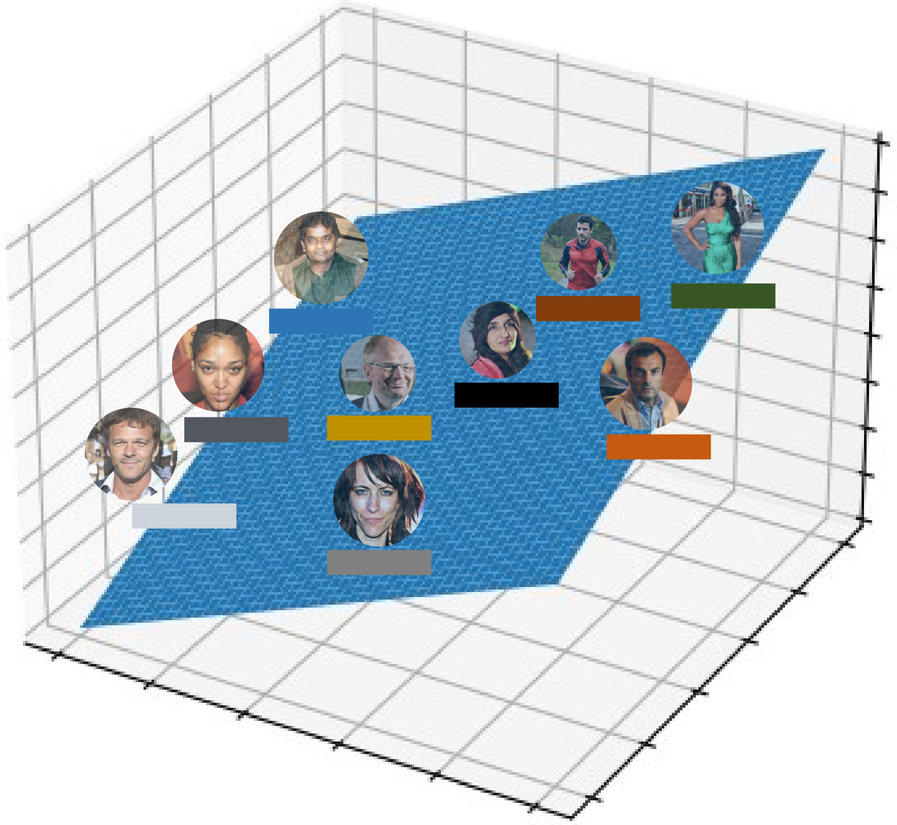
Model Weight Space



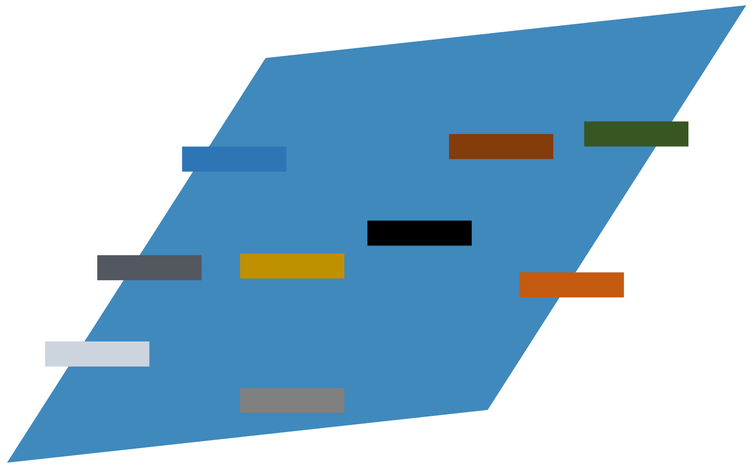
60,000+ Diffusion Models!

weights2weights (w2w) Space

Model Weight Space



weights2weights (w2w) Space



weights2weights (w2w) Space

PCA Basis: $\boldsymbol{w} = \{w_1, \ldots w_m\}$

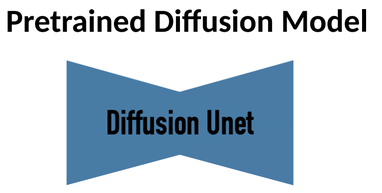


"Meta"-Latent Space

How Do We Get These Models?

Fine-Tune for each Task

Task = Generate particular
visual concept (e.g., identity)



Repeat over 60,000 Times





Small Set of Training Images



Diffusion Unet

Diffusion Unet

Fine-Tune Diffusion Model

Diffusion Unet

Pretrained Diffusion Model

Diffusion Unet



Diffusion Unet





Diffusion Unet

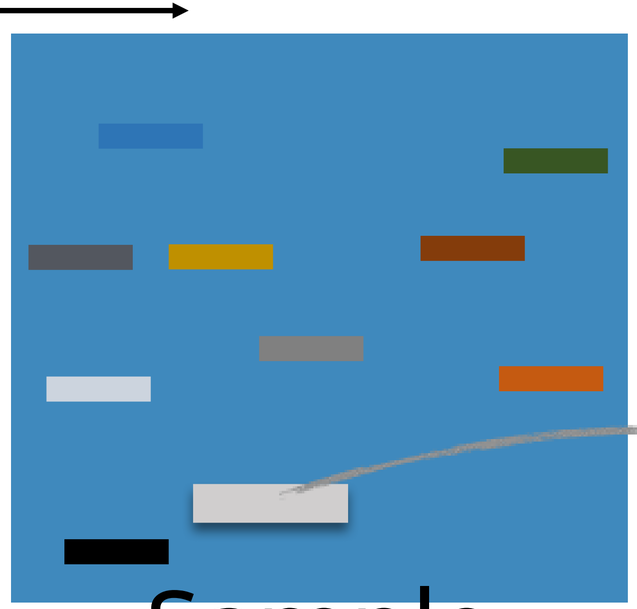
Diffusion Unet

Synthetic dataset of identities based on CelebA Dataset:
Deep Learning Face Attributes in the Wild, Liu et al., 2015

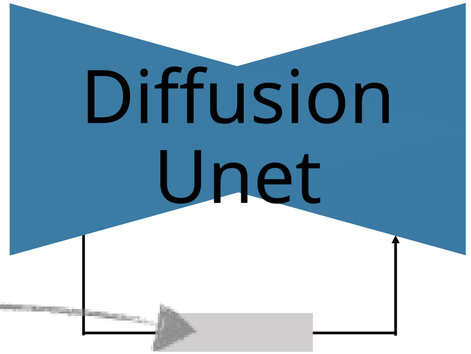
Sampling New Models

weights2weights Space

PCA Basis: $\boldsymbol{w} = \{w_1, \ldots w_m\}$



Sample weights



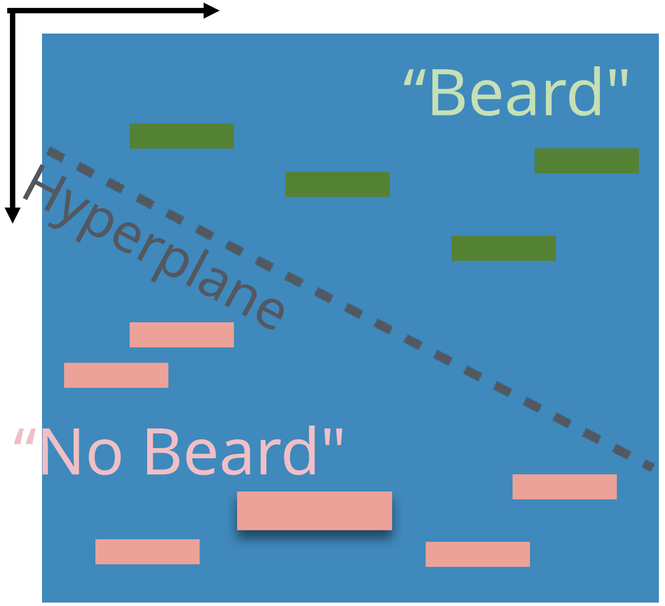
Generate Identity

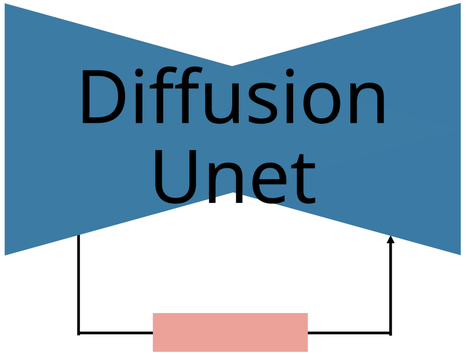


Editing Models

weights2weights Space

PCA Basis: $\boldsymbol{w} = \{w_1, \ldots w_m\}$





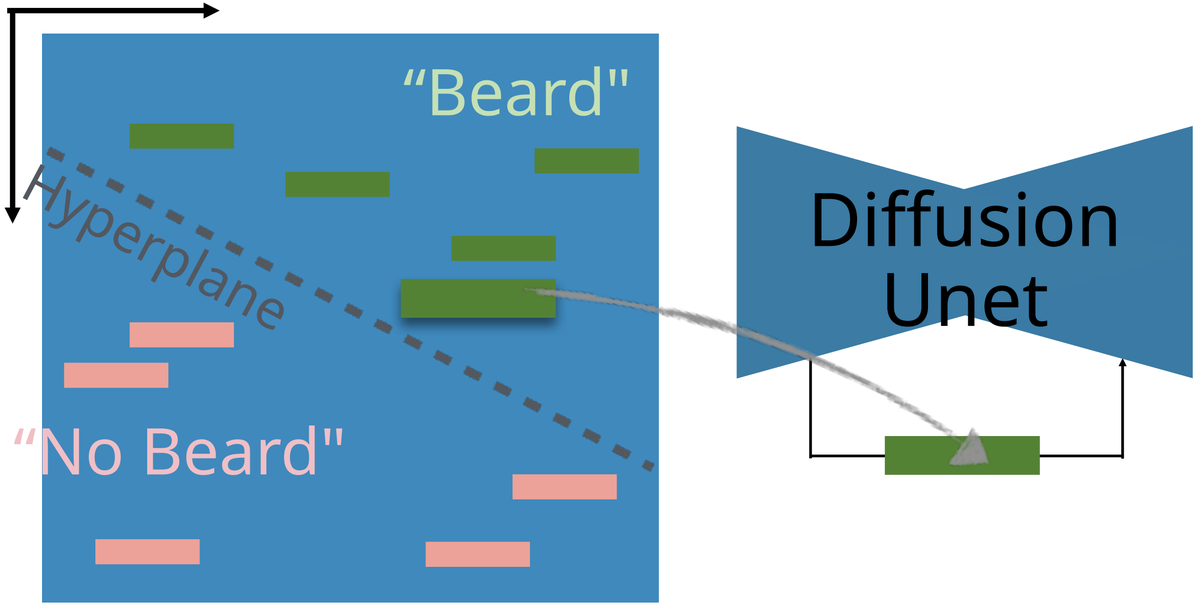
Generate Identity



Editing Models

weights2weights Space

PCA Basis: $\boldsymbol{w} = \{w_1, \ldots w_m\}$



Generate Identity



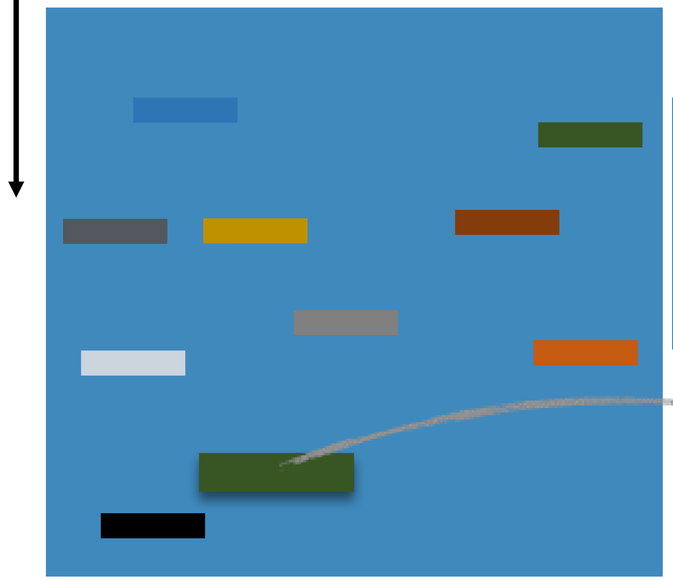
Inversion into *w2w* Space

weights2weights Space

PCA Basis: $\boldsymbol{w} = \{w_1, \ldots w_m\}$

Input





Diffusion Unet

Generate Identity







Invert into weights

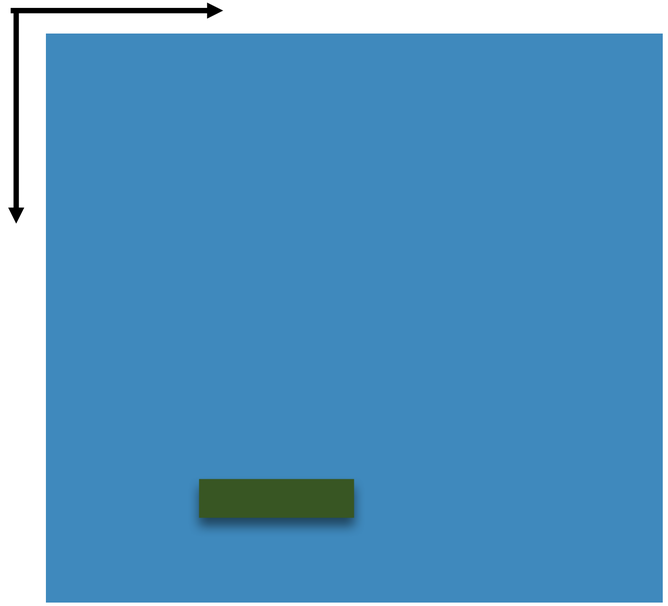
Inversion into *w2w* Space

weights2weights Space

PCA Basis: $\boldsymbol{w} = \{w_1, \ldots w_m\}$

Input





Invert into weights

$$\max_{\theta} \mathbb{E}_{\mathbf{x},\mathbf{c},\epsilon,t}\left[w_t \left|\left|\epsilon - \epsilon_\theta(\mathbf{x}_t, \mathbf{c}, t)\right|\right|_2^2\right] \quad \text{s.t. } \theta \in w2w$$

- Standard diffusion denoising objective with a subspace constraint
- Enforce constraint by operating directly on principal component basis

Inversion into *w2w* Space



Out-of-Distribution Inversion

Input



Projection



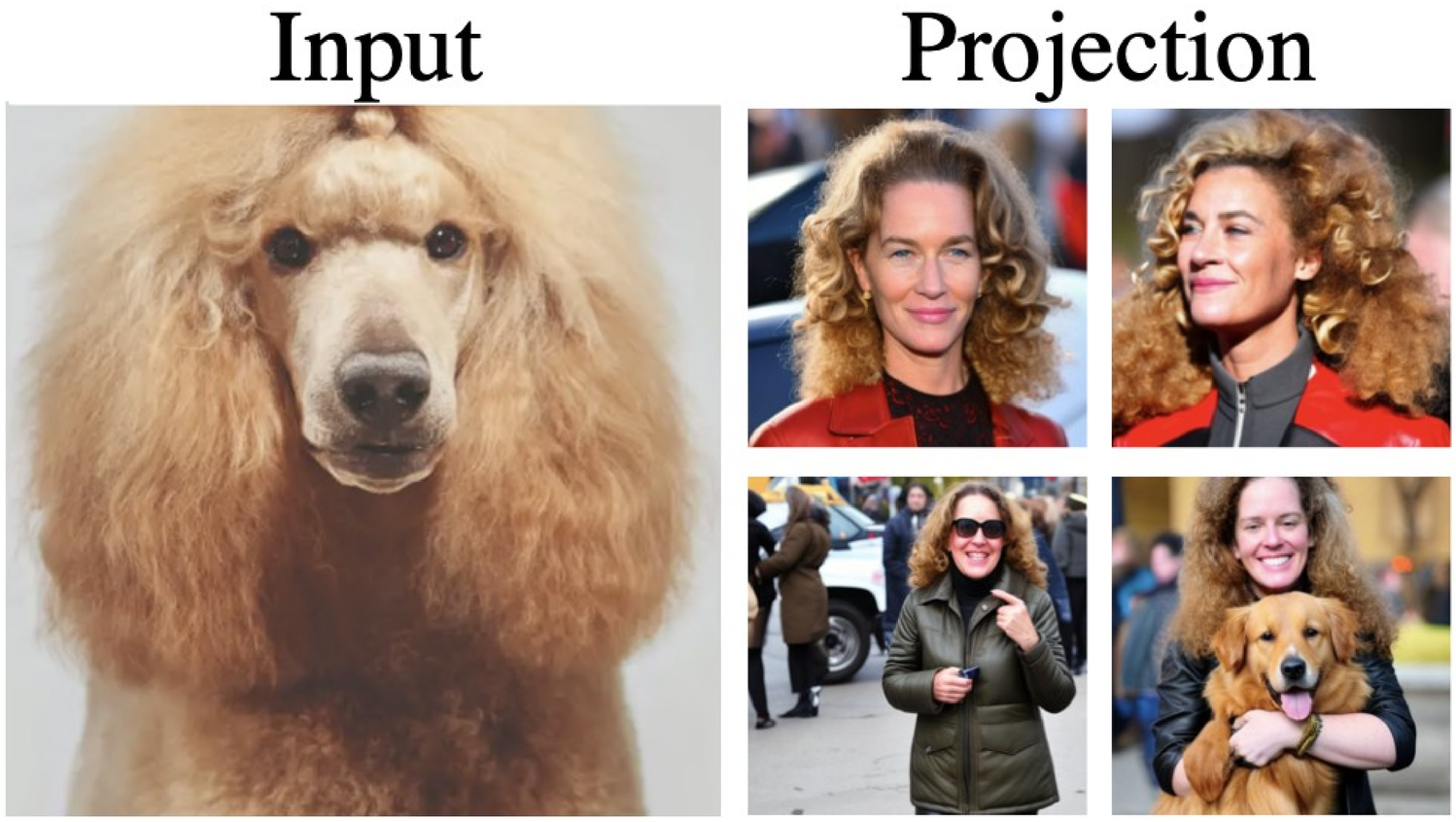
Input



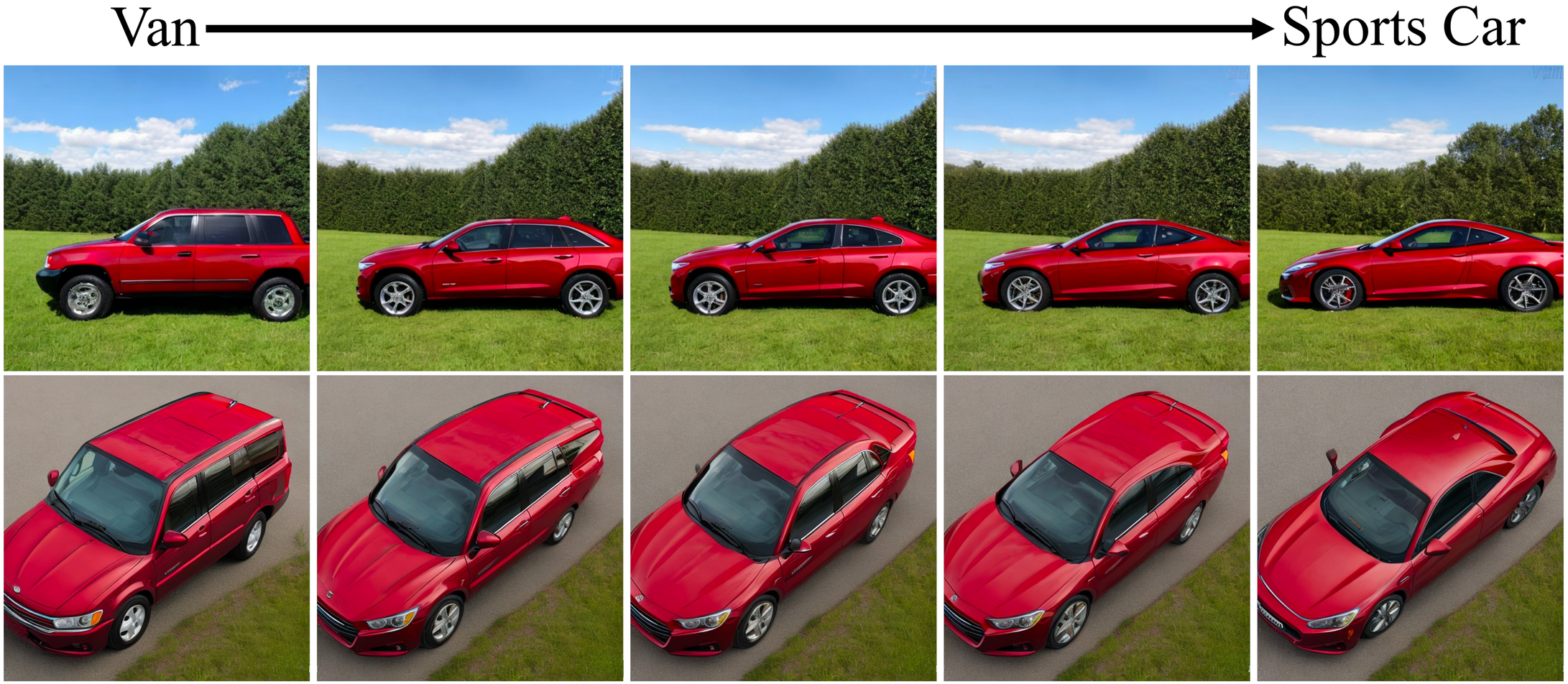
Projection



Out-of-Distribution Inversion



Extending to Other Tasks



Linear subspaces exist for other visual concepts

Train using cars from: *3D Object Representations for Fine-Grained Categorization* (Stanford Cars Dataset), Krause et al., 2013

Datasets and their discontents

Are datasets measuring the right thing?

- In Machine Learning:

  Dataset is The World

- In Recognition

  Dataset is a *representation* of The World

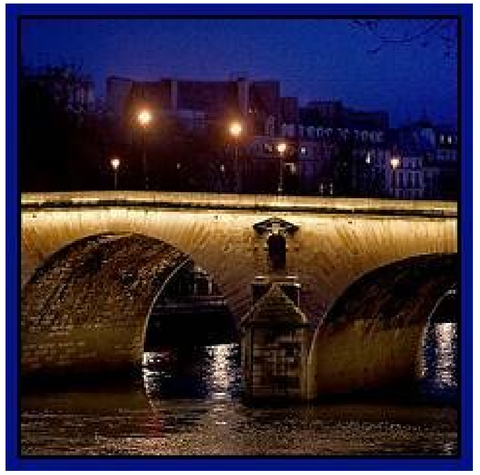
- Do datasets provide a good representation?

Visual Data is Inherently Biased

- Internet is a tremendous repository of visual data (Flickr, YouTube, Picassa, etc)
- But it’s not random samples of visual world
  - And is “random” the right thing anyway?

Flickr Paris







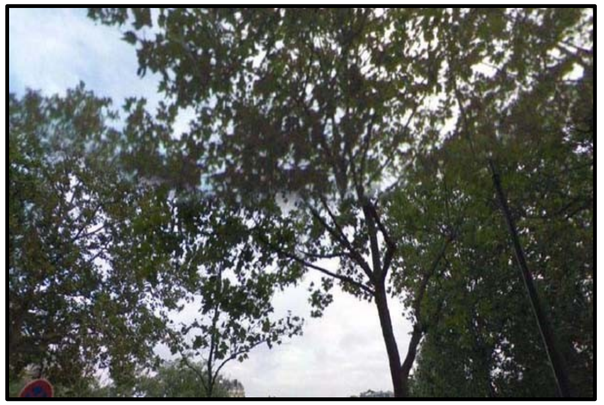








Google StreetView Paris















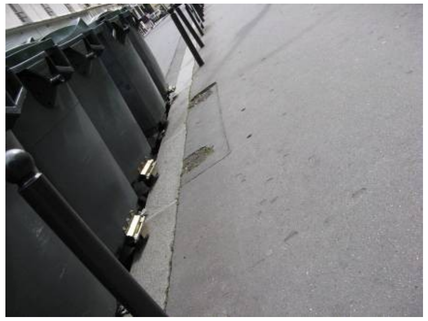


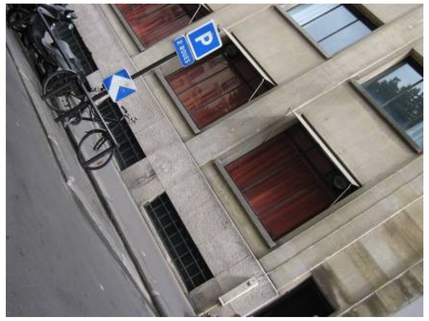
Knopp, Sivic, Pajdla, ECCV 2010

“Uniformly Sampled” Paris











Most “collection bias” is subcontious

mug Search SafeSearch moderate ▼

About 10,100,000 results (0.09 seconds) Advanced search

59¢ Logo Coffee Mugs
www.DiscountMugs.com Lead Free & Dishwasher Safe. Save 40-50%. No Catch. Factory Direct !

Custom Mugs On Sale
www.Vistaprint.com Order Now & Save 50% On Custom Mugs No Minimums. Upload Photos & Logos.

Promotional Mugs from 69¢ Sponsored
www.4imprint.com/Mugs Huge Selection of Style Colors- Buy 72 Mugs @ $1.35 ea-24hr Service

Related searches: white mug coffee mug mug root beer mug shot


Representational
500 × 429 - 91k - jpg
eagereyes.org
Find similar images


Ceramic Happy Face
300 × 300 - 77k - jpg
larose.com
Find similar images


Here I go then, trying
600 × 600 - 35k - jpg
beeper.wordpress.com
Find similar images



The Chalk Mug »
304 × 314 - 17k - jpg
coolest-gadgets.com
Find similar images



mug
300 × 279 - 54k - jpg
reynosawatch.org
Find similar images


Bring your own
500 × 451 - 15k - jpg
cookstownunited.ca
Find similar images


ceramic mug
980 × 1024 - 30k - jpg
diytrade.com


Dual Purpose Drinking
490 × 428 - 16k - jpg
freshome.com
Find similar images


This coffee mug,
300 × 300 - 22k - jpg
gizmodo.com
Find similar images


Back to Ceramic
400 × 400 - 11k - jpg
freshpromotions.com.au
Find similar images


Back to Ceramic
400 × 400 - 8k - jpg
freshpromotions.com.au
Find similar images


Coffee Mug as a
303 × 301 - 10k - jpg
dustbowl.wordpress.com
Find similar images


SASS Life Member
300 × 302 - 6k - jpg
sassnet.com



personalized coffee
400 × 343 - 15k - jpg
walyou.com
Find similar images


We like our mugs
290 × 290 - 6k - jpg
kitchencontraptions.com
Find similar images

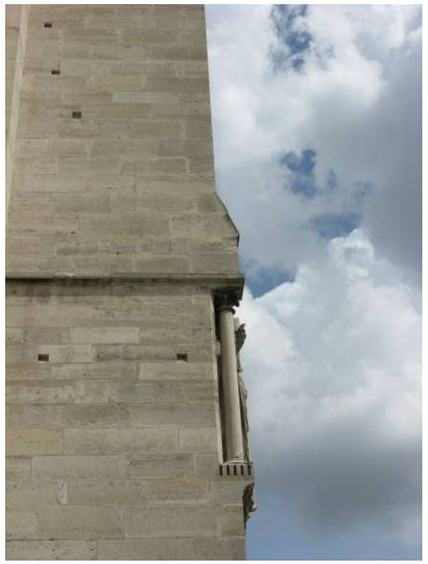
Sampling Bias

- People like to take pictures on vacation



Photographer Bias

- People want their pictures to be recognizable and/or interesting



vs.



Social Bias

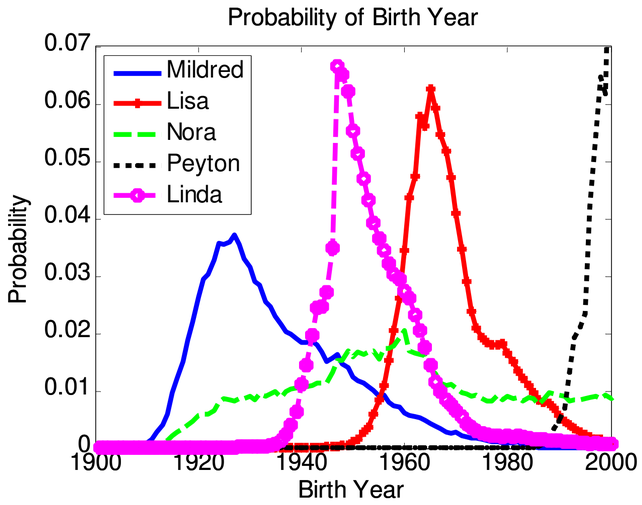


“100 Special Moments” by Jason Salavon

Social Bias



Mildred and Lisa



Source: U.S. Social Security Administration

Gallagher et al CVPR 2008

How much does this bias affect standard recognition datasets?

Poor Cross-Dataset Generalization

MSRC



Classifier trained on MSRC cars

[Torralba & Efros, 2011]

ImageNet vs. ImageNetV2

ImageNet

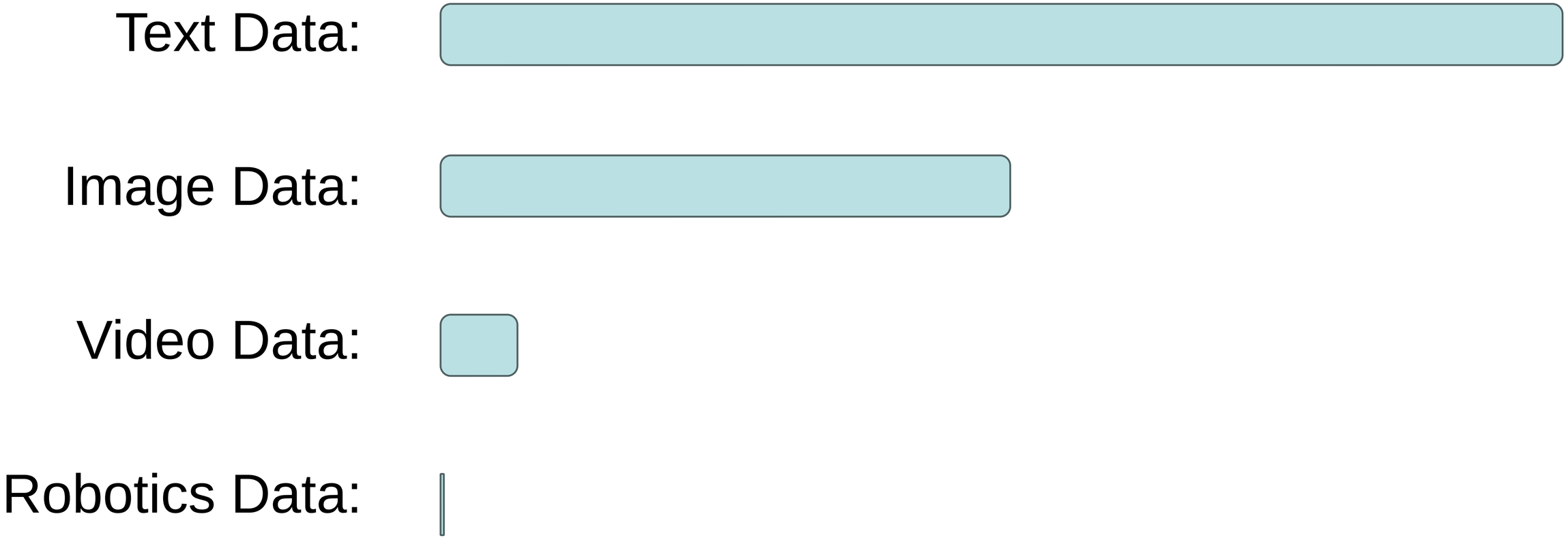


ImageNetV2



Chapter 4: Epilogue

Data is best predictor for success:

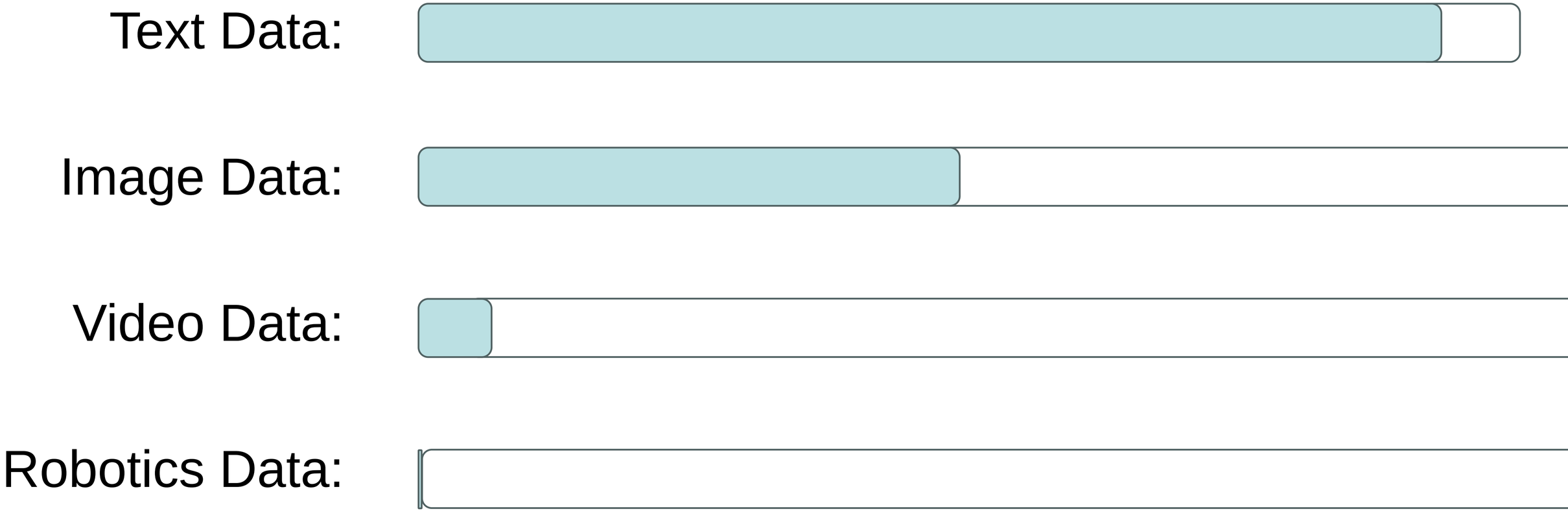


Moravec's Paradox lives on!

“In AI, the hard things are easy while the easy things are hard”



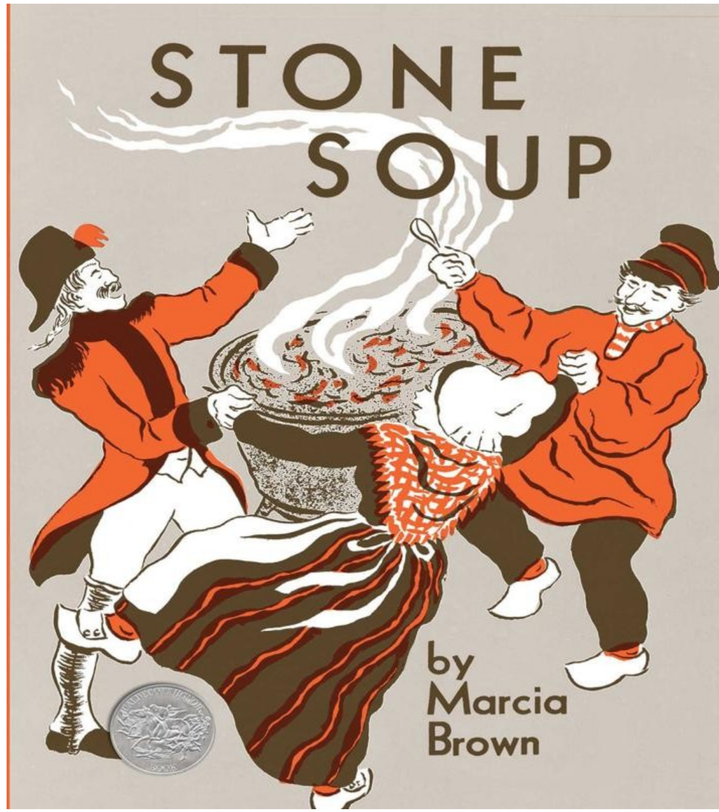
Data availability going forward:





Gopnik’s Parable of Stone Soup AI

- Tech execs come to village:
  - *“We will make amazing AI with just a few magical algorithms (e.g. gradient descent, transformers, next token prediction)”*
  - *…but would be so much better with more data*
  - *…even more magical with human feedback*
  - *…even better if humans would learn to ask questions in the right way*
- Villagers: wow, what magical AI just from a few algorithms!



https://simons.berkeley.edu/news/stone-soup-ai

Take Aways

- Is it all just data, with no place for algorithms?
  - No! Large-scale data is clearly **necessary**, but it is **not sufficient**

- Need to learn to **be humble** and give data the credit it is due