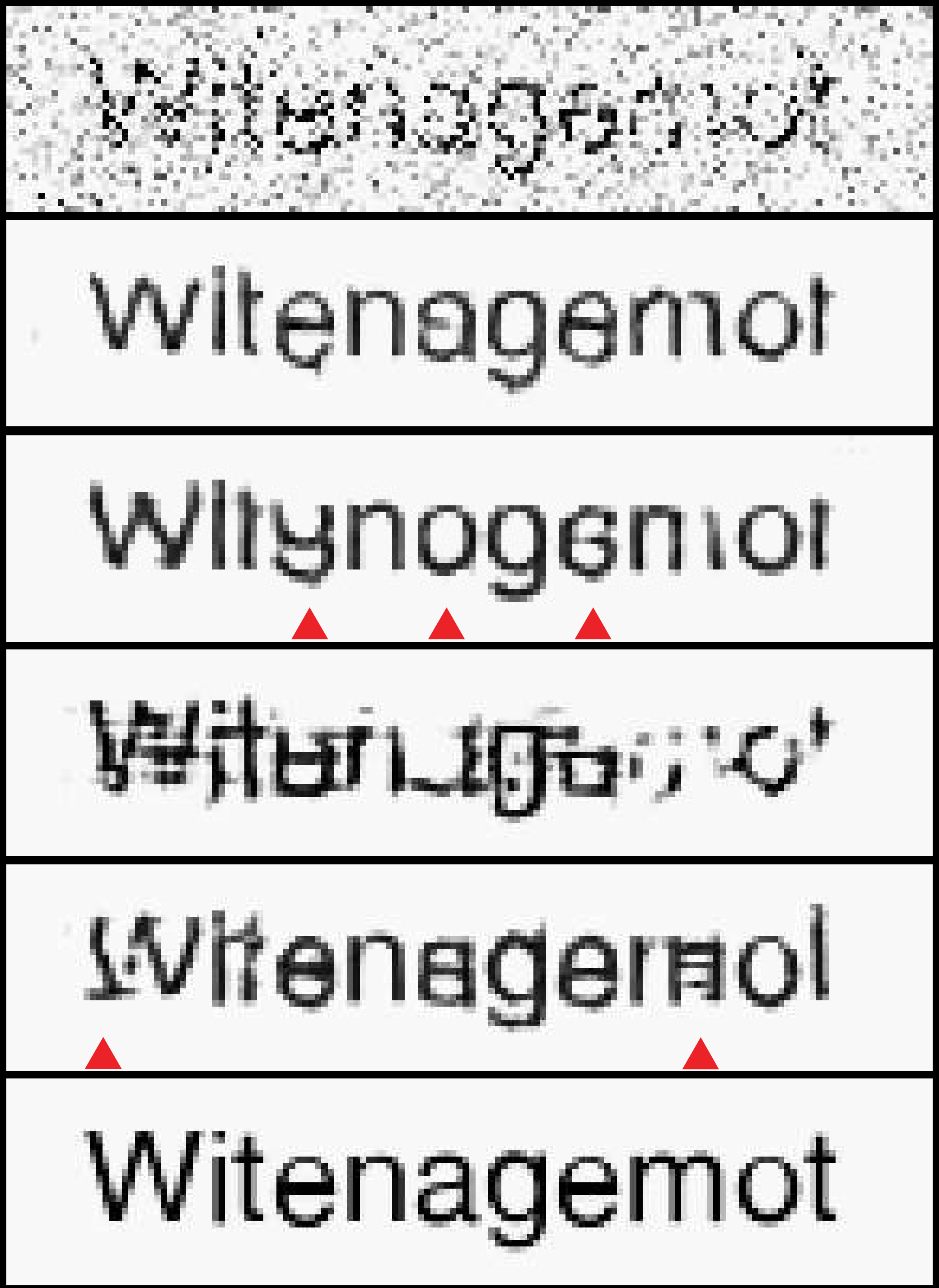
Limits and Pitfalls of Deep Learning for biological imaging

Loïc A. Royer
@loicaroyer

CHAN ZUCKERBERG
BIOHUB



The Ground Truth Problem

The Training Problem

The Overfitting Problem

The Adversarial Fragility Problem

The Hallucination Problem

The Cheating Problem

The Black Box Problem

The ‘not-that-smart’ Problem

KEEP
CALM
AND
STAY
POSITIVE

The Ground Truth Problem

The Training Problem

The Overfitting Problem

The Adversarial Fragility Problem

The Hallucination Problem

The Cheating Problem

The Black Box Problem

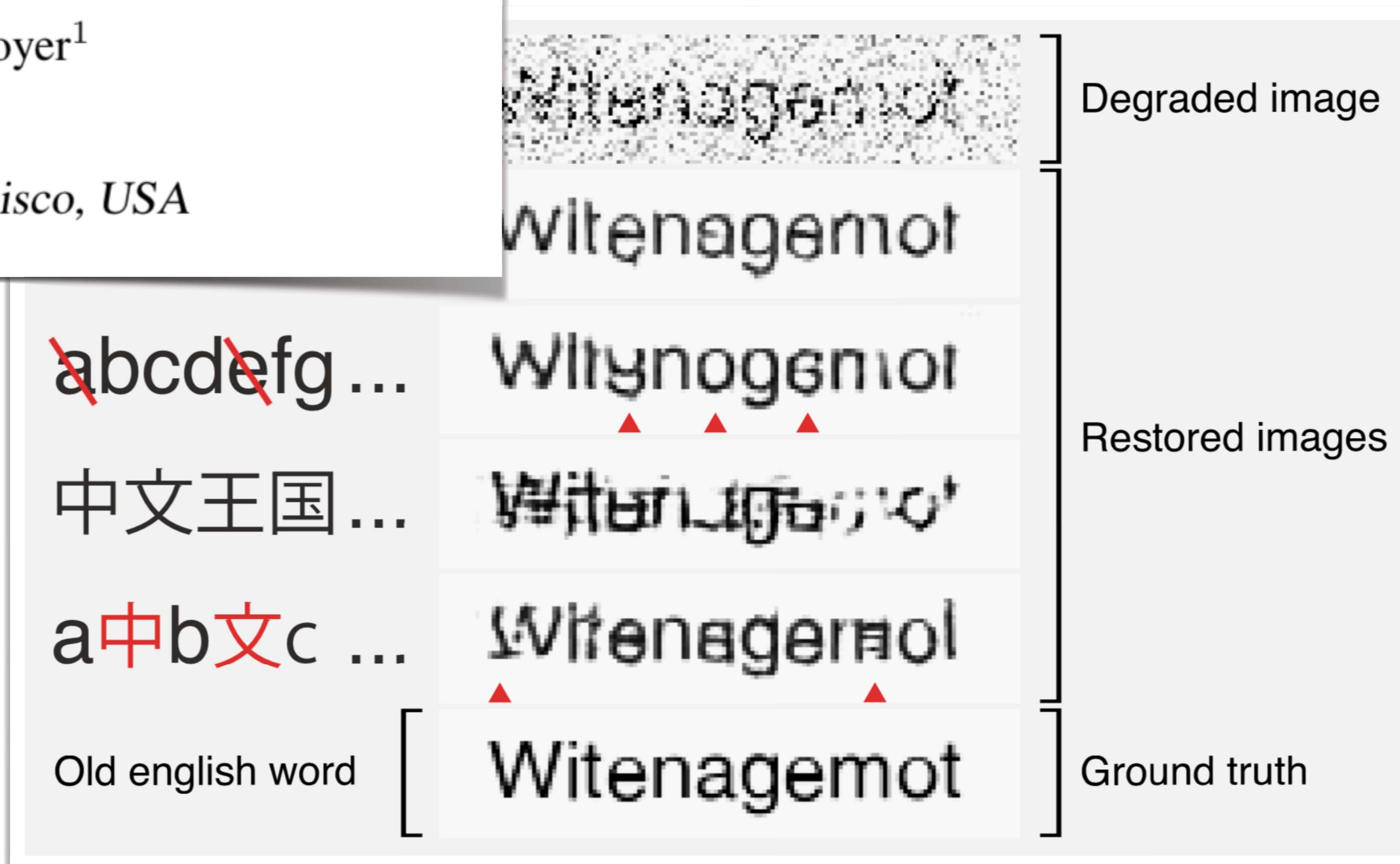
The ‘not-that-smart’ Problem

The Ground Truth Problem

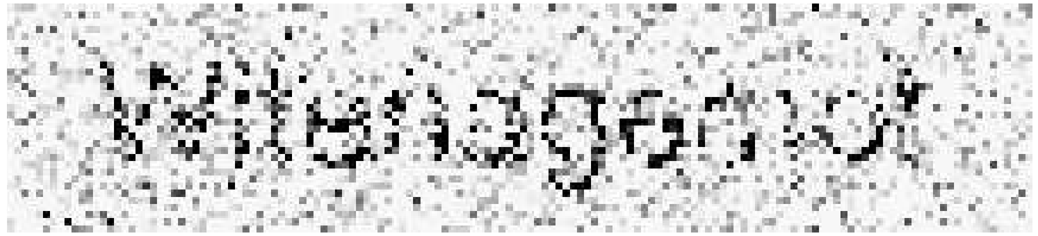
Applications, Promises, and Pitfalls of Deep Learning for Fluorescence Image Reconstruction

Chinmay Belthangady[1], & Loic A. Royer[1]

[1] *Chan Zuckerberg Biohub, San Francisco, USA*



Belthangady & Royer, Nature Methods, 2019



Belthangady & Royer, Nature Methods, 2019

Training datasets

abcdefg …

Belthangady & Royer, Nature Methods, 2019

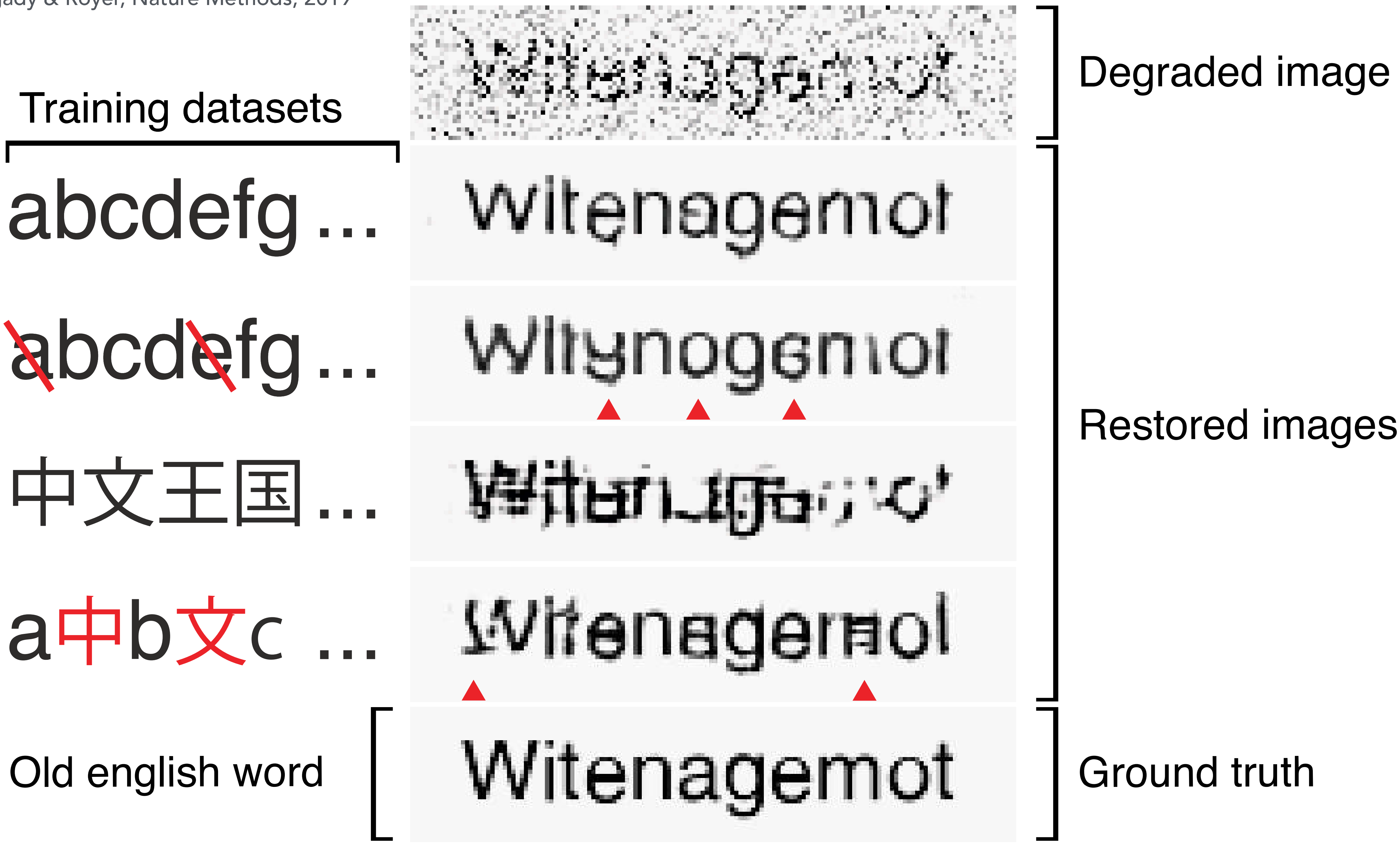
Training datasets

abcdefg …

~~a~~bcd~~e~~fg …

Belthangady & Royer, Nature Methods, 2019

Belthangady & Royer, Nature Methods, 2019

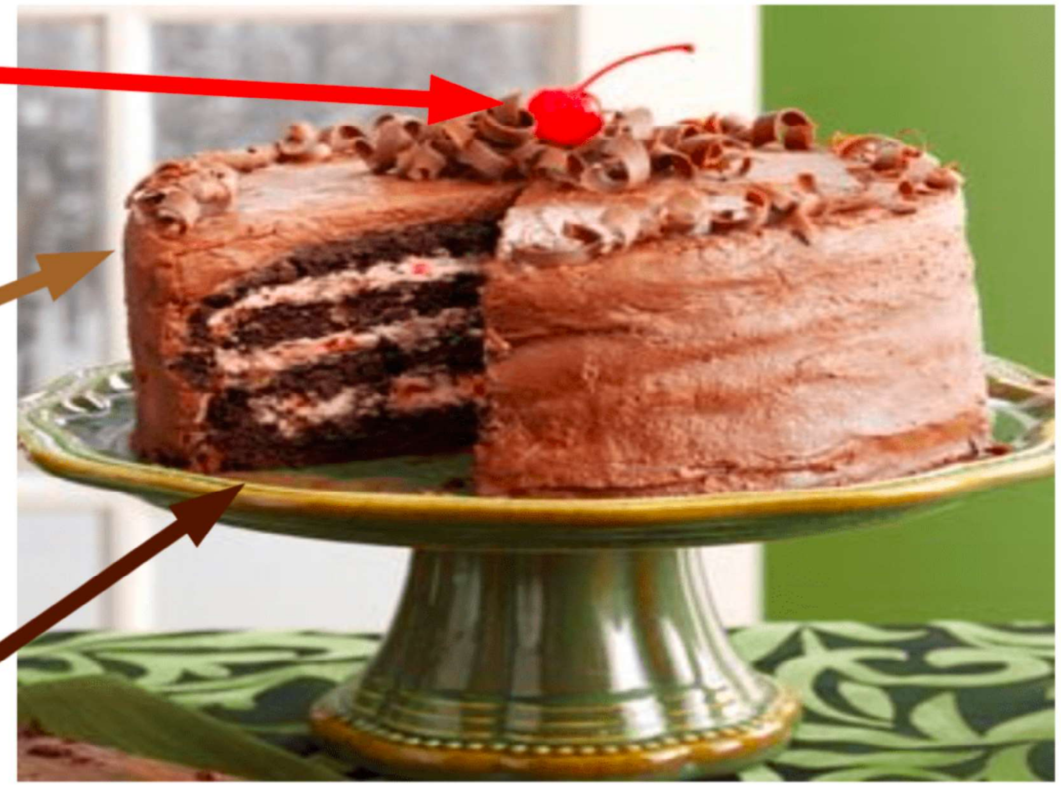


Ideally, we would not need ‘Ground Truth’

How Much Information is the Machine Given during Learning?

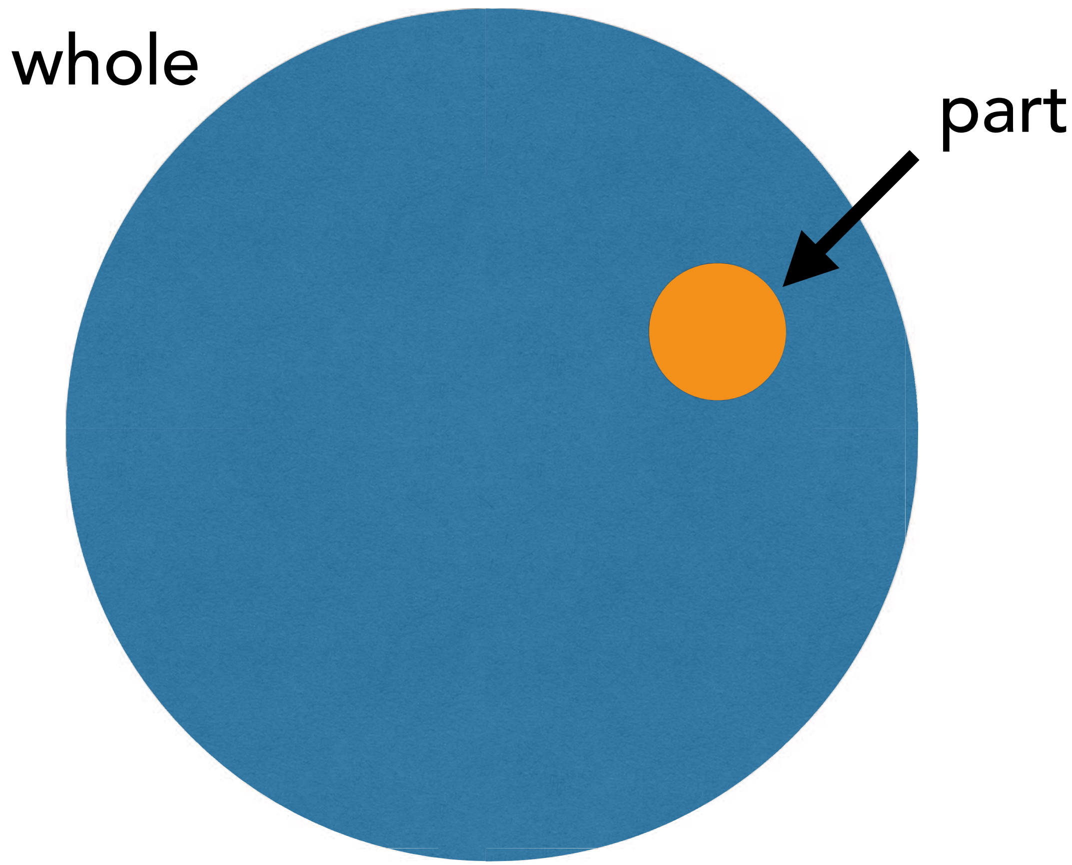
Y. LeCun

- **“Pure” Reinforcement Learning (cherry)**
  - The machine predicts a scalar reward given once in a while.
  - **A few bits for some samples**
- **Supervised Learning (icing)**
  - The machine predicts a category or a few numbers for each input
  - Predicting human-supplied data
  - **10→10,000 bits per sample**
- **Self-Supervised Learning (cake génoise)**
  - The machine predicts any part of its input for any observed part.
  - Predicts future frames in videos
  - **Millions of bits per sample**



Self Supervised Learning

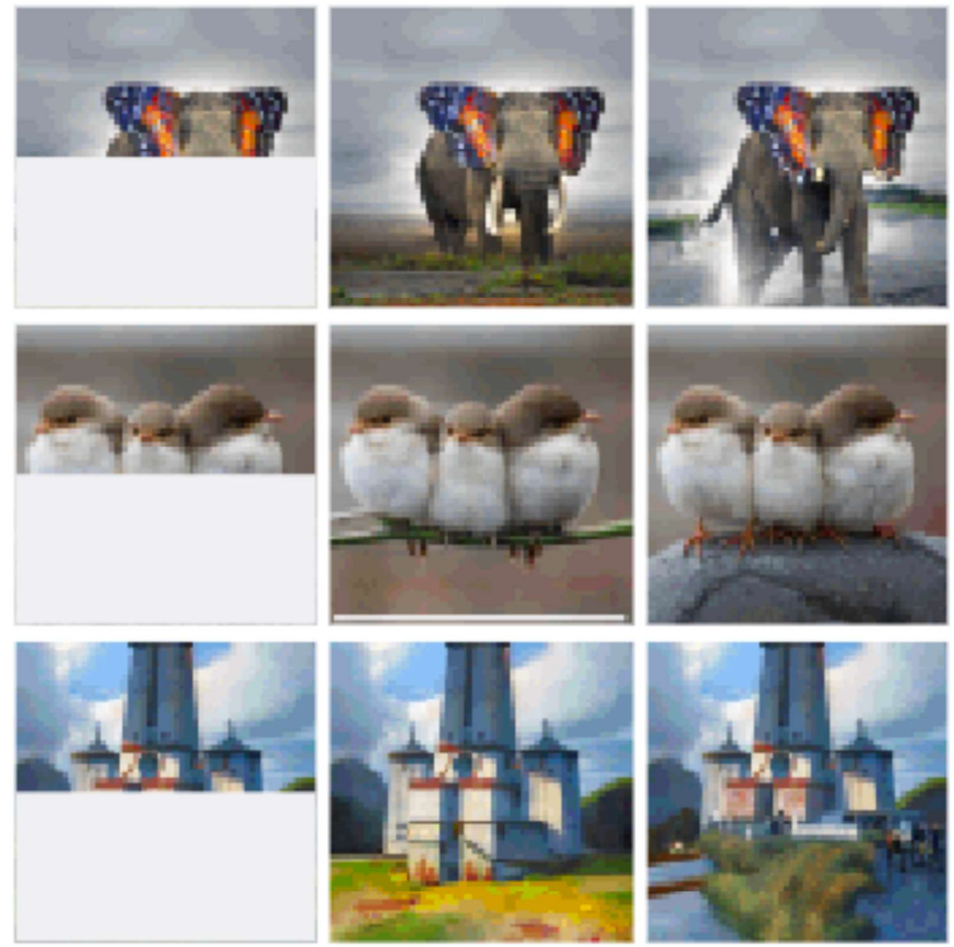
predict part from the whole



Self Supervised Learning

predict part from the whole

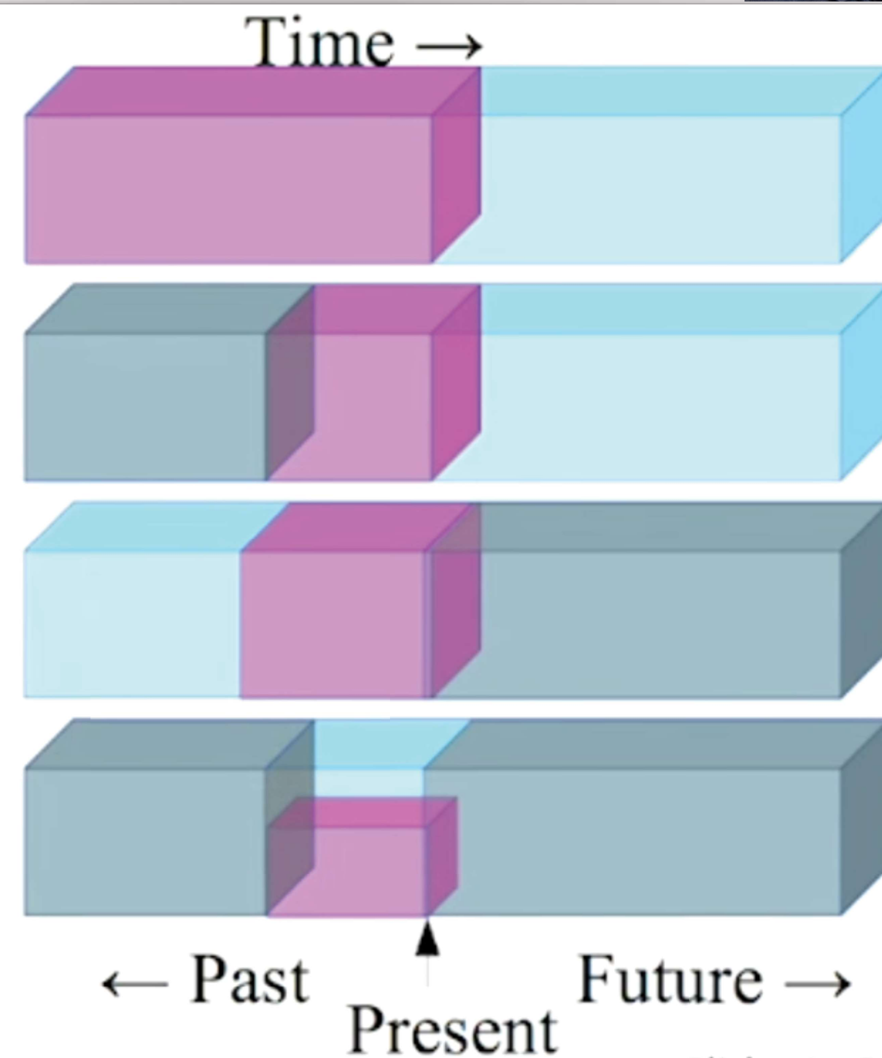
Chen *et al.* PMLR 2020



Self Supervised Learning

predict part from whole

- Predict any part of the input from any other part.
- Predict the future from the past.
- Predict the future from the recent past.
- Predict the past from the present.
- Predict the top from the bottom.
- Predict the occluded from the visible
- Pretend there is a part of the input you don't know and predict that.



Slide: LeCun

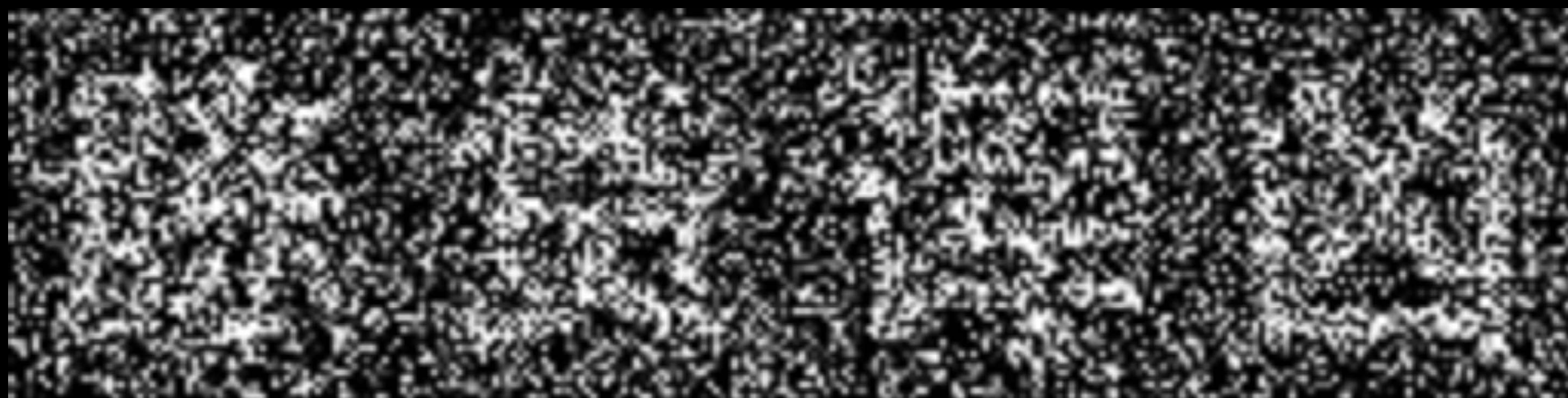
Image Denoising
with self-supervised deep learning

Joshua Batson

Noise2Self: Blind Denoising by Self-Supervision, ***Joshua Batson & Loic Royer,*** ICML (2019)
Noise2Void - Learning Denoising from Single Noisy Images, ***Krull, Buchholz, Jug,*** CVPR (2019)

Image Denoising
with self-supervised deep learning

Joshua Batson



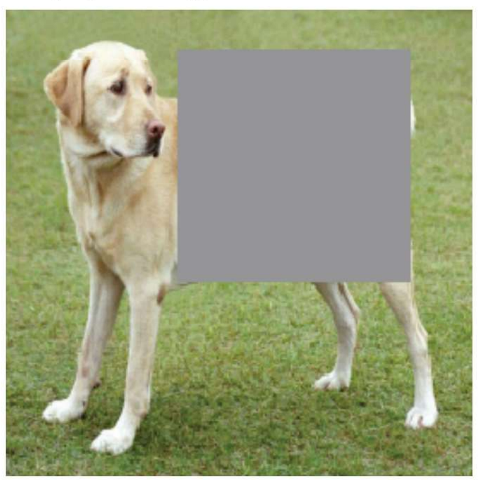
Noise2Self: Blind Denoising by Self-Supervision, ***Joshua Batson & Loic Royer,*** ICML (2019)
Noise2Void - Learning Denoising from Single Noisy Images, ***Krull, Buchholz, Jug,*** CVPR (2019)

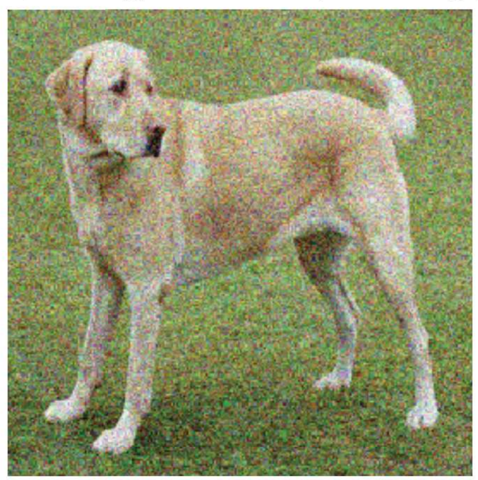
Self Supervised Learning

representations

(b) Crop and resize

(c) Crop, resize (and flip)





Chen *et al.* arXiv 2020

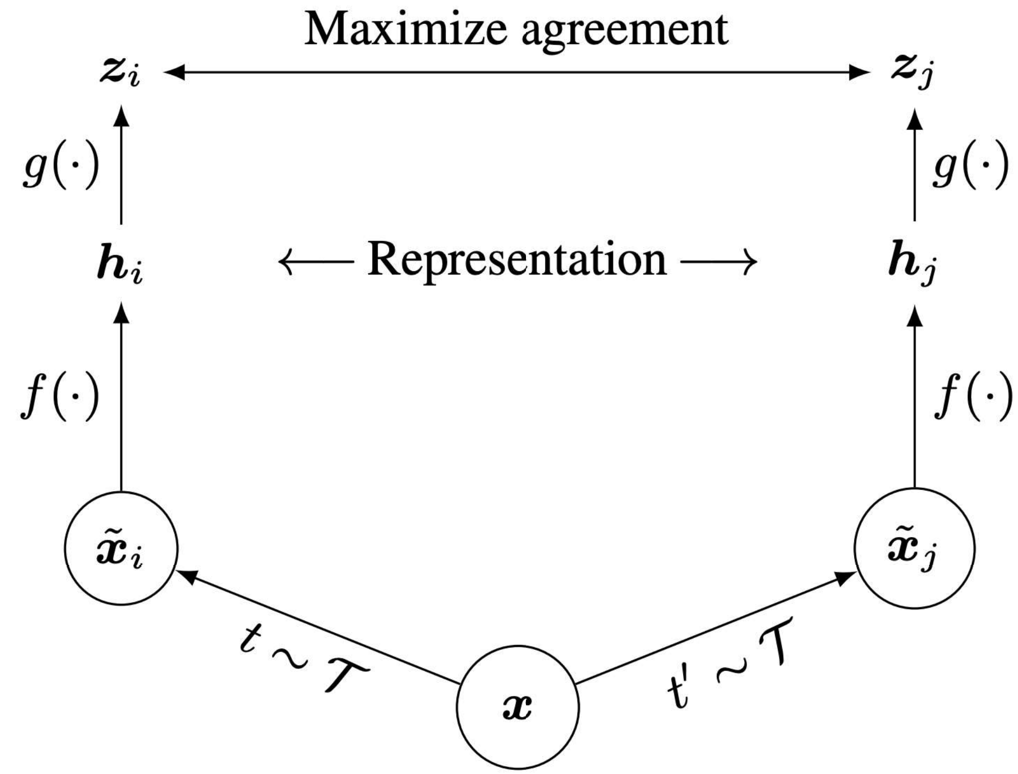


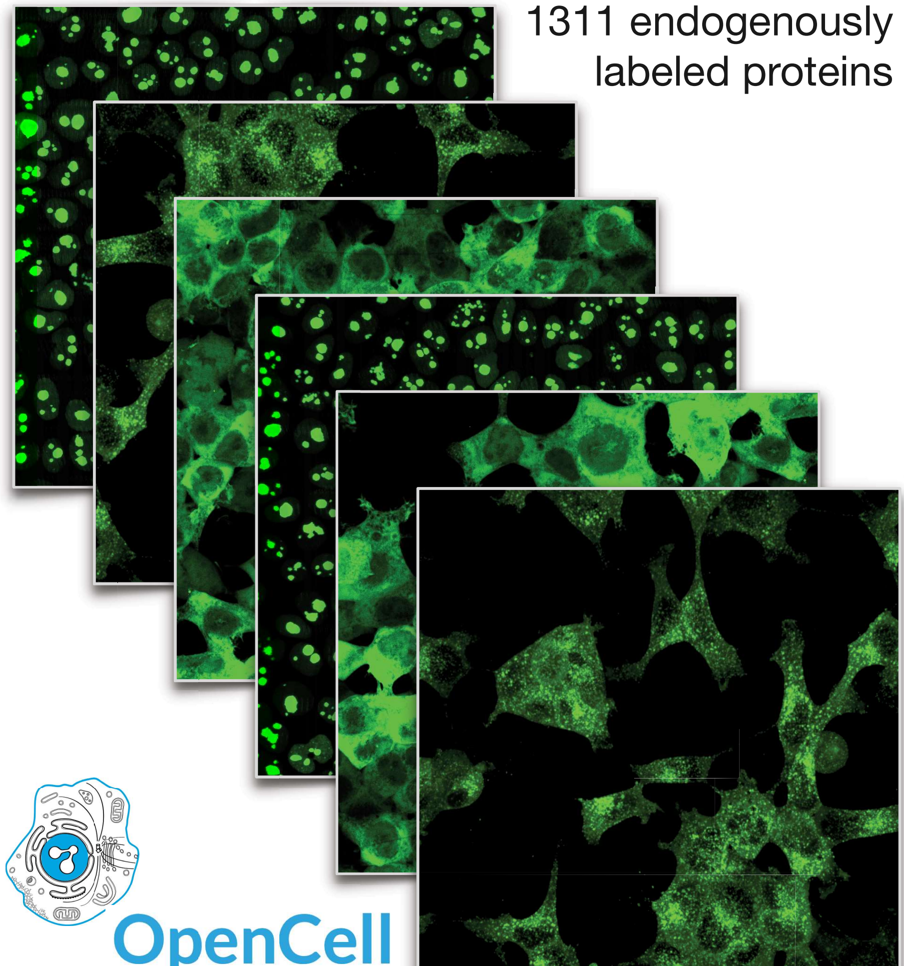
Figure 2. A simple framework for contrastive learning of visual representations. Two separate data augmentation operators are sampled from the same family of augmentations ($t \sim \mathcal{T}$ and $t' \sim \mathcal{T}$) and applied to each data example to obtain two correlated views. A base encoder network $f(\cdot)$ and a projection head $g(\cdot)$ are trained to maximize agreement using a contrastive loss. After training is completed, we throw away the projection head $g(\cdot)$ and use encoder $f(\cdot)$ and representation $\boldsymbol{h}$ for downstream tasks.

Self-Supervised Deep Learning Encodes High-Resolution Features of Protein Subcellular Localization

Hirofumi Kobayashi, Keith C. Cheveralls, Manuel D. Leonetti, Loic A. Royer

doi: https://doi.org/10.1101/2021.03.29.437595

1311 endogenously labeled proteins

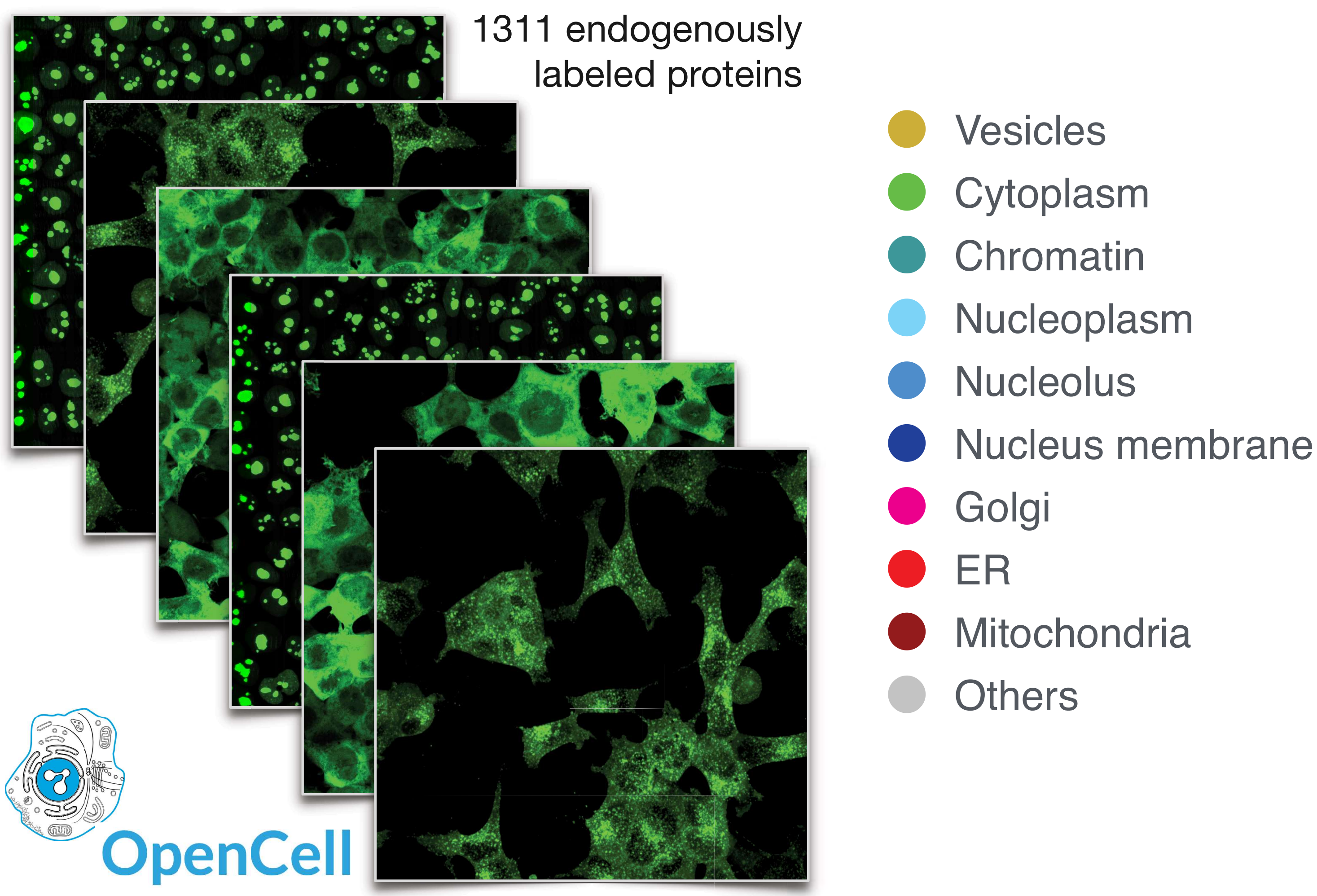


OpenCell

Self-Supervised Deep Learning Encodes High-Resolution Features of Protein Subcellular Localization

Hirofumi Kobayashi, Keith C. Cheveralls, Manuel D. Leonetti, Loic A. Royer

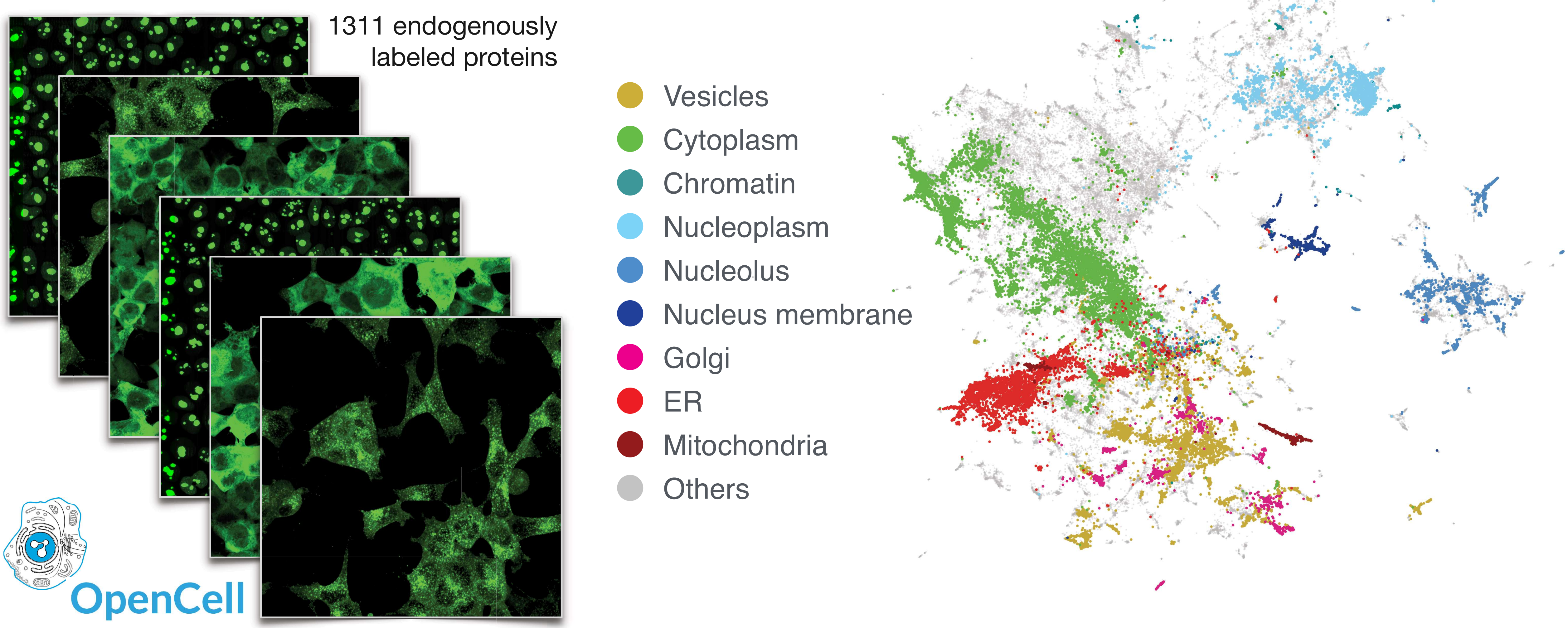
doi: https://doi.org/10.1101/2021.03.29.437595



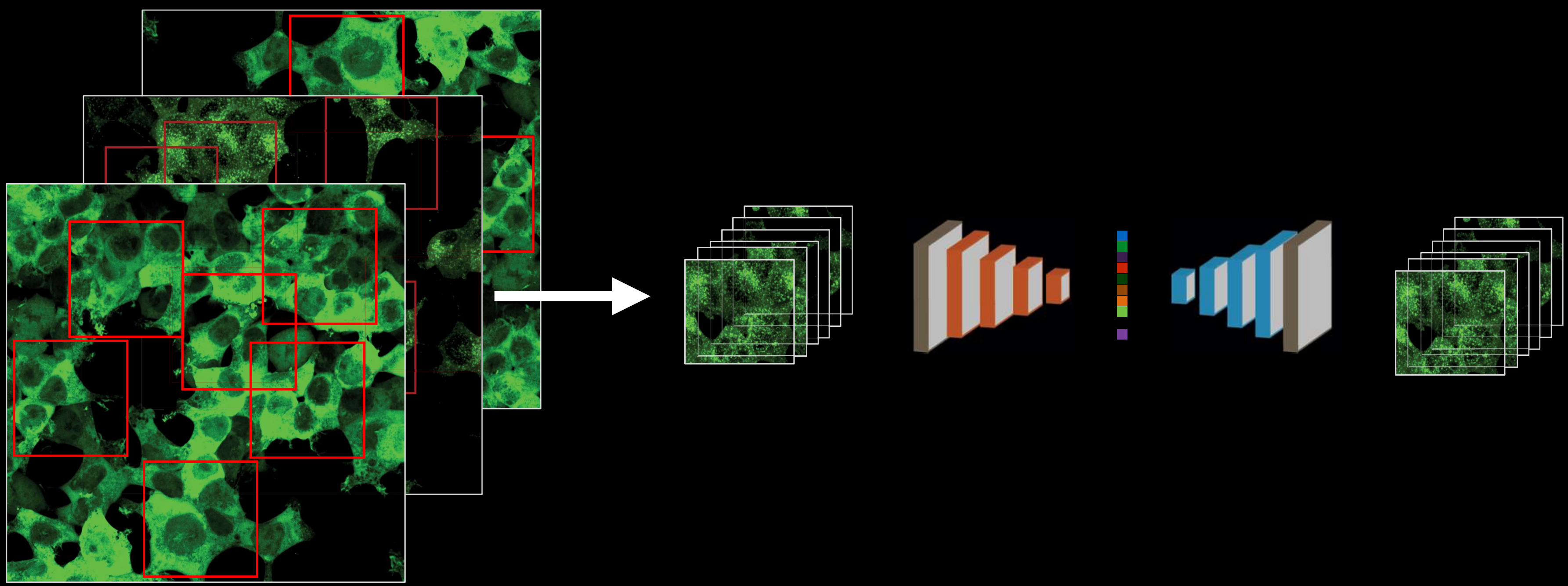
Self-Supervised Deep Learning Encodes High-Resolution Features of Protein Subcellular Localization

Hirofumi Kobayashi, Keith C. Cheveralls, Manuel D. Leonetti, Loic A. Royer

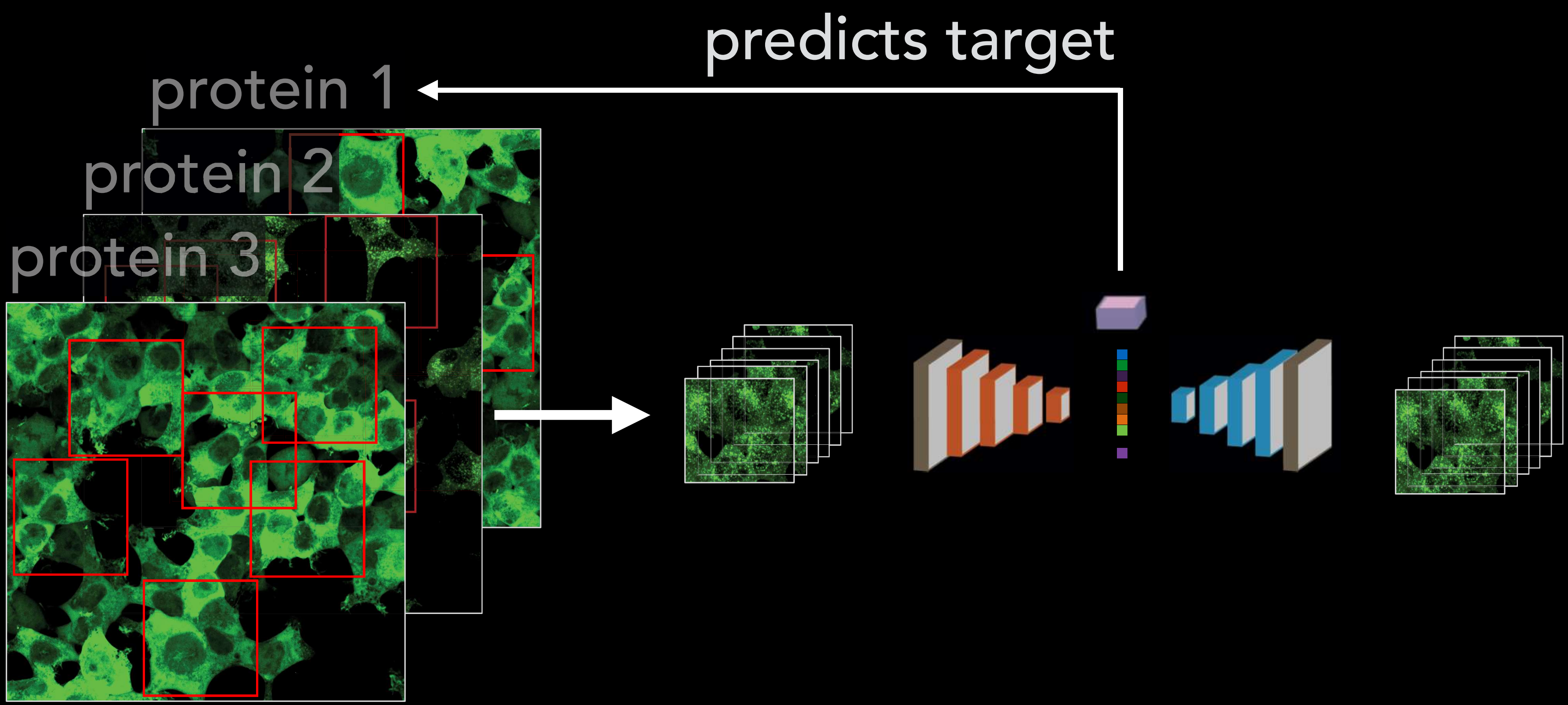
doi: https://doi.org/10.1101/2021.03.29.437595



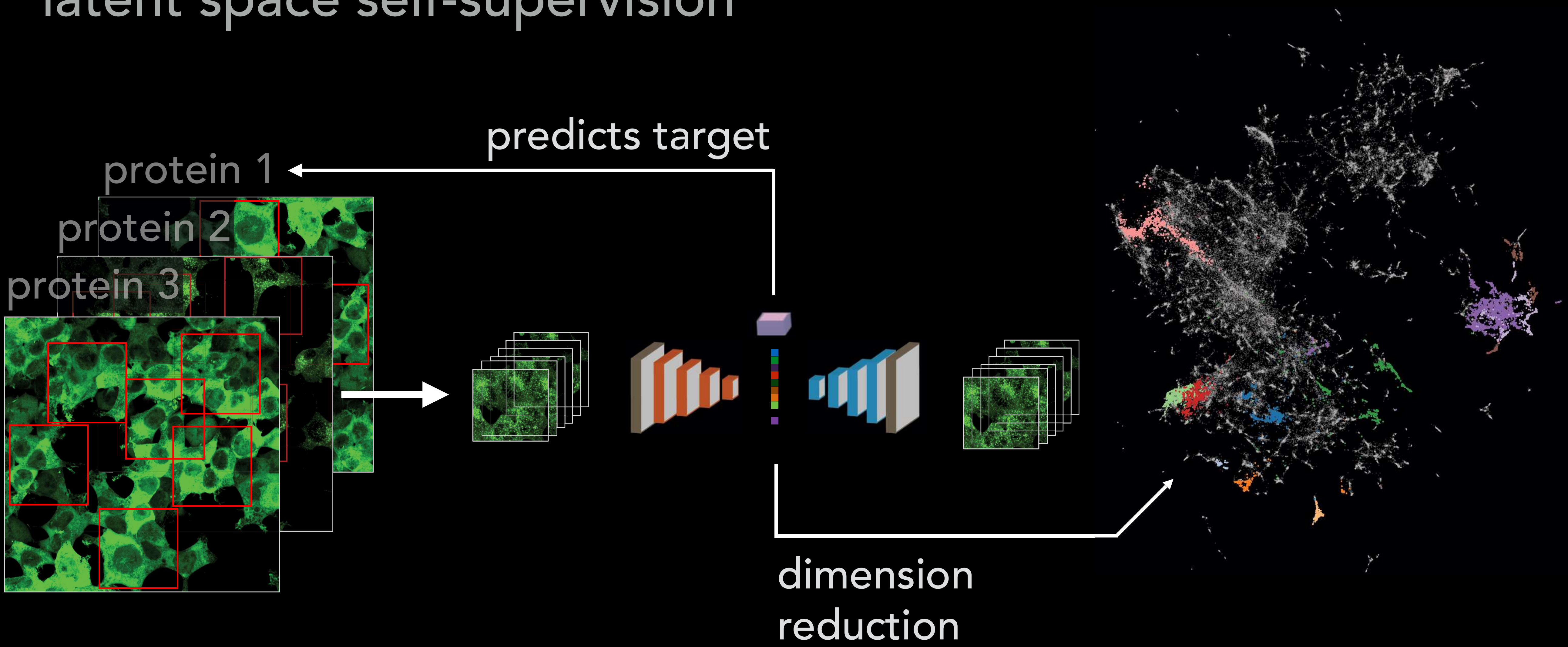
Our Approach
latent space self-supervision



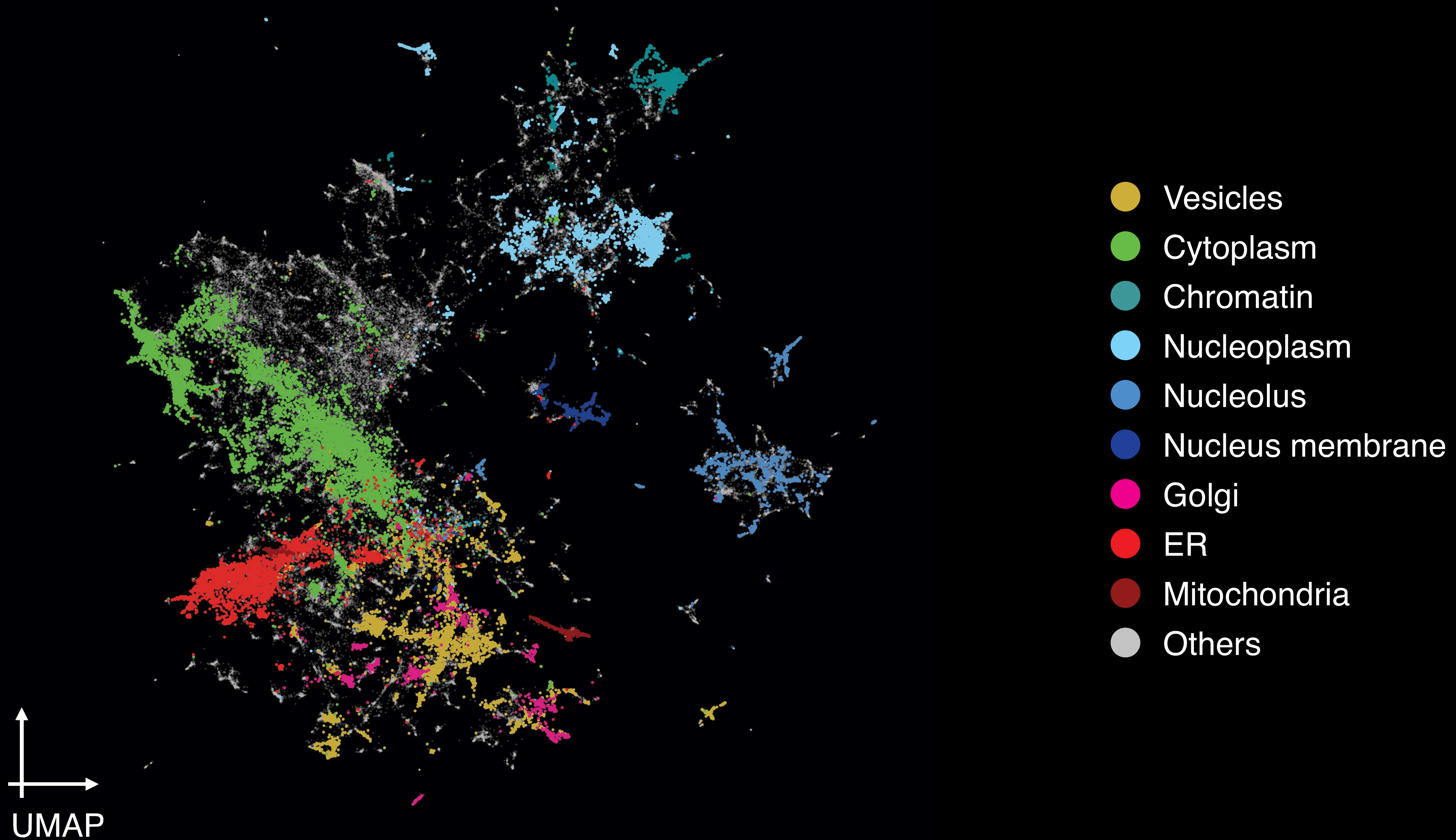
Our Approach
latent space self-supervision

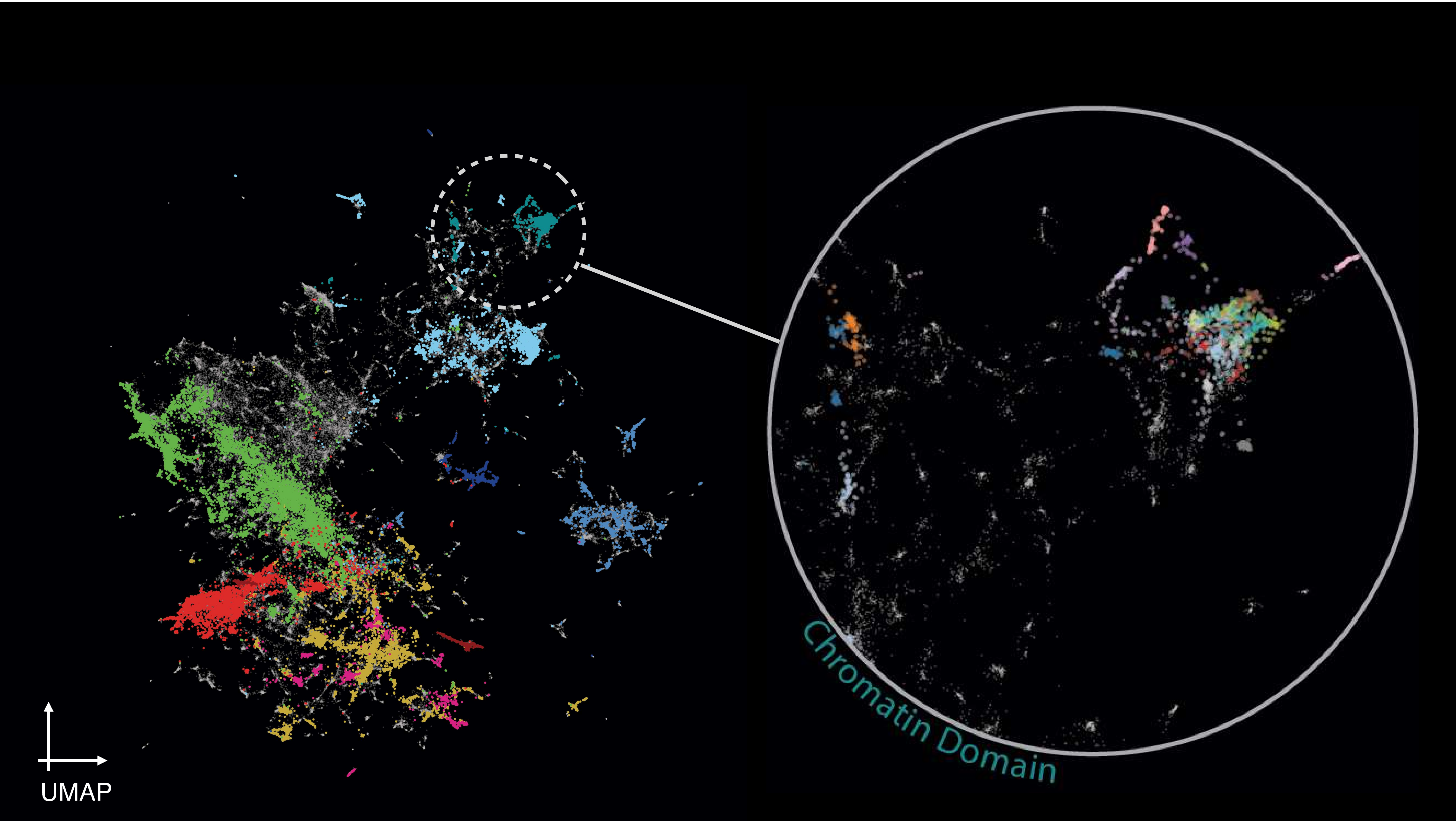


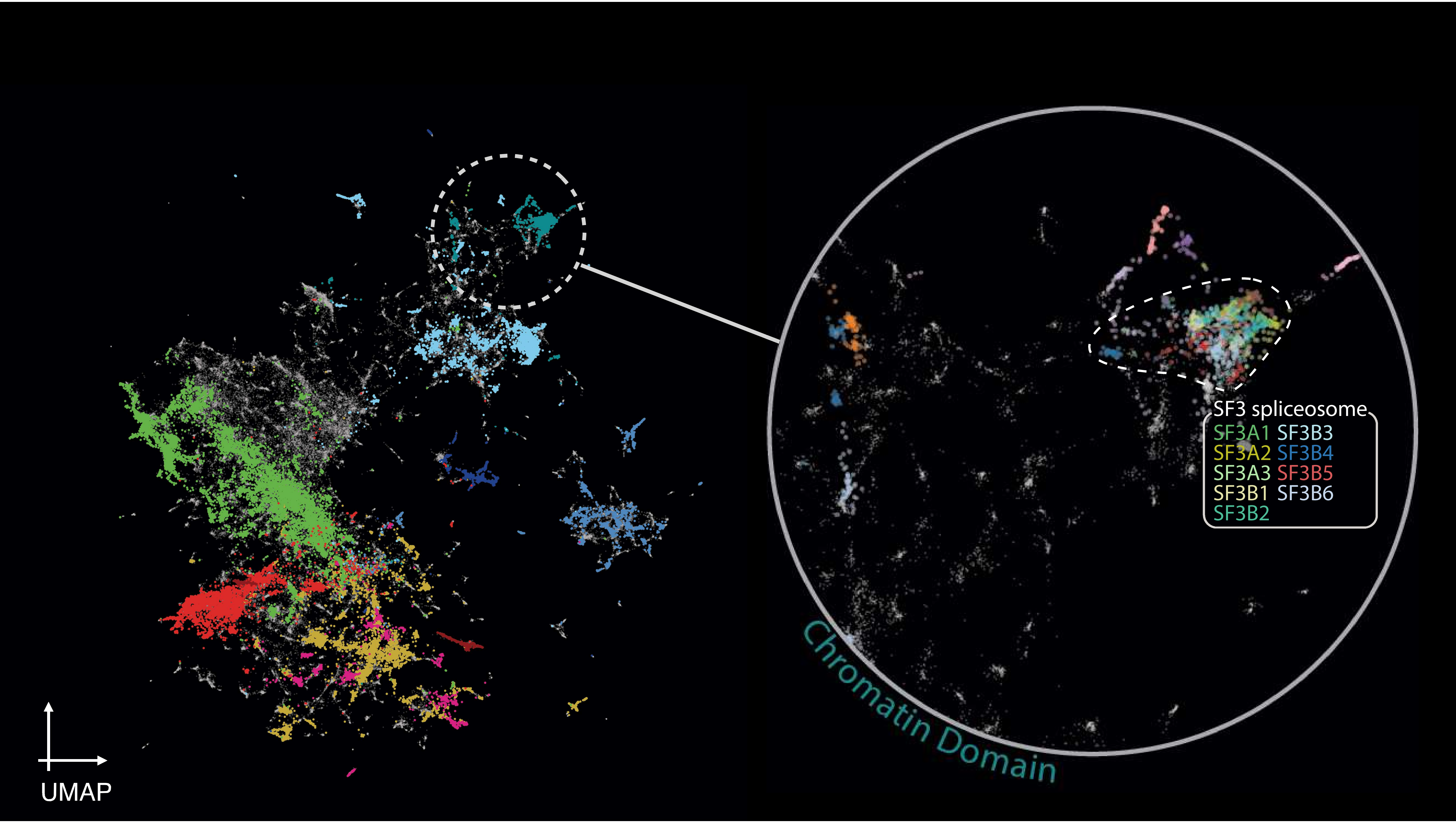
Our Approach
latent space self-supervision



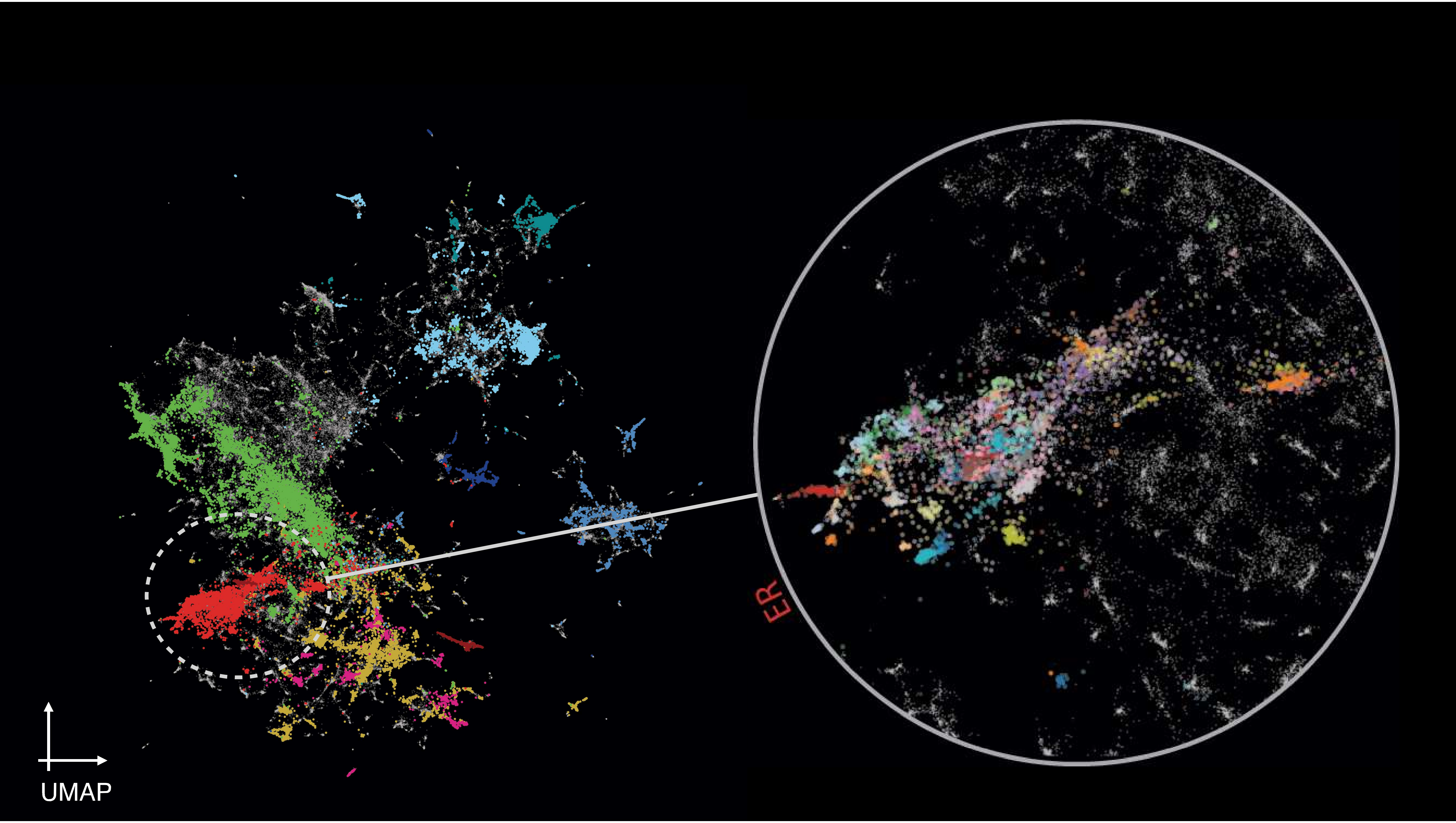
A New Map of Protein Sub-cellular Localisation



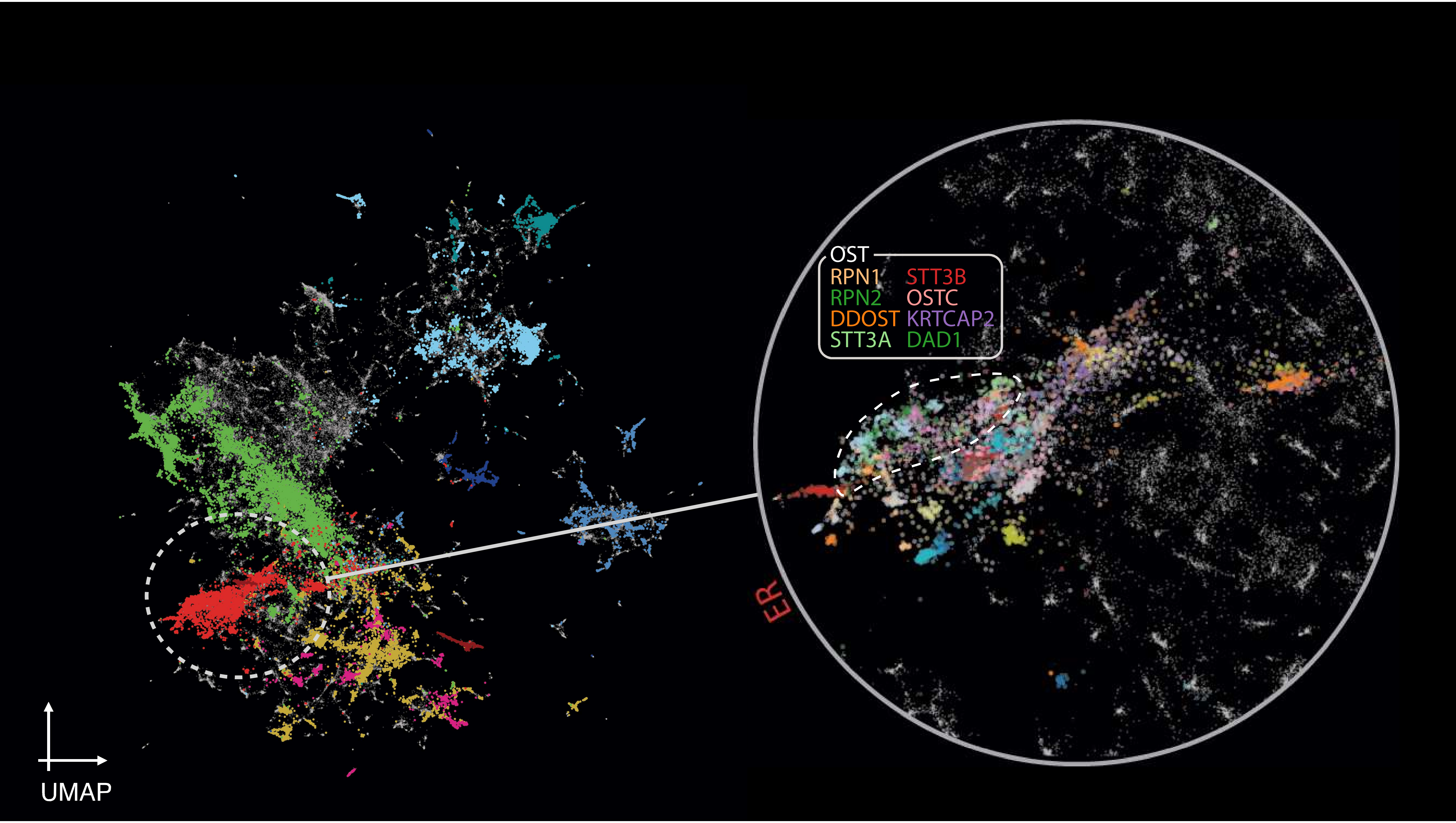

UMAP
Chromatin Domain

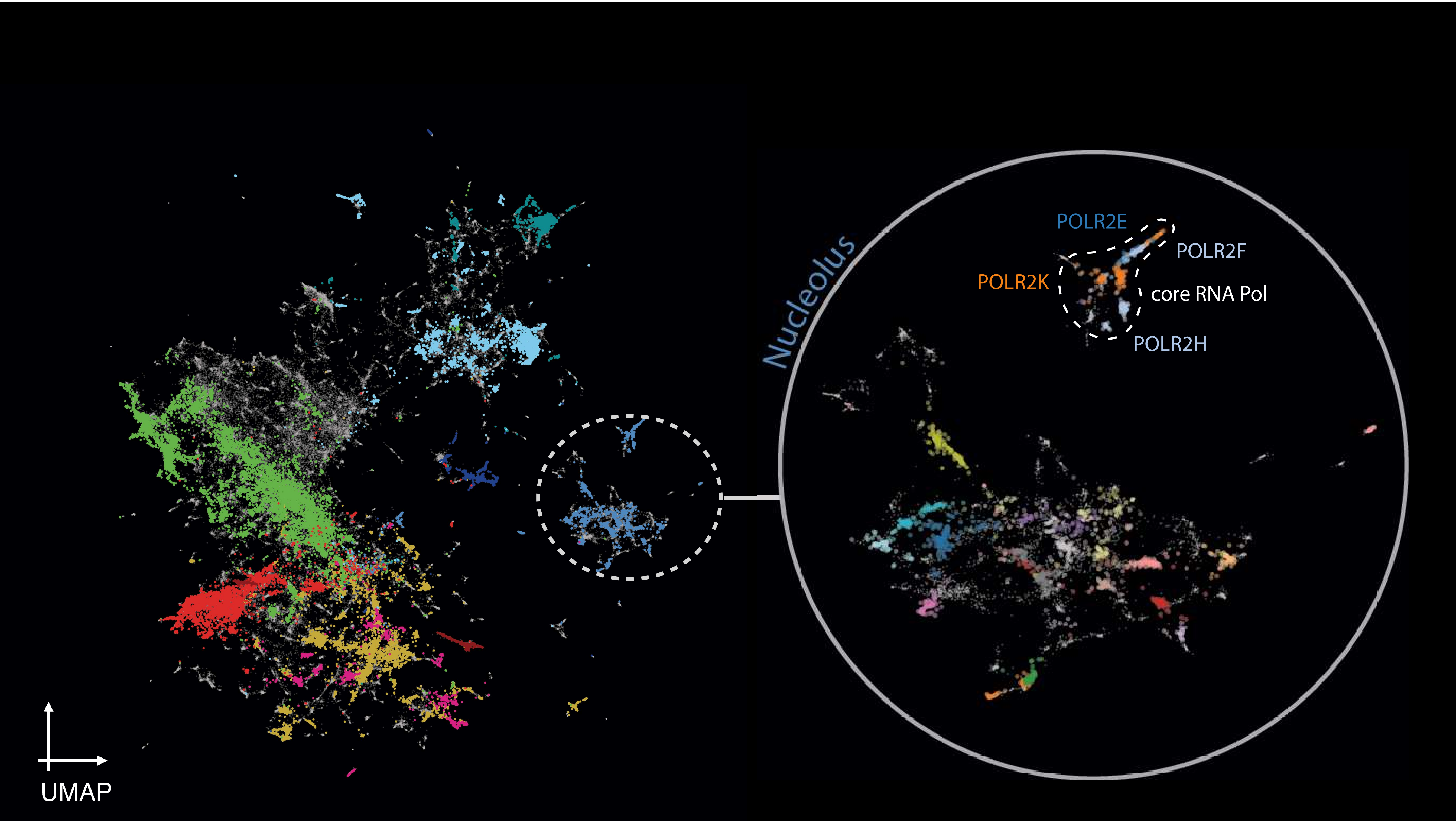


SF3 spliceosome
SF3A1 SF3B3
SF3A2 SF3B4
SF3A3 SF3B5
SF3B1 SF3B6
SF3B2
Chromatin Domain
UMAP



ER
UMAP


OST
RPN1
STT3B
RPN2
OSTC
DDOST
KRTCAP2
STT3A
DAD1
ER
UMAP

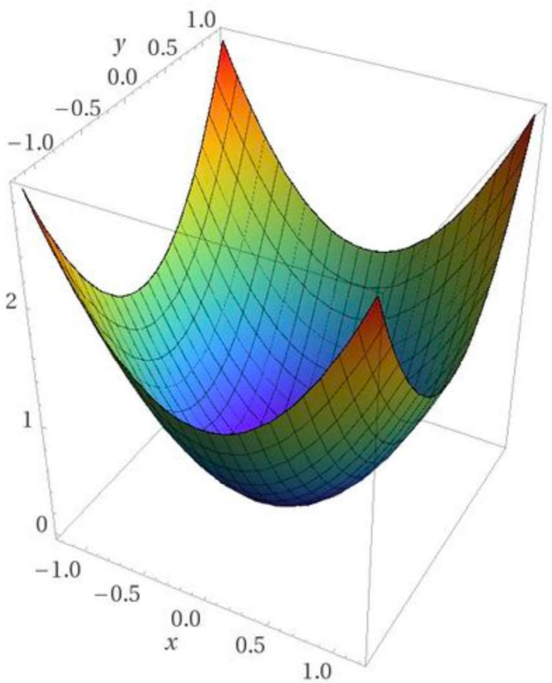


Nucleolus
POLR2E
POLR2F
POLR2K
core RNA Pol
POLR2H
UMAP

The Training Problem

The Complex Objective (Loss) of Deep Models

a different output every time you train…

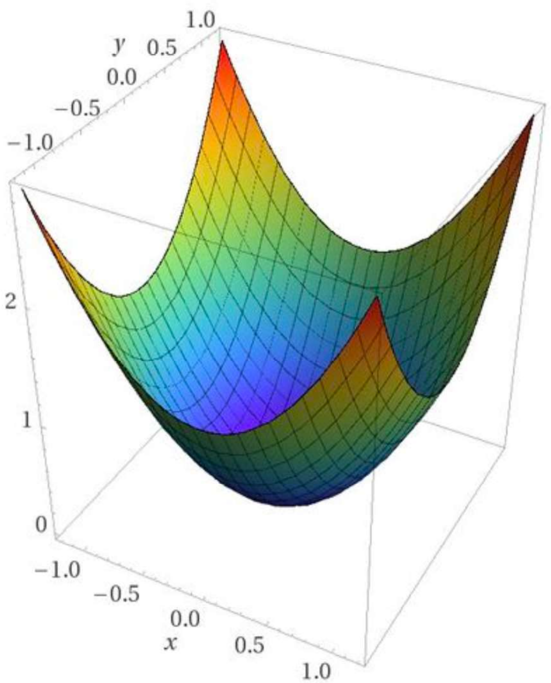


convex objective

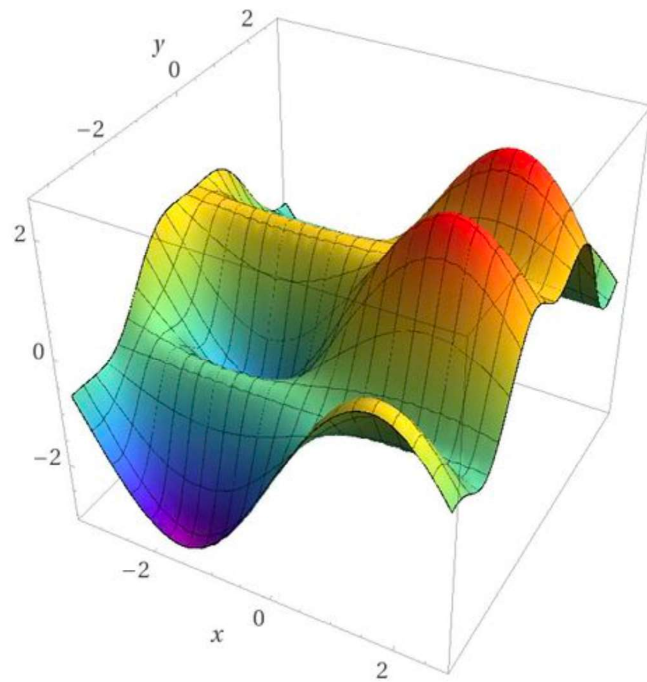
Ashwin Singh

The Complex Objective (Loss) of Deep Models

a different output every time you train…



convex objective

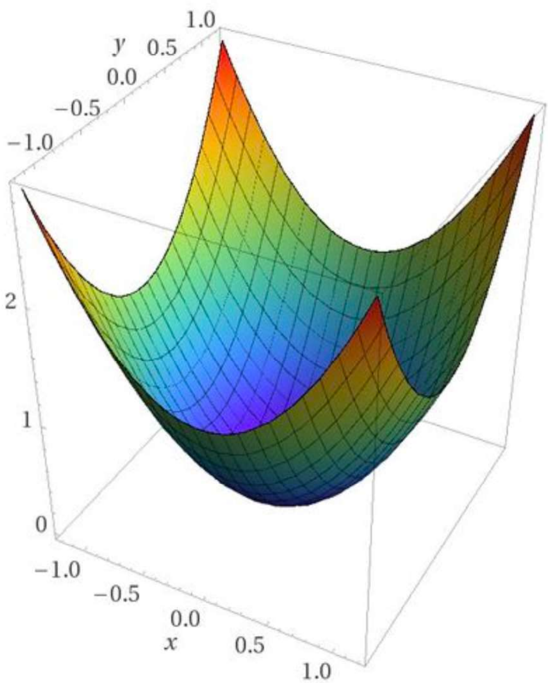


complex objective

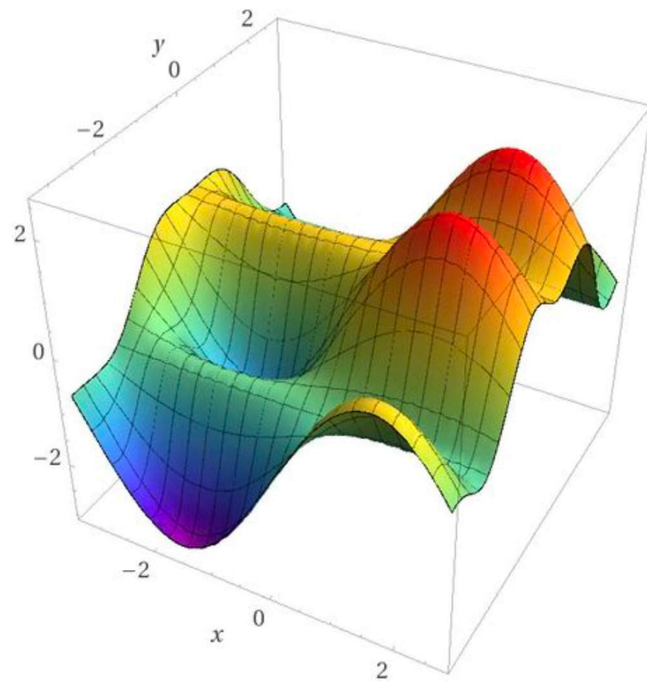
Ashwin Singh

The Complex Objective (Loss) of Deep Models

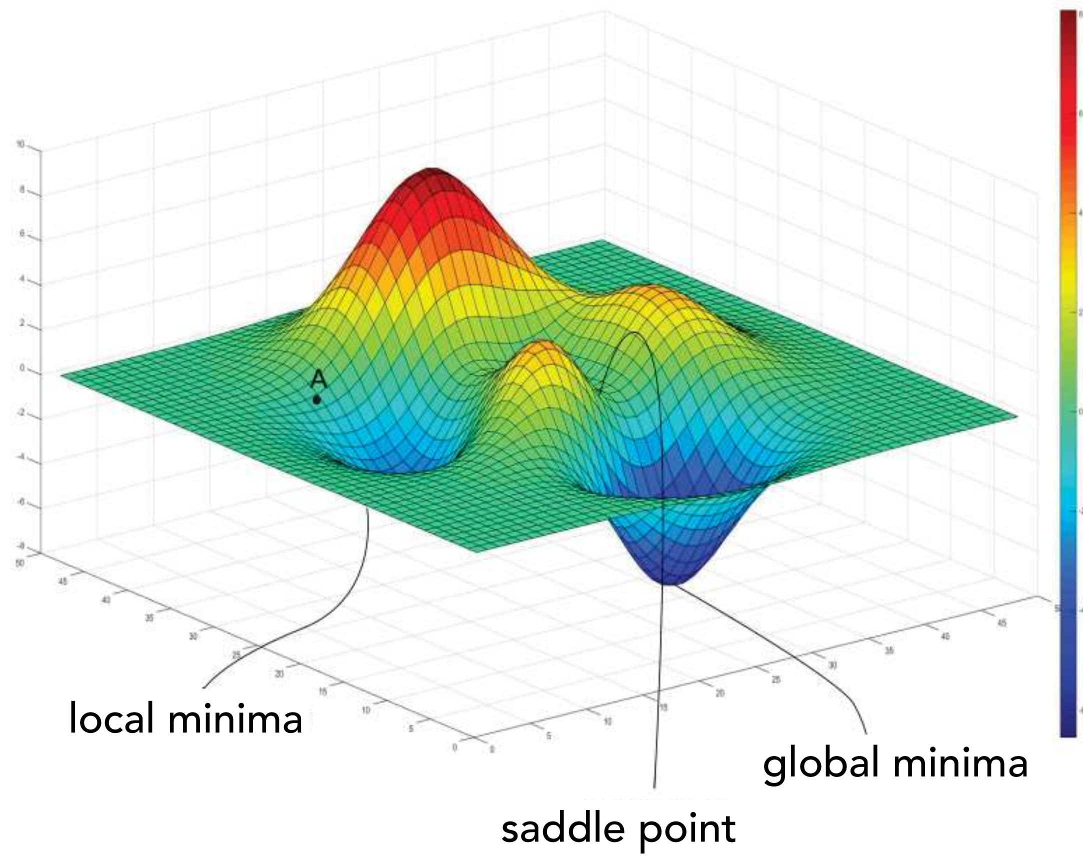
a different output every time you train…



convex objective



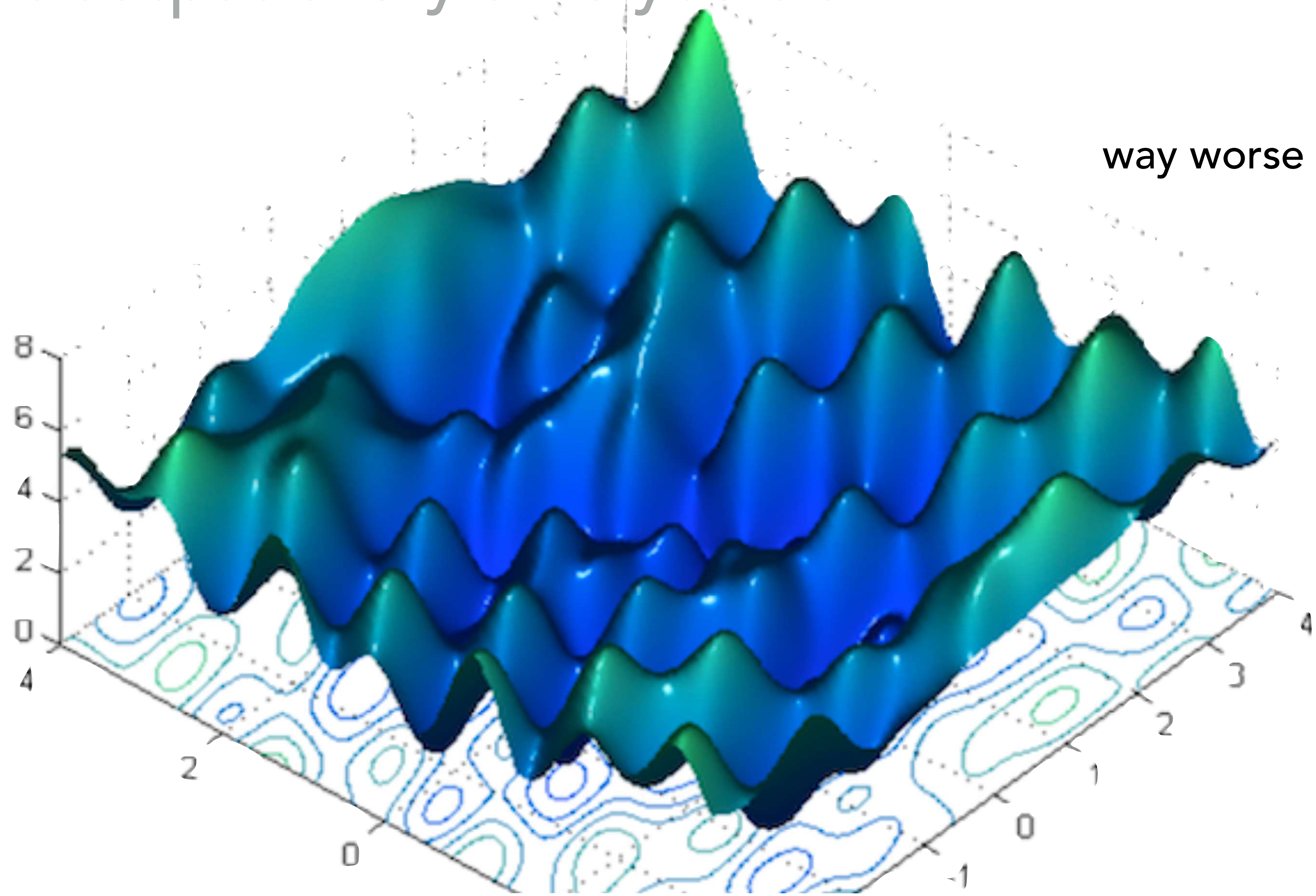
complex objective



Ashwin Singh

The Complex Objective (Loss) of Deep Models

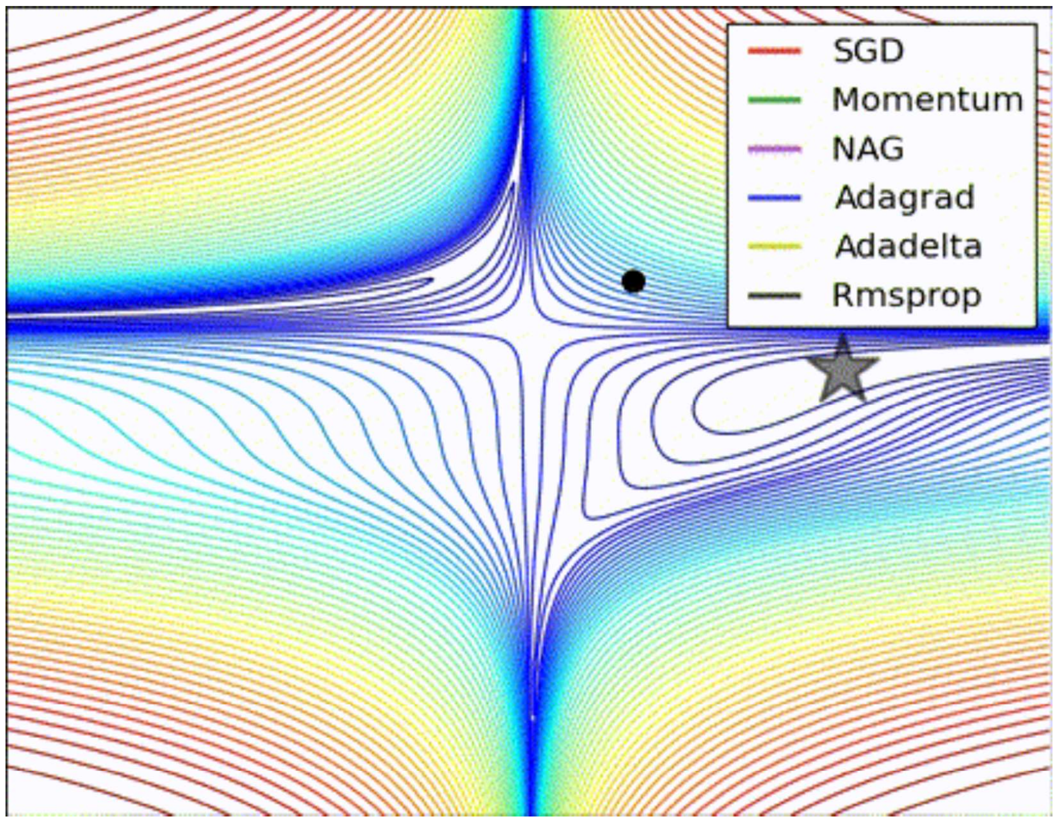
a different output every time you train…

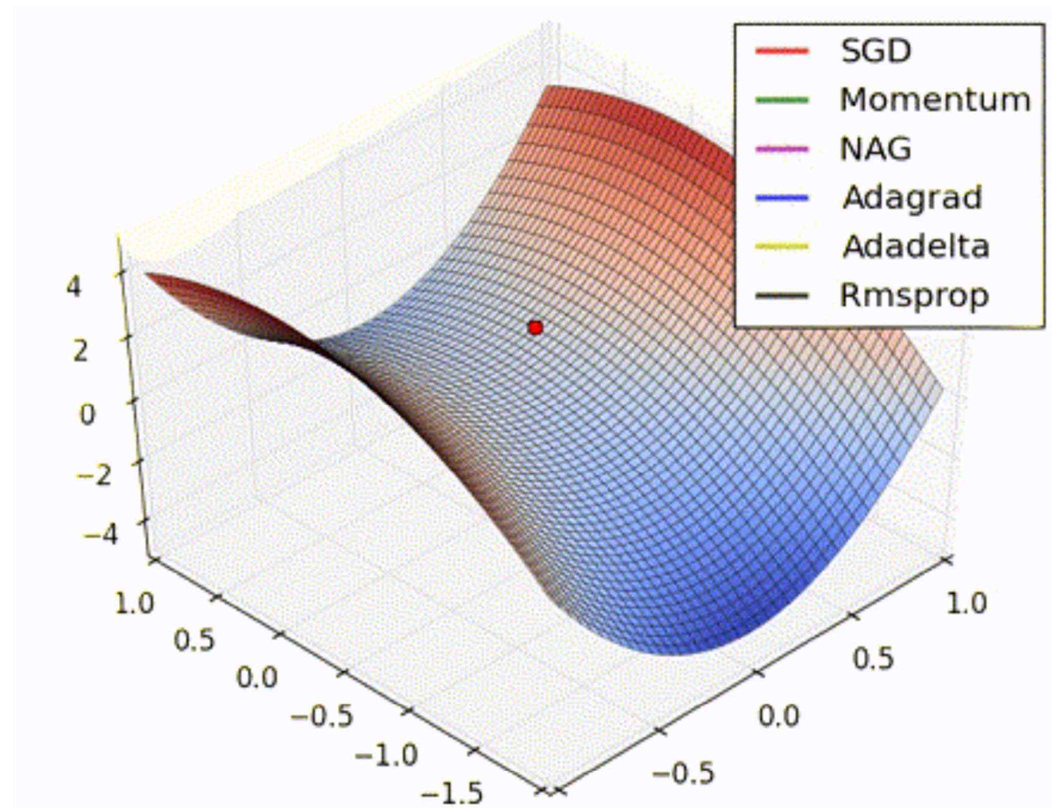


way worse than that…

The Black Magic of Optimisation

How to choose algorithm and hyper-parameters?

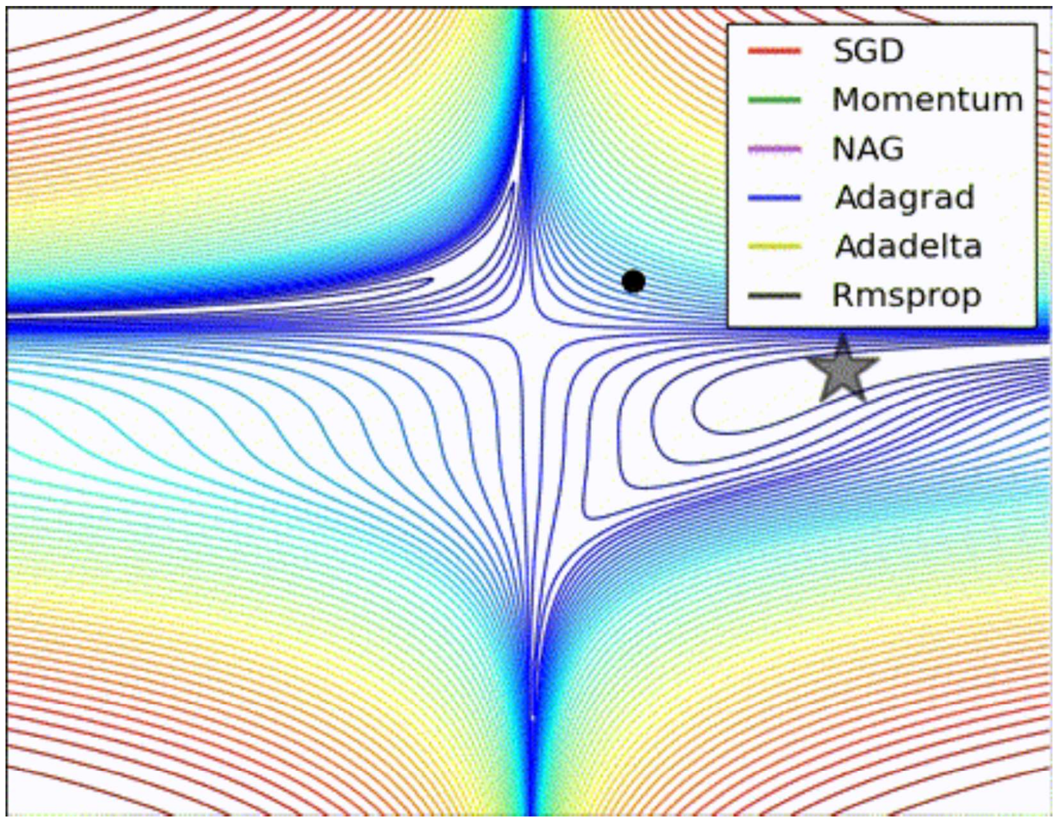


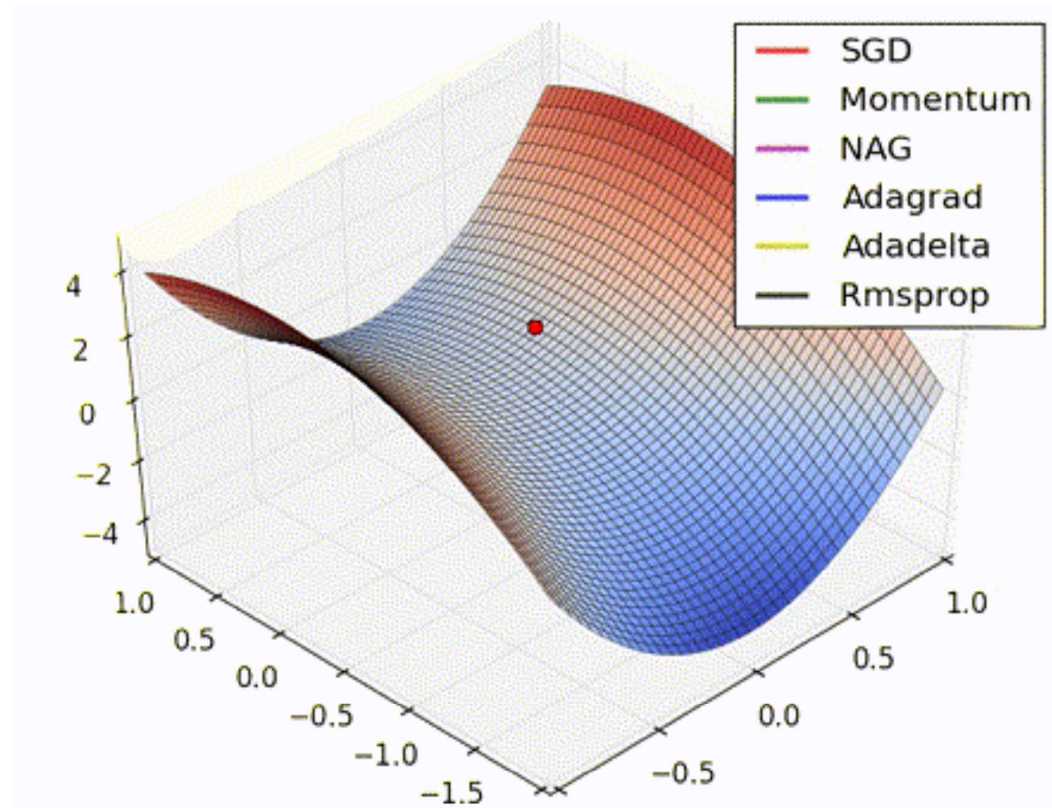


Alec Radford (twitter.com/alecrad)

The Black Magic of Optimisation

How to choose algorithm and hyper-parameters?

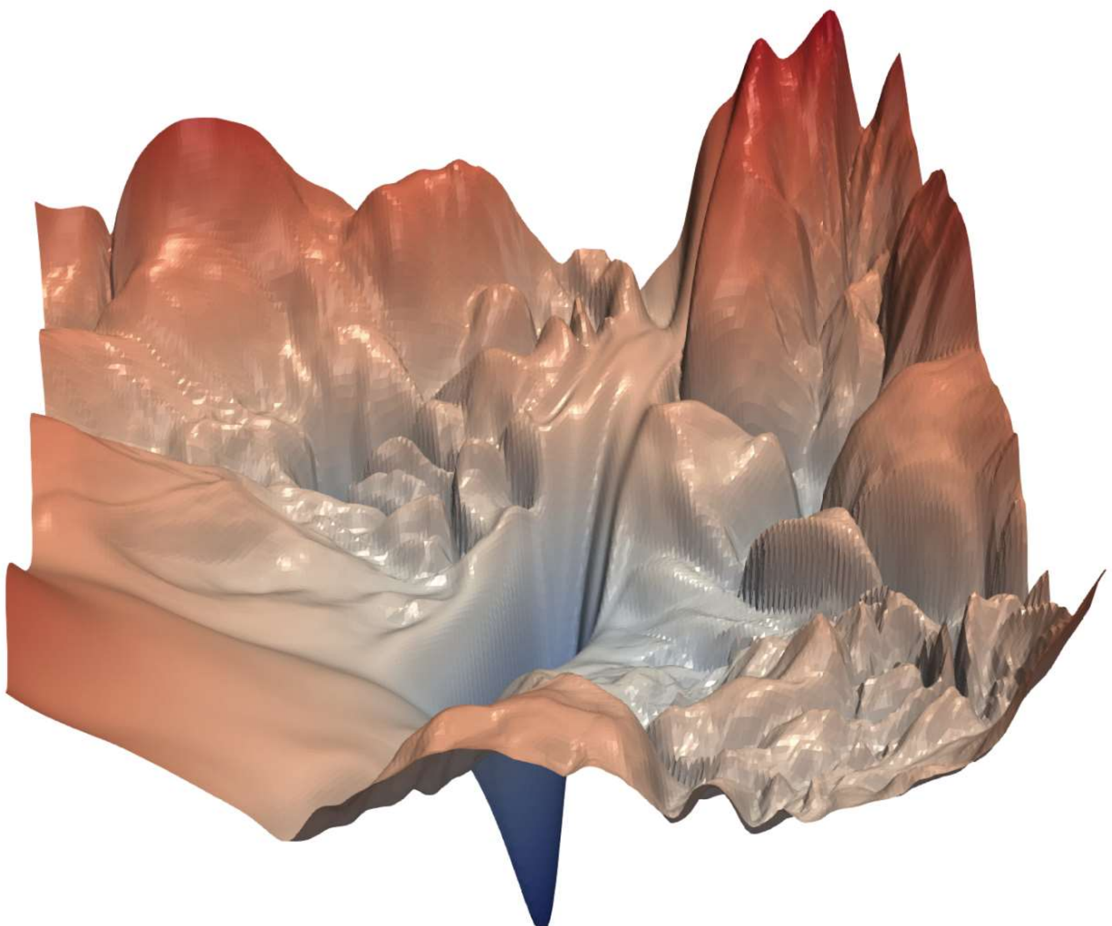




Alec Radford (twitter.com/alecrad)

The Art of Neural Architecture Design

and the loss landscape of neural networks

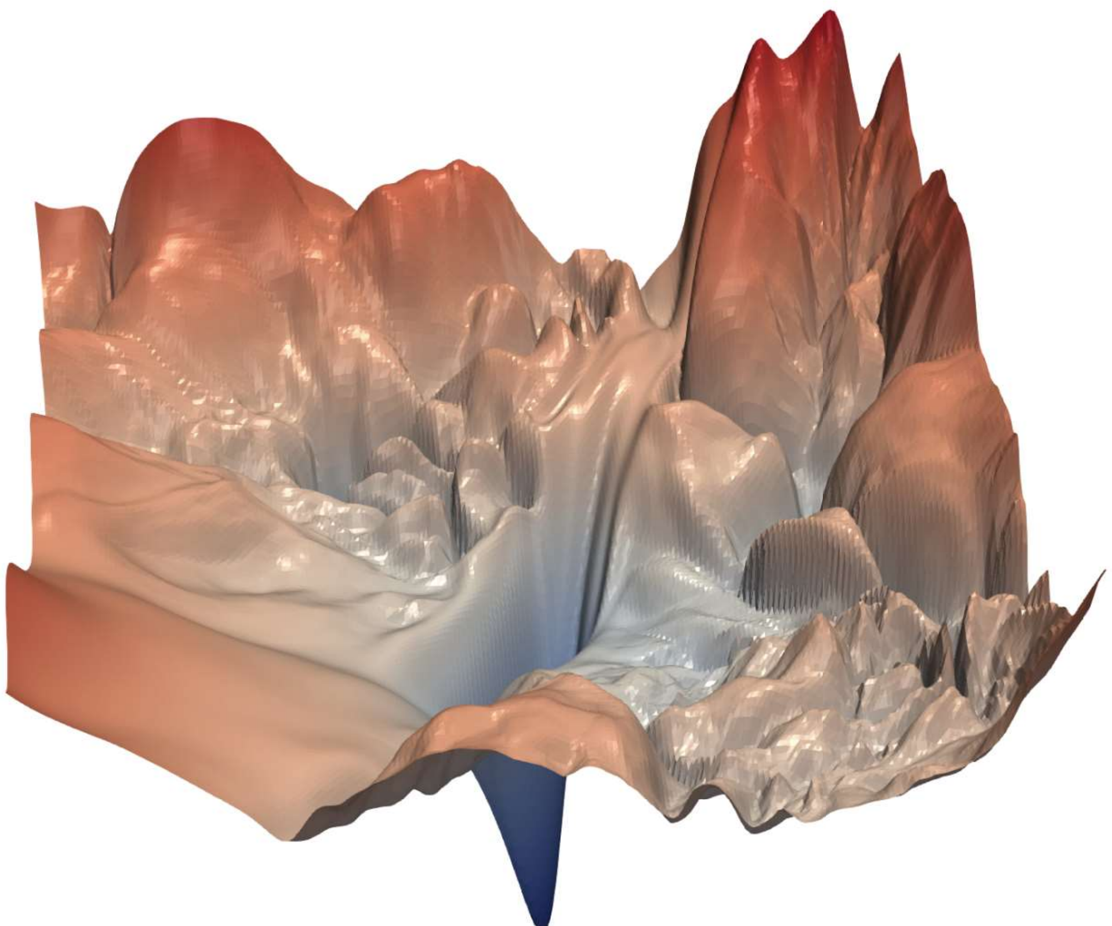


without skip connections

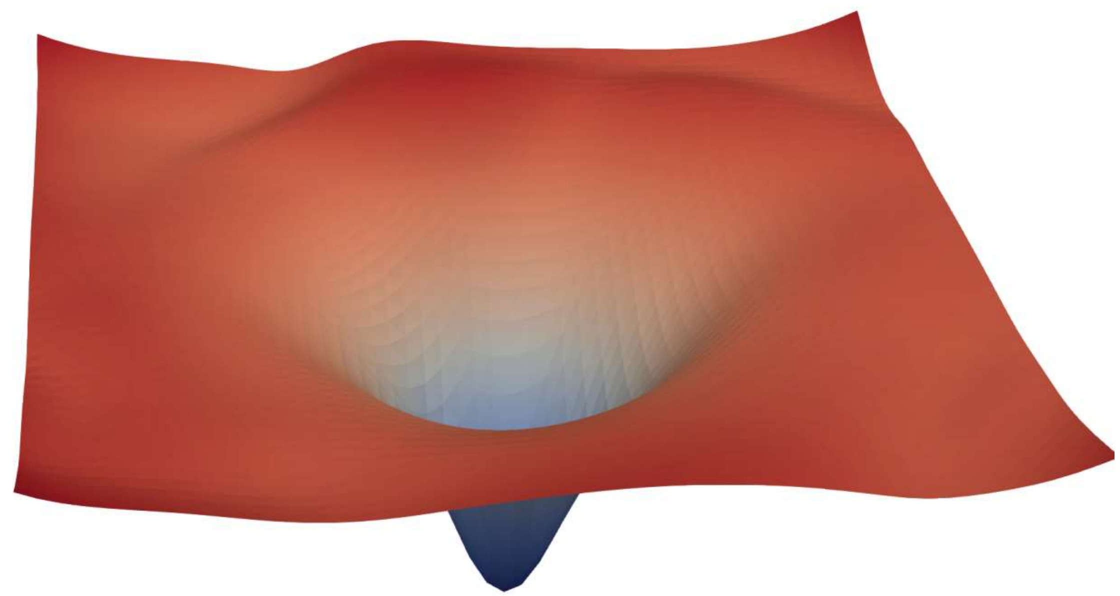
Li et al. NeurIPS 2018

The Art of Neural Architecture Design

and the loss landscape of neural networks



without skip connections



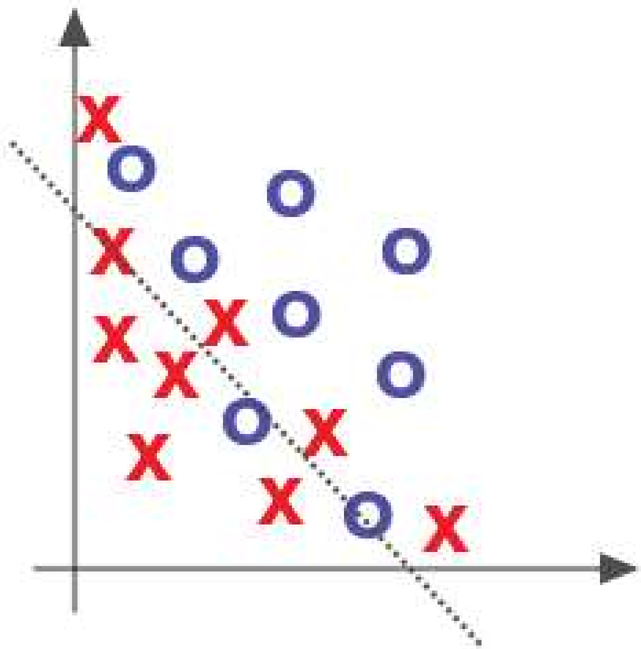
with skip connections

Li et al. NeurIPS 2018

The Overfitting Problem

Under- and Overfitting

intuition

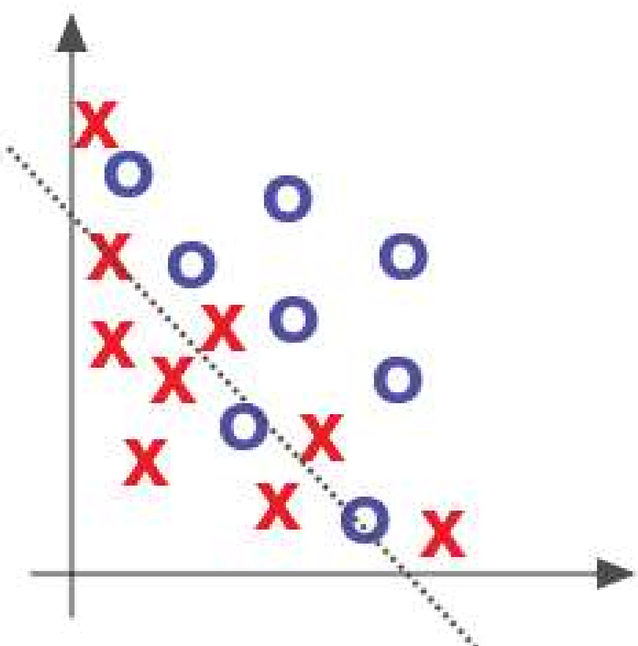


under-fitting

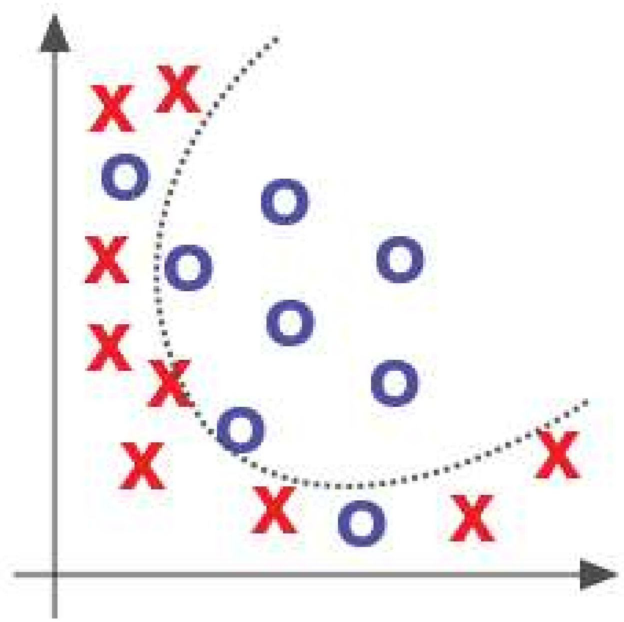
Deep Learning, Gibson & Patterson

Under- and Overfitting

intuition



under-fitting

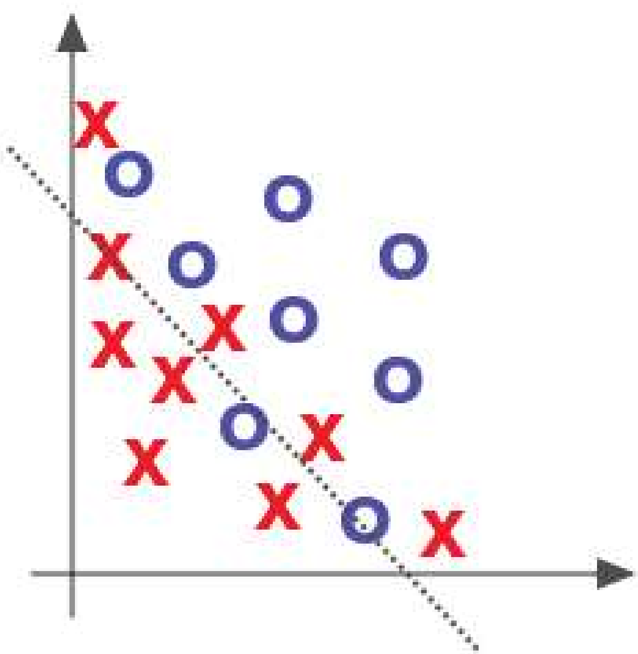


desired

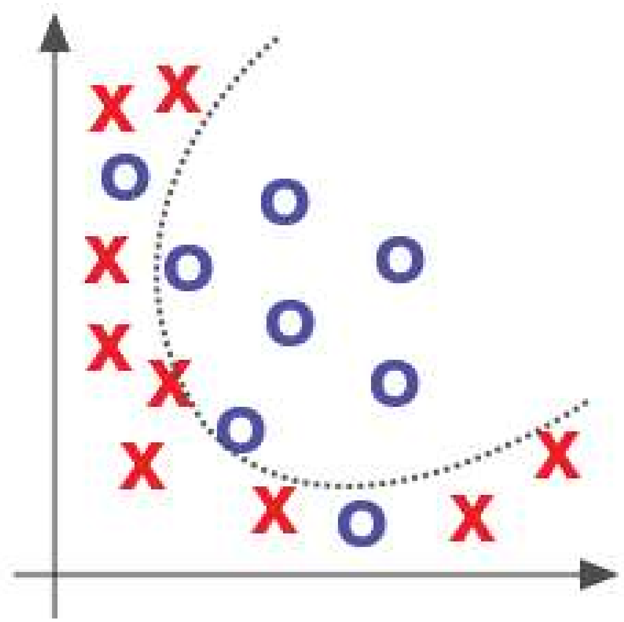
Deep Learning, Gibson & Patterson

Under- and Overfitting

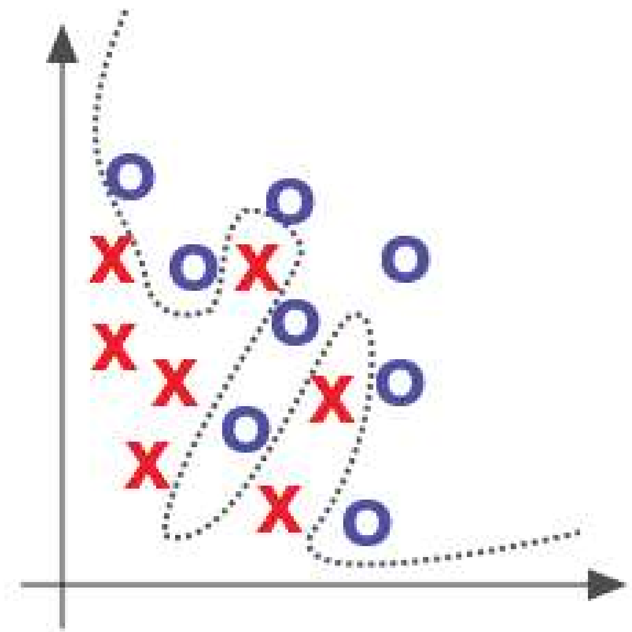
intuition



under-fitting



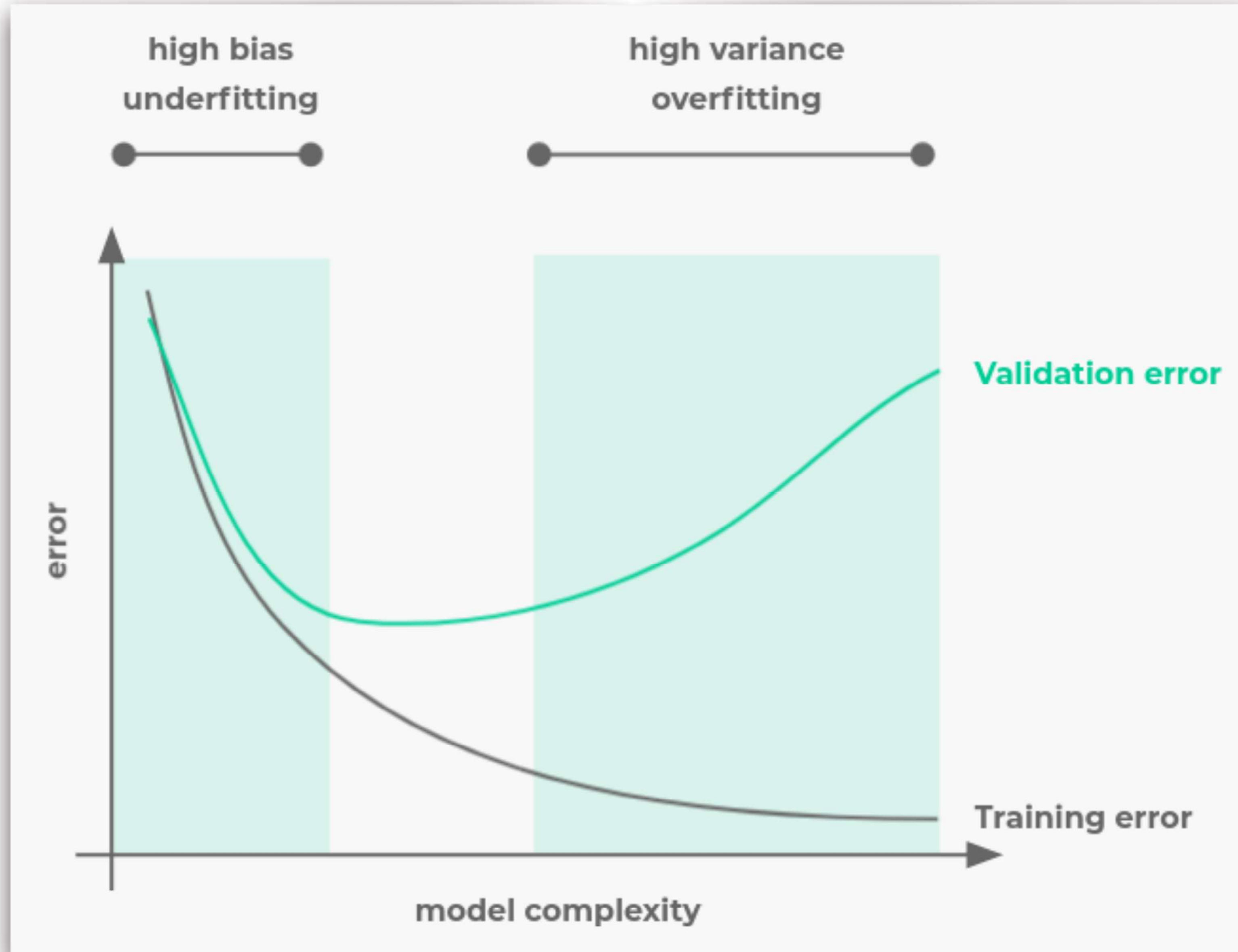
desired



over-fitting

Deep Learning, Gibson & Patterson

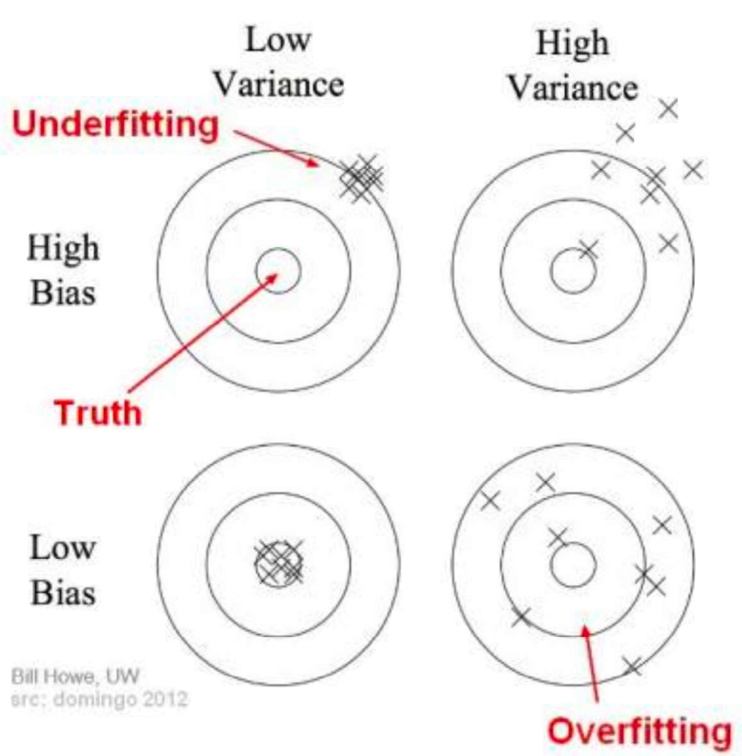
Under- and Overfitting
bias versus variance

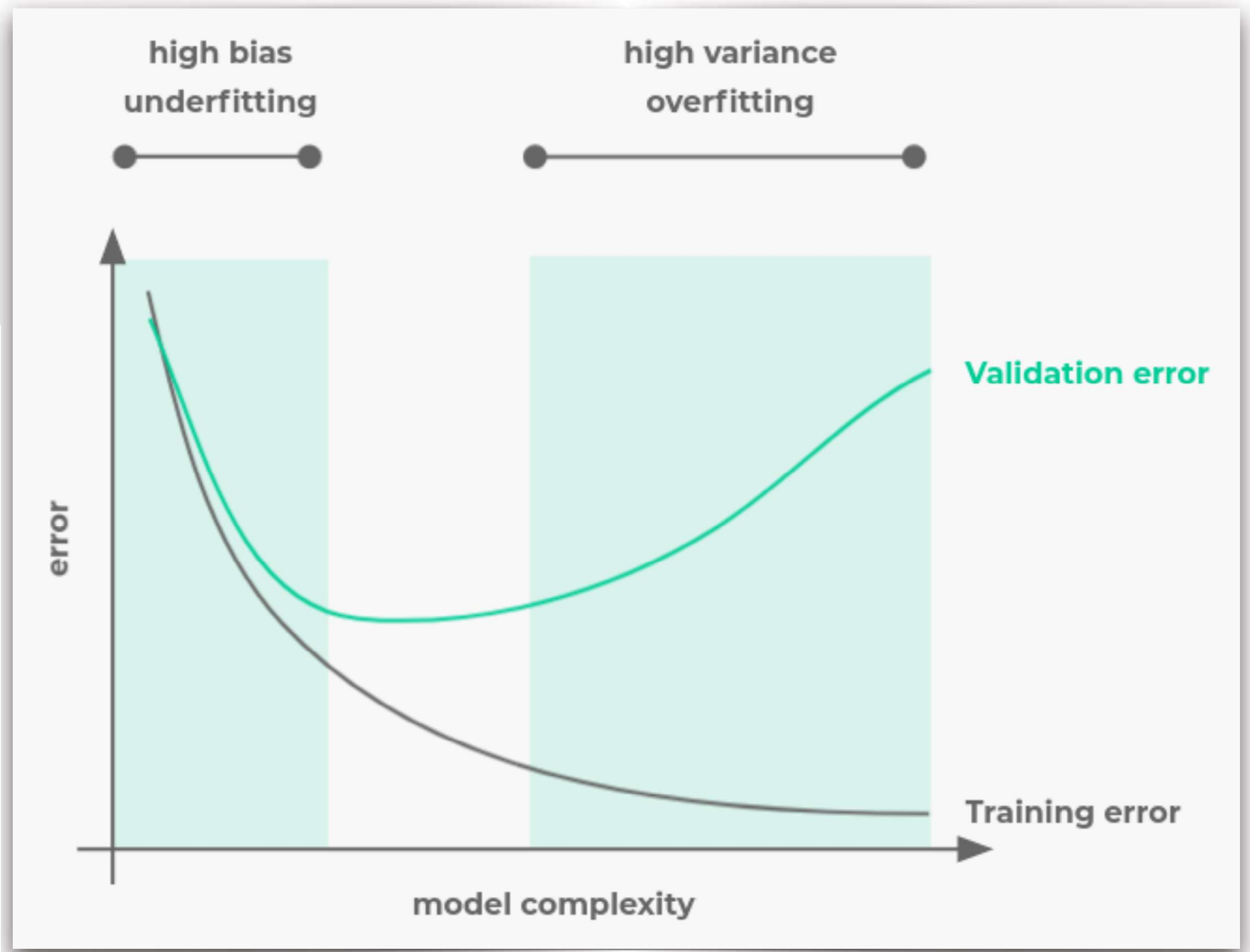


Quantdare, Pablo Sánchez

Under- and Overfitting

bias versus variance





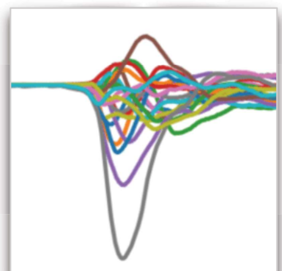
Quantdare, Pablo Sánchez

Overfitting

solutions

$$\lambda \| W \|_1$$

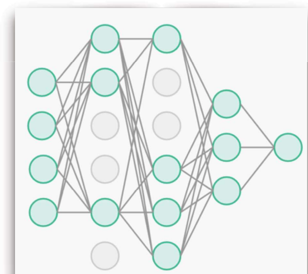
Weight Regularisation



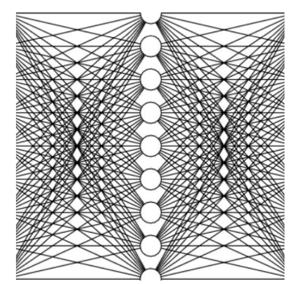
Activity Regularisation

Data Augmentation

Early Stopping



Dropout

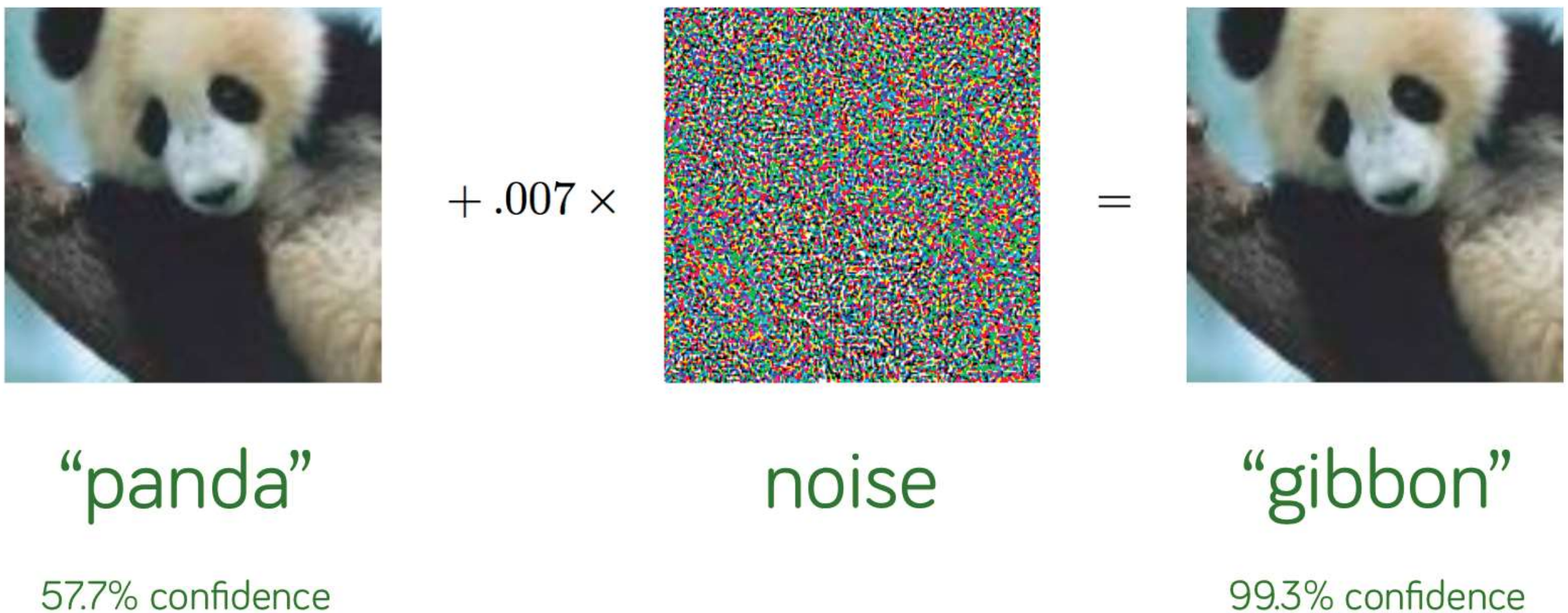


Simplify Model

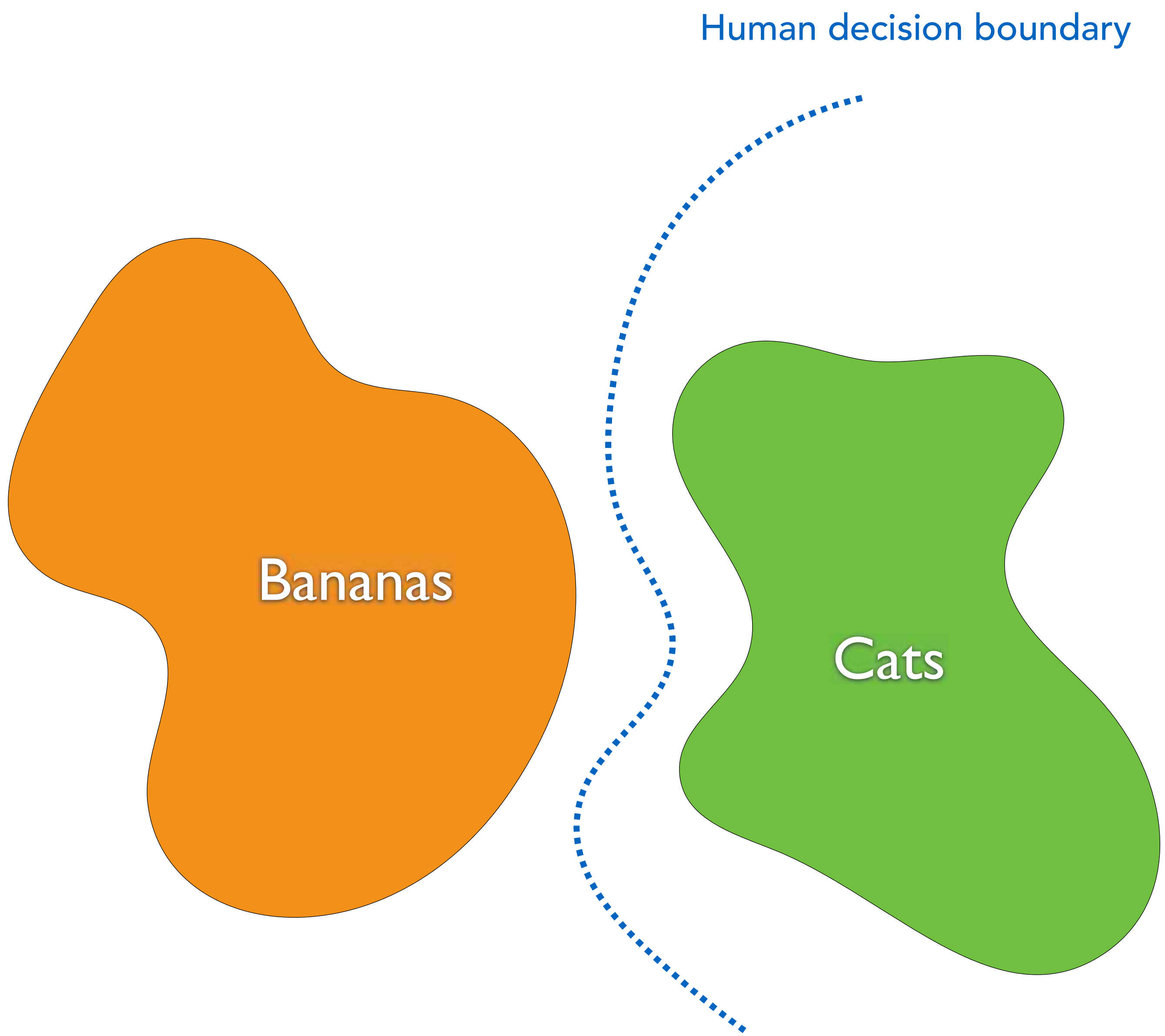
The Adversarial Fragility Problem

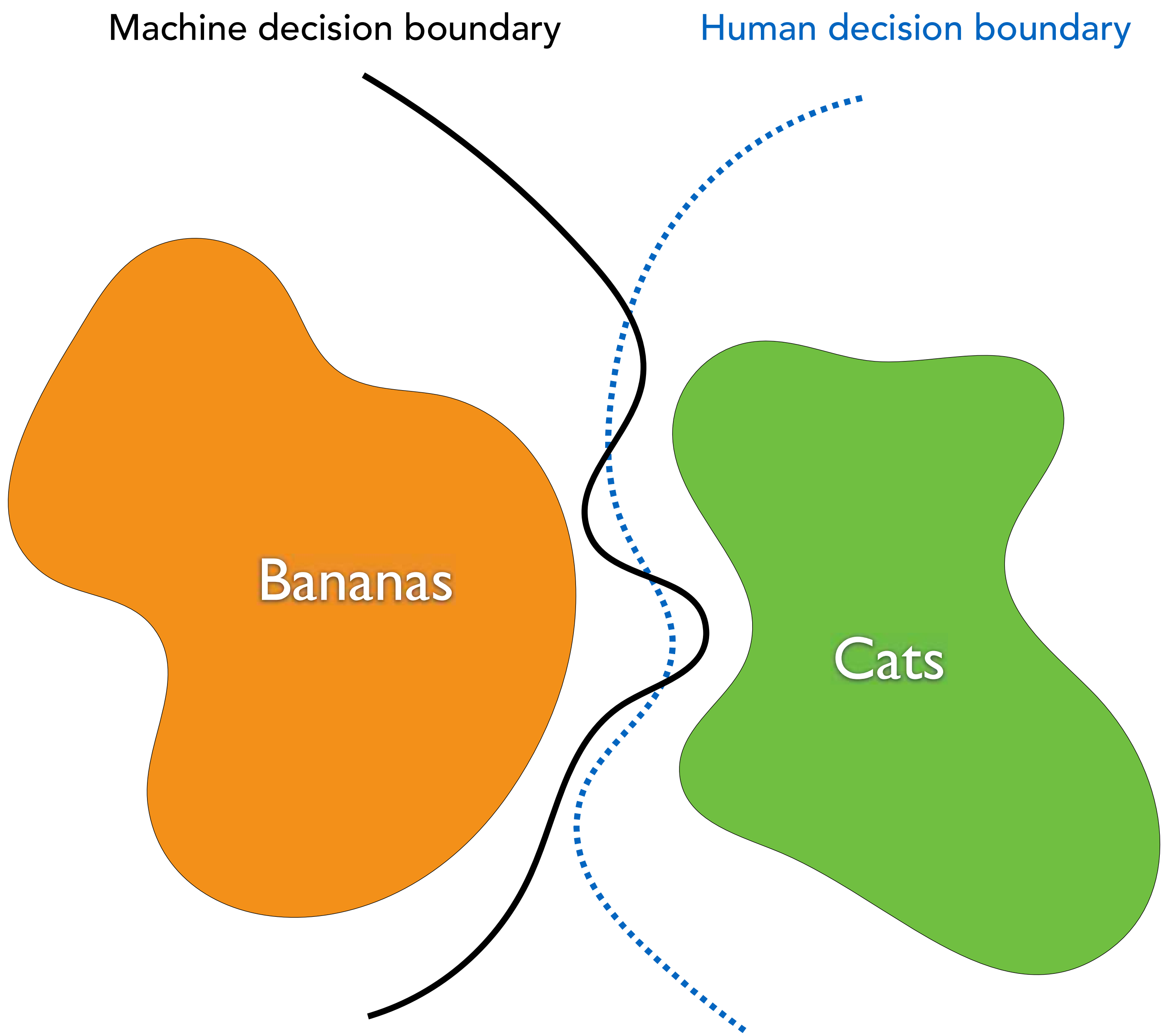
Adversarial Attacks

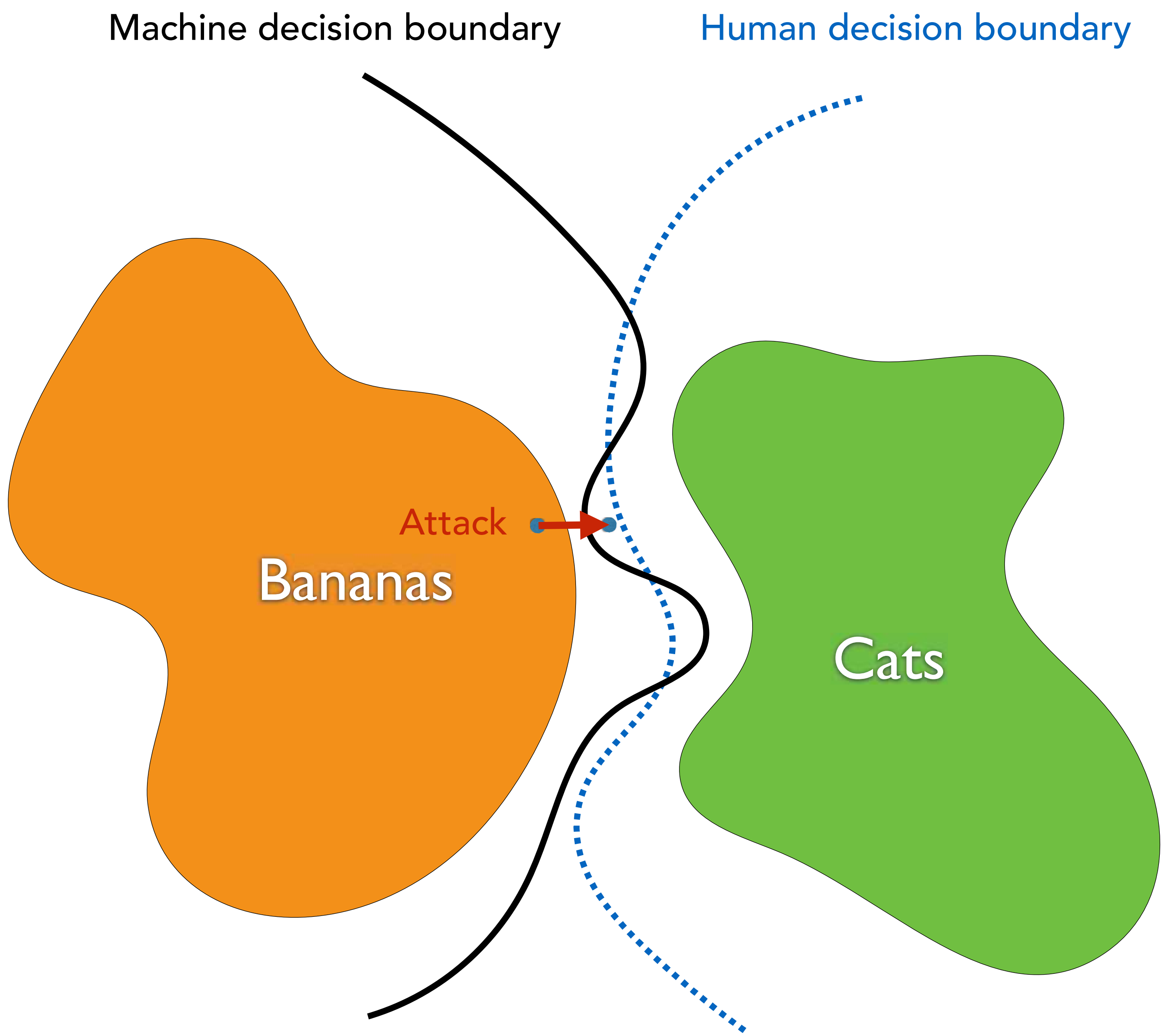
imperceptible noise turns panda into gibbon



Goodfellow et al, ICLR 2015

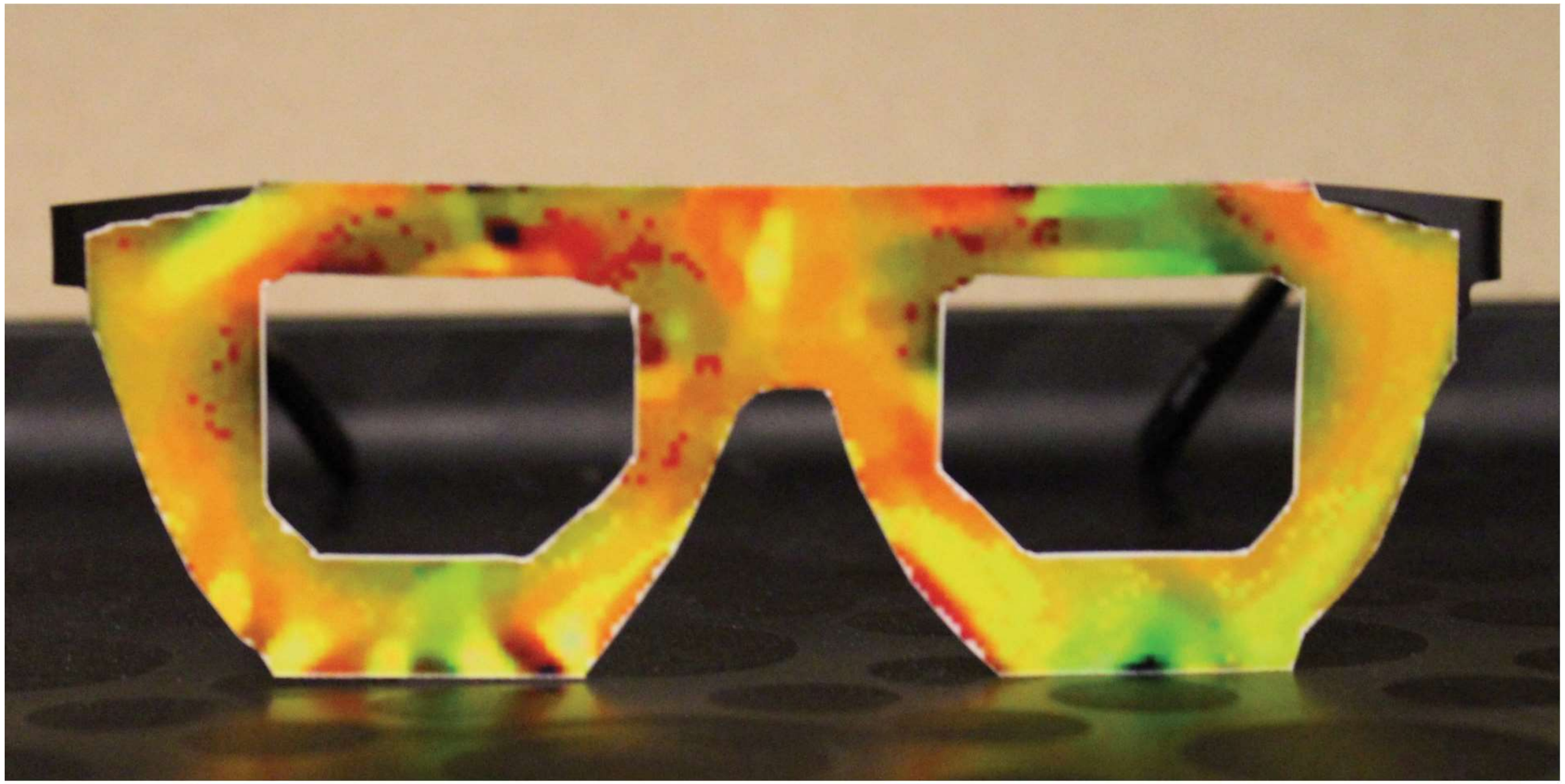

Human decision boundary
Bananas
Cats


Machine decision boundary
Human decision boundary
Bananas
Cats


Machine decision boundary
Human decision boundary
Attack
Bananas
Cats

Adversarial Attacks

fooling face recognition with ugly glasses



Sharif et al. CCS 2016

Adversarial Attacks

this is not the stop sign you have been looking for

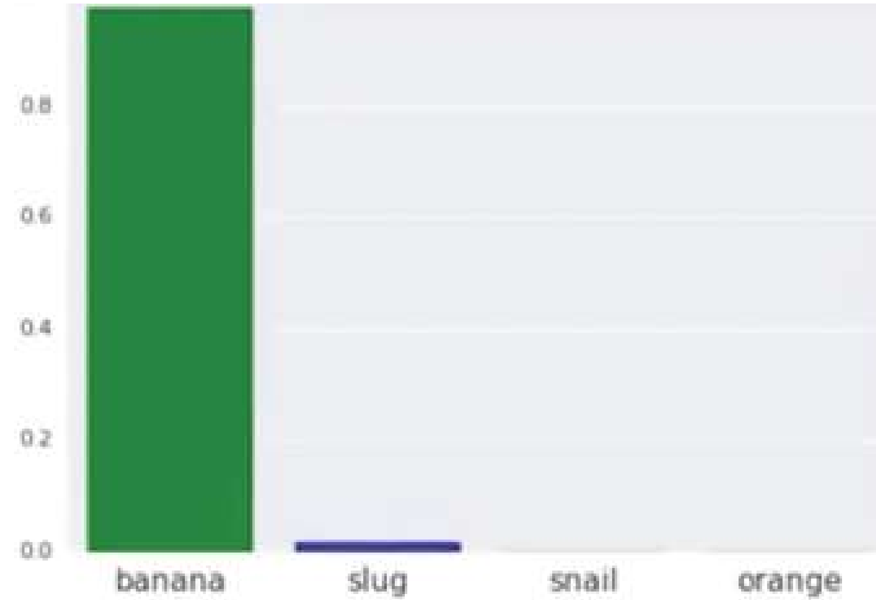
Eykholt et al. CVPR 2018

classifier input

Brown et al. NeurIPS 2017

classifier input

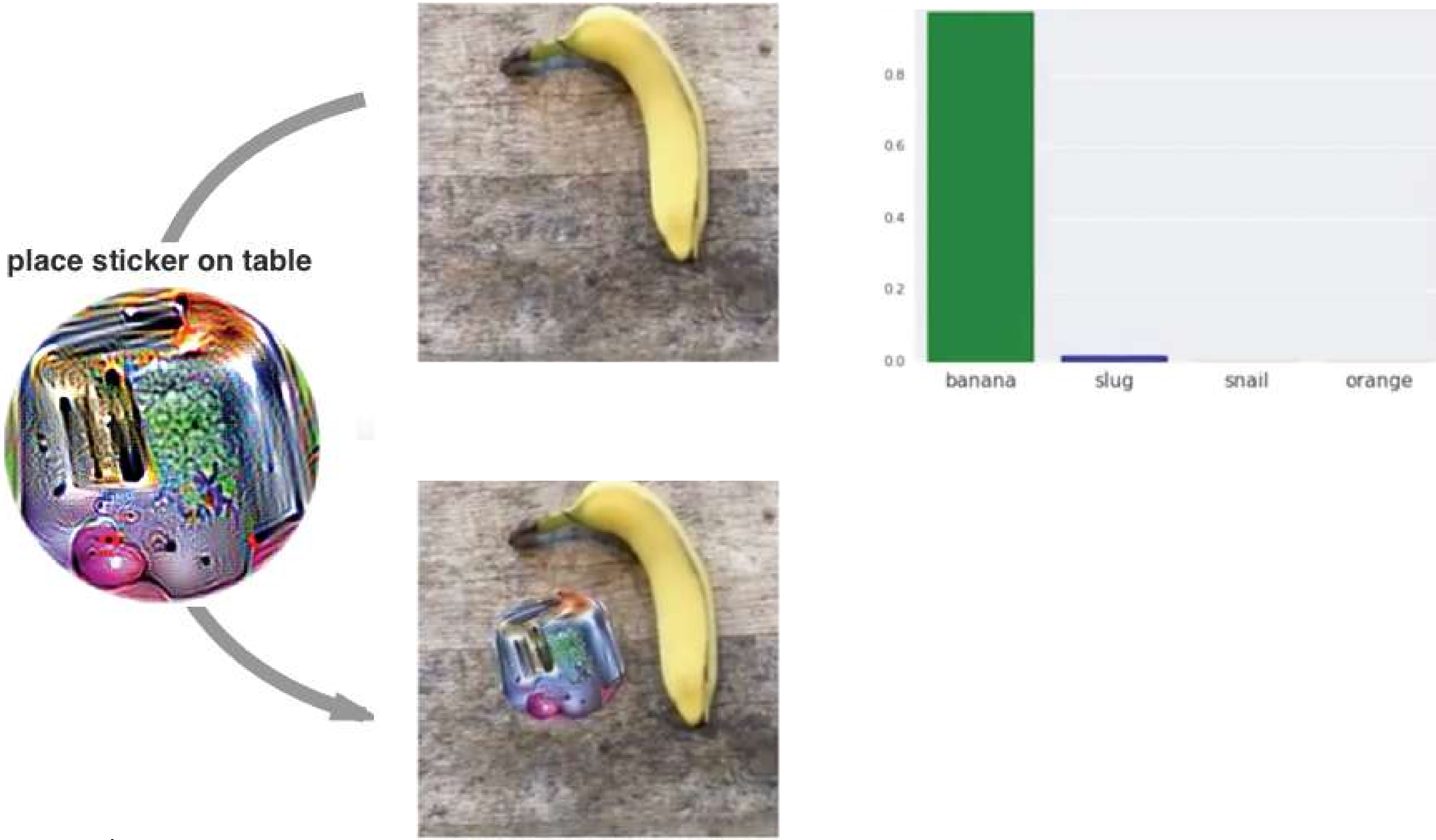
classifier output



Brown et al. NeurIPS 2017

classifier input

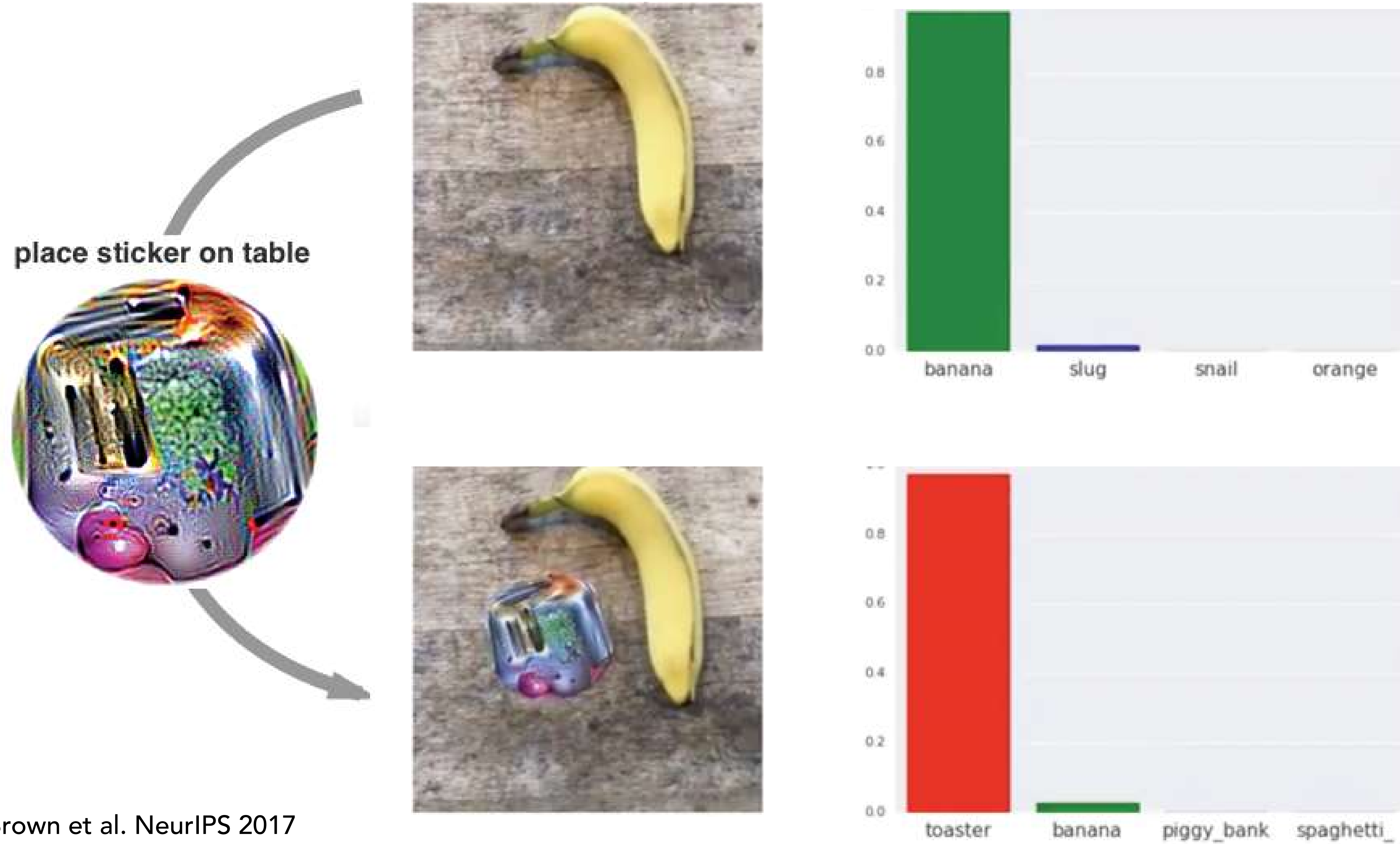
classifier output



Brown et al. NeurIPS 2017

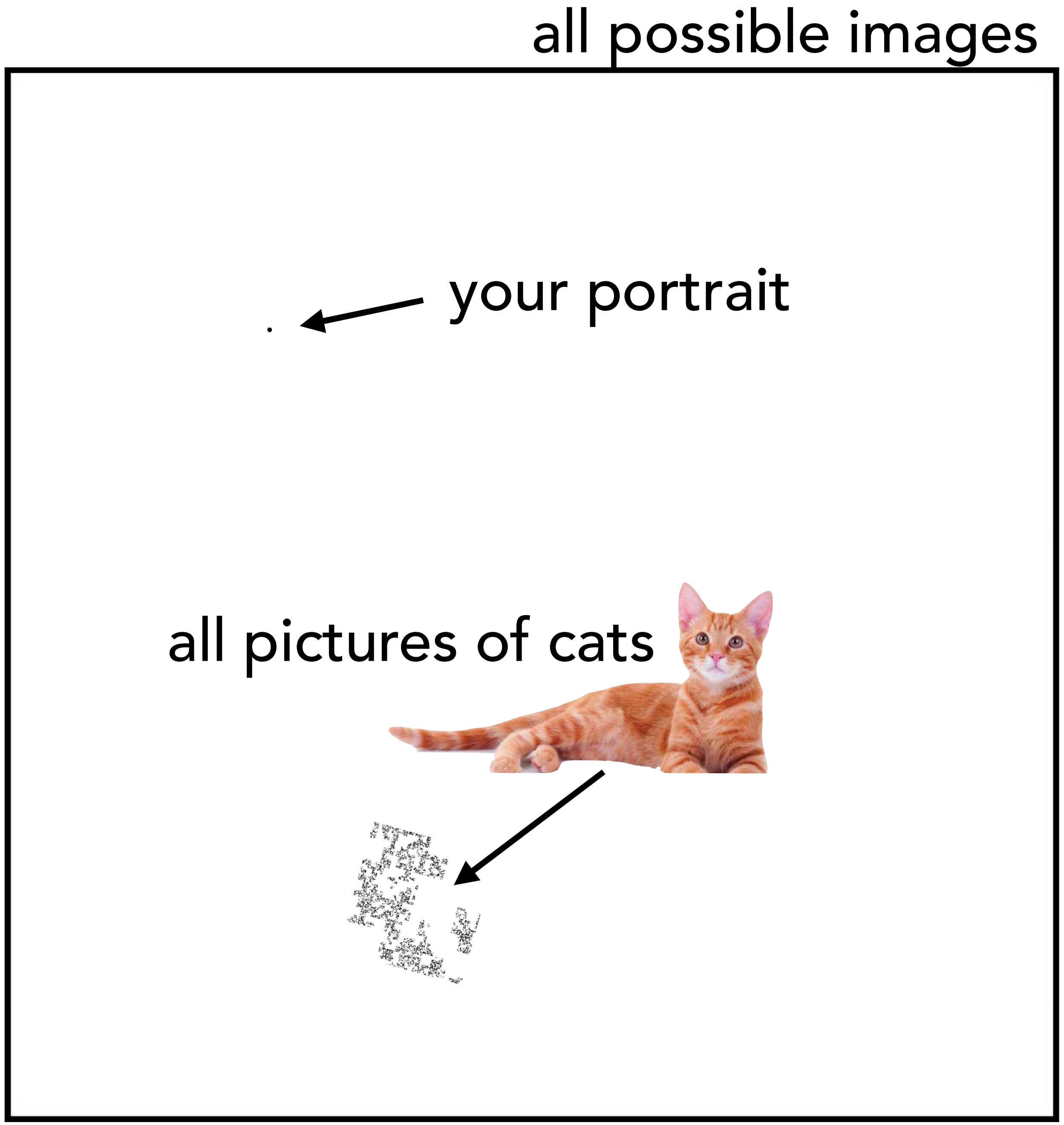
classifier input

classifier output



Brown et al. NeurIPS 2017

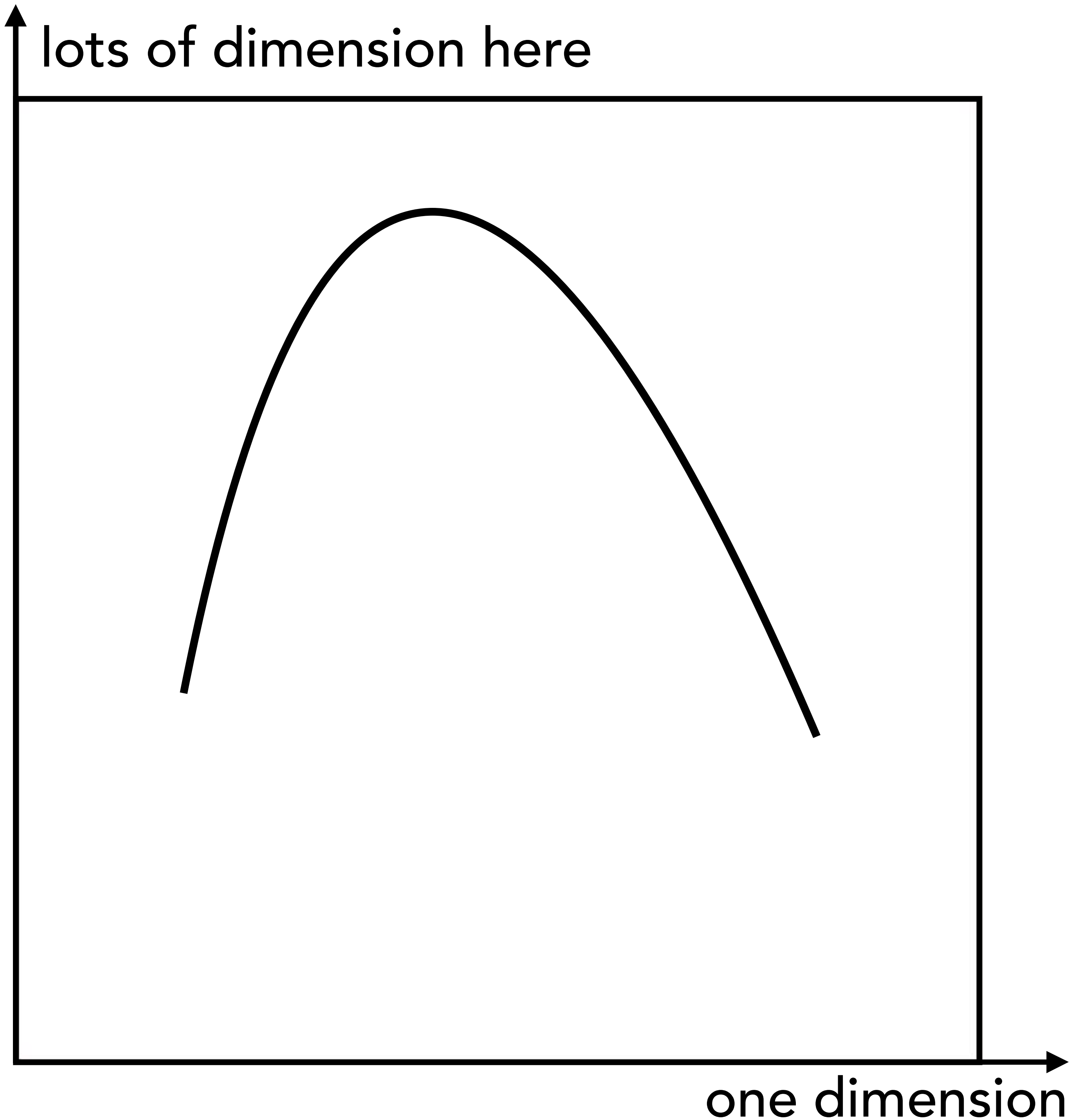
The Hallucination Problem

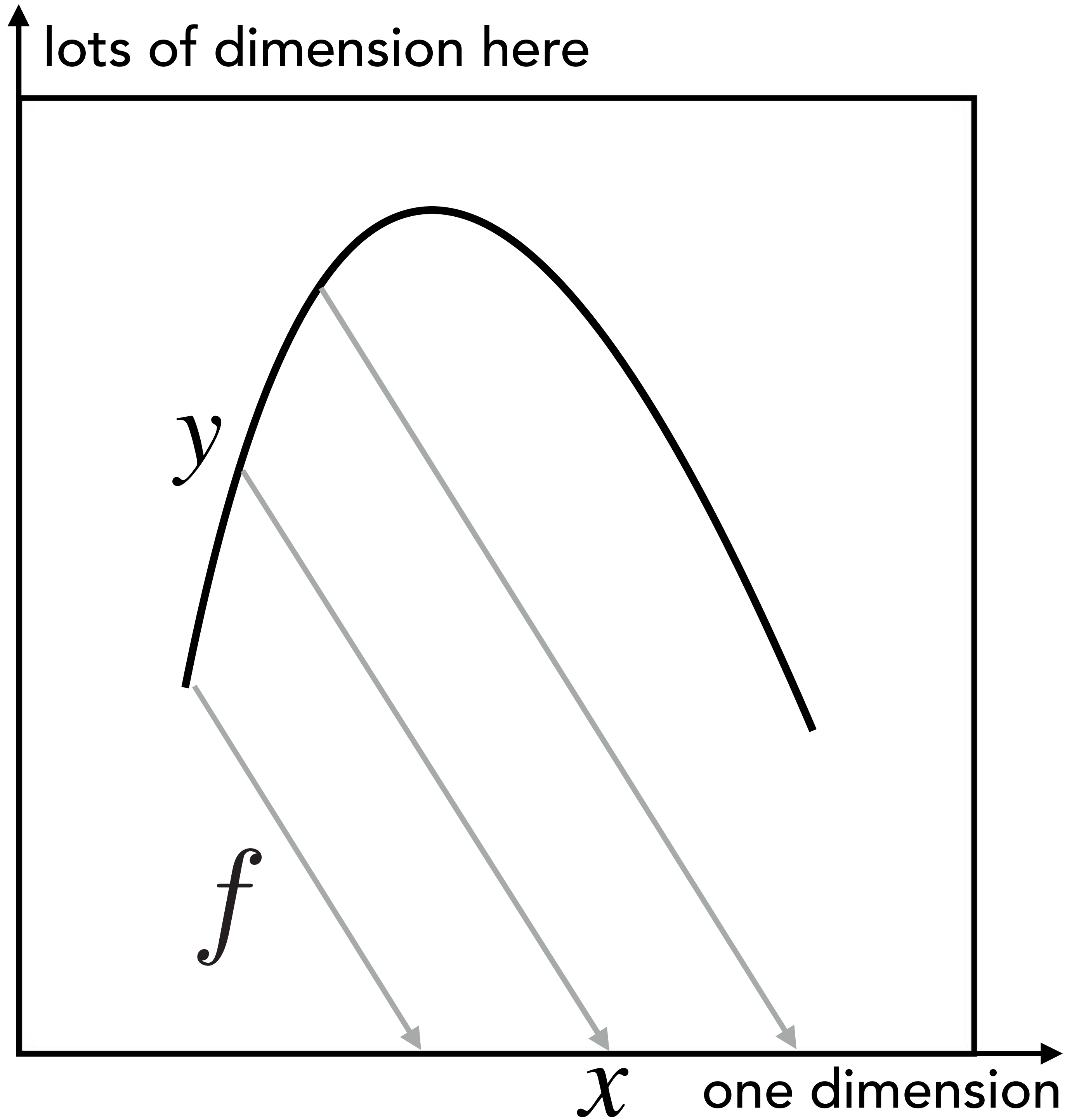


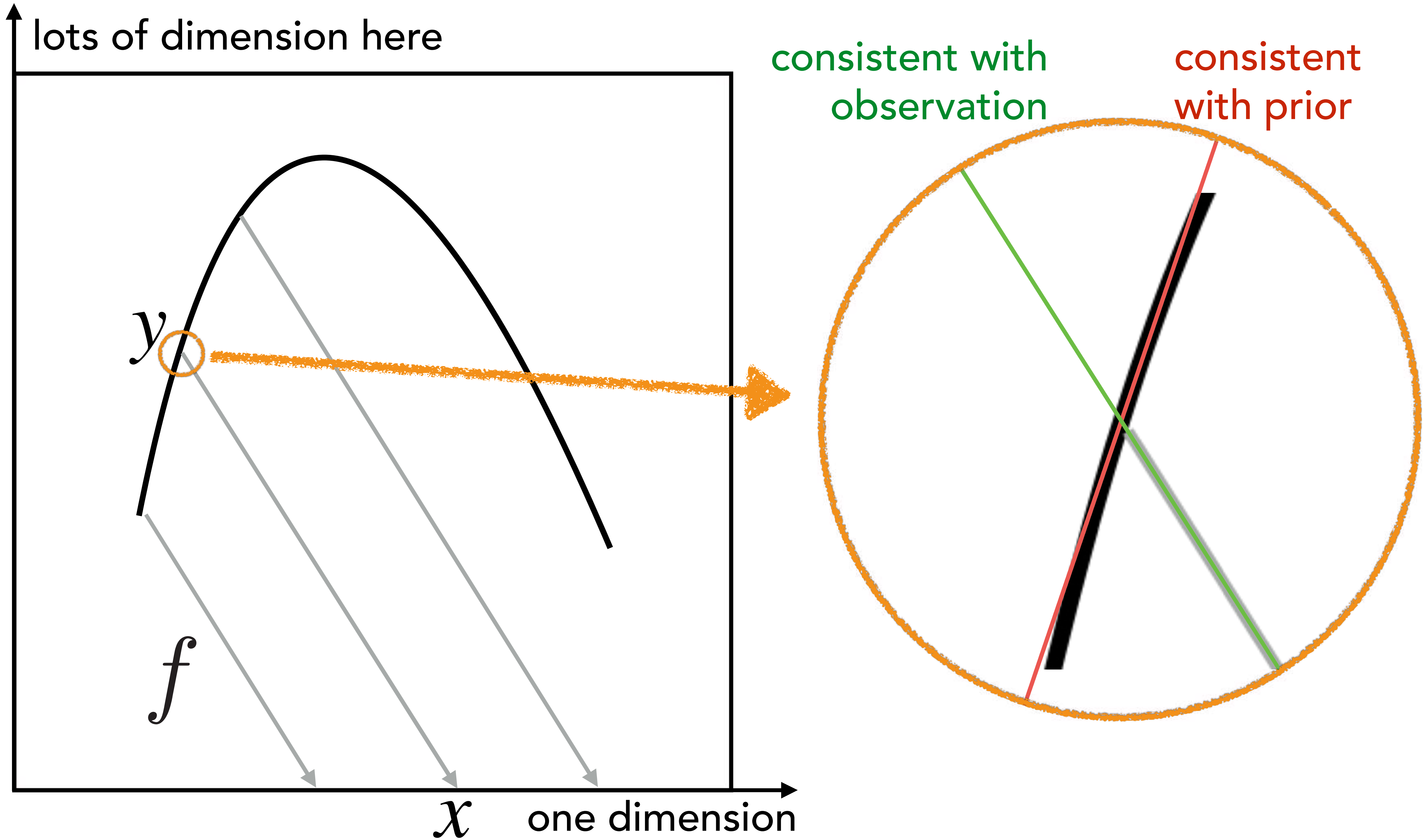
$$\left(2^{24}\right)^{1024\times1024}$$

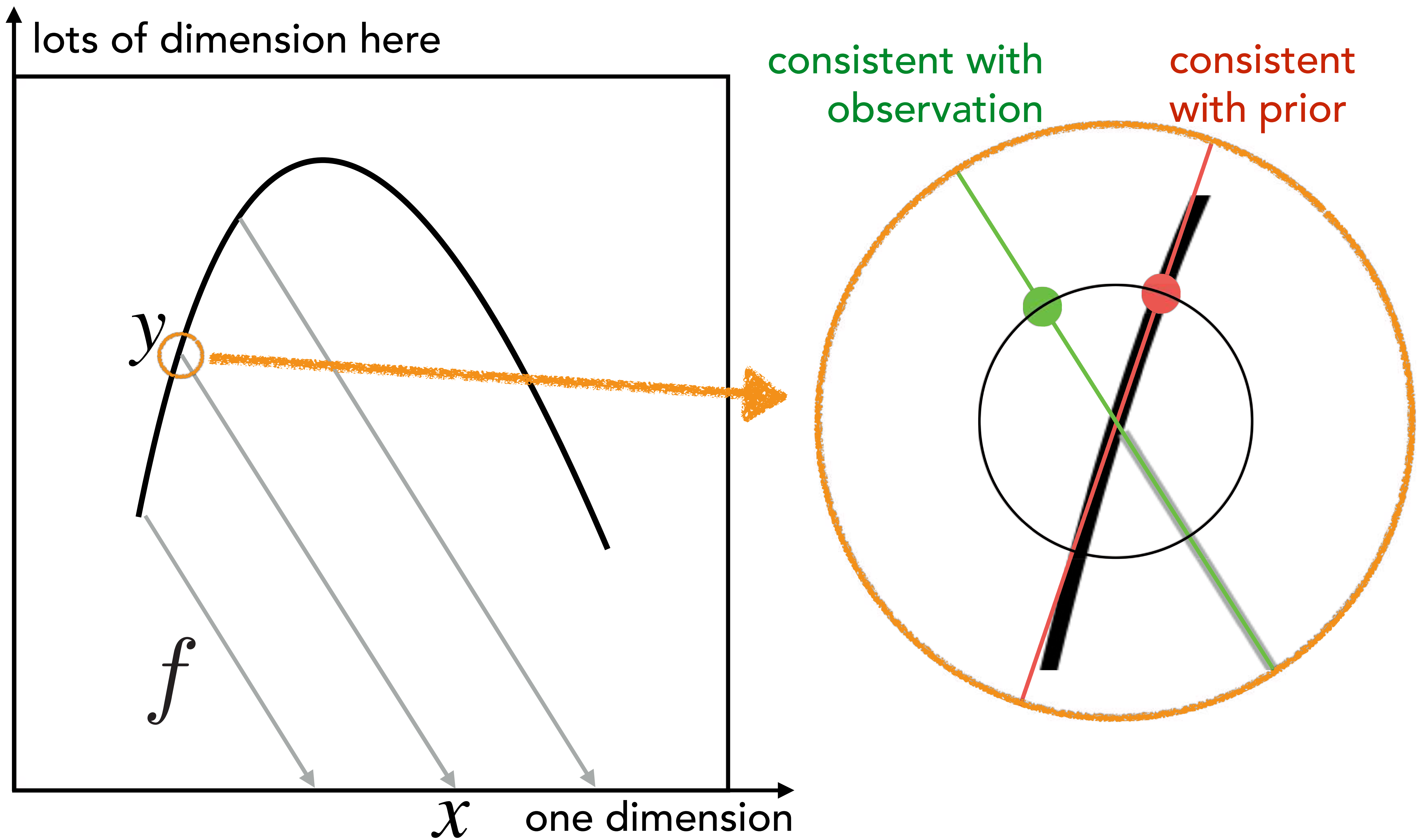
$$= 1000000\ldots$$

7 million zeros


lots of dimension here
one dimension

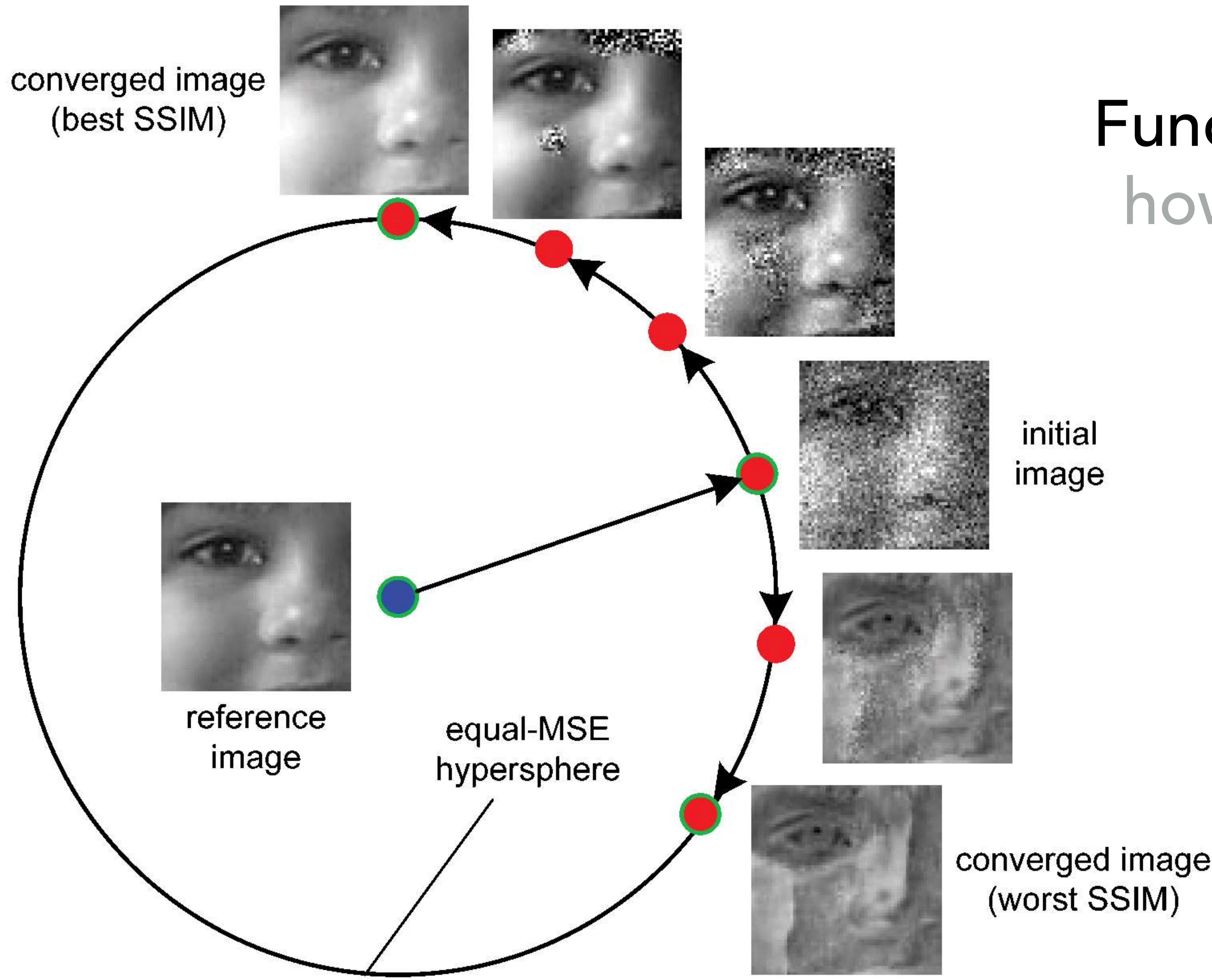

lots of dimension here
y
f
x
one dimension


lots of dimension here
y
f
x
one dimension
consistent with observation
consistent with prior


lots of dimension here
y
f
x
one dimension
consistent with observation
consistent with prior

Fundamental Problem

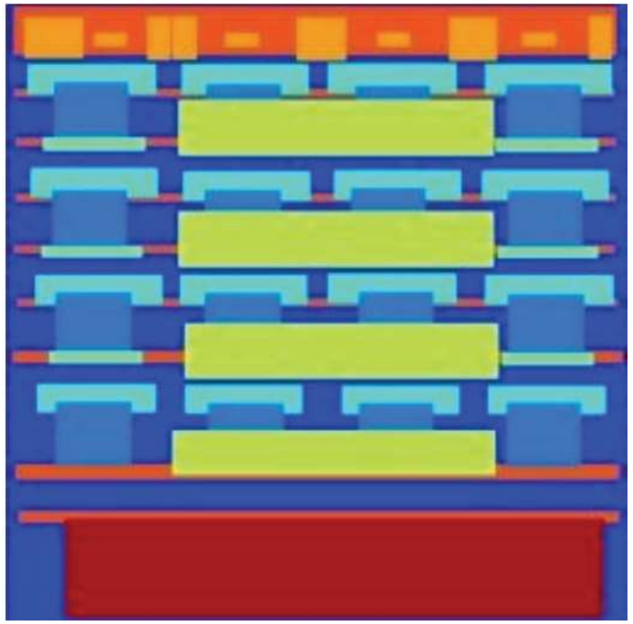
how to estimate how close we are from the solution?

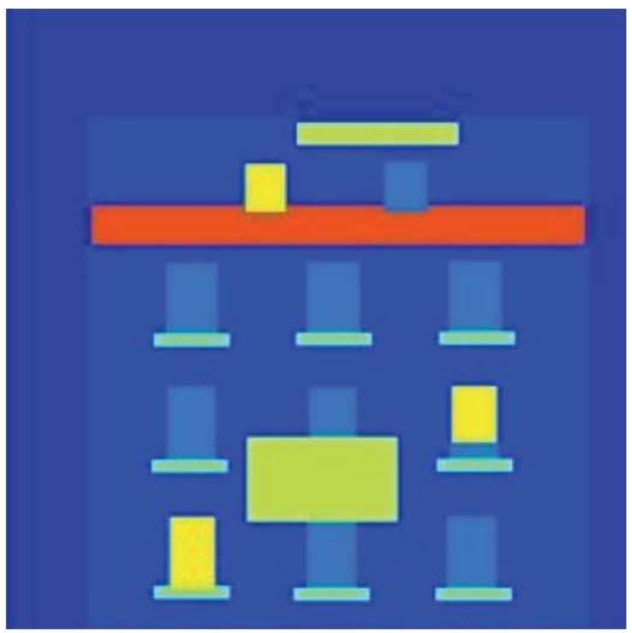


Wang et al. IEEE TIP 2004

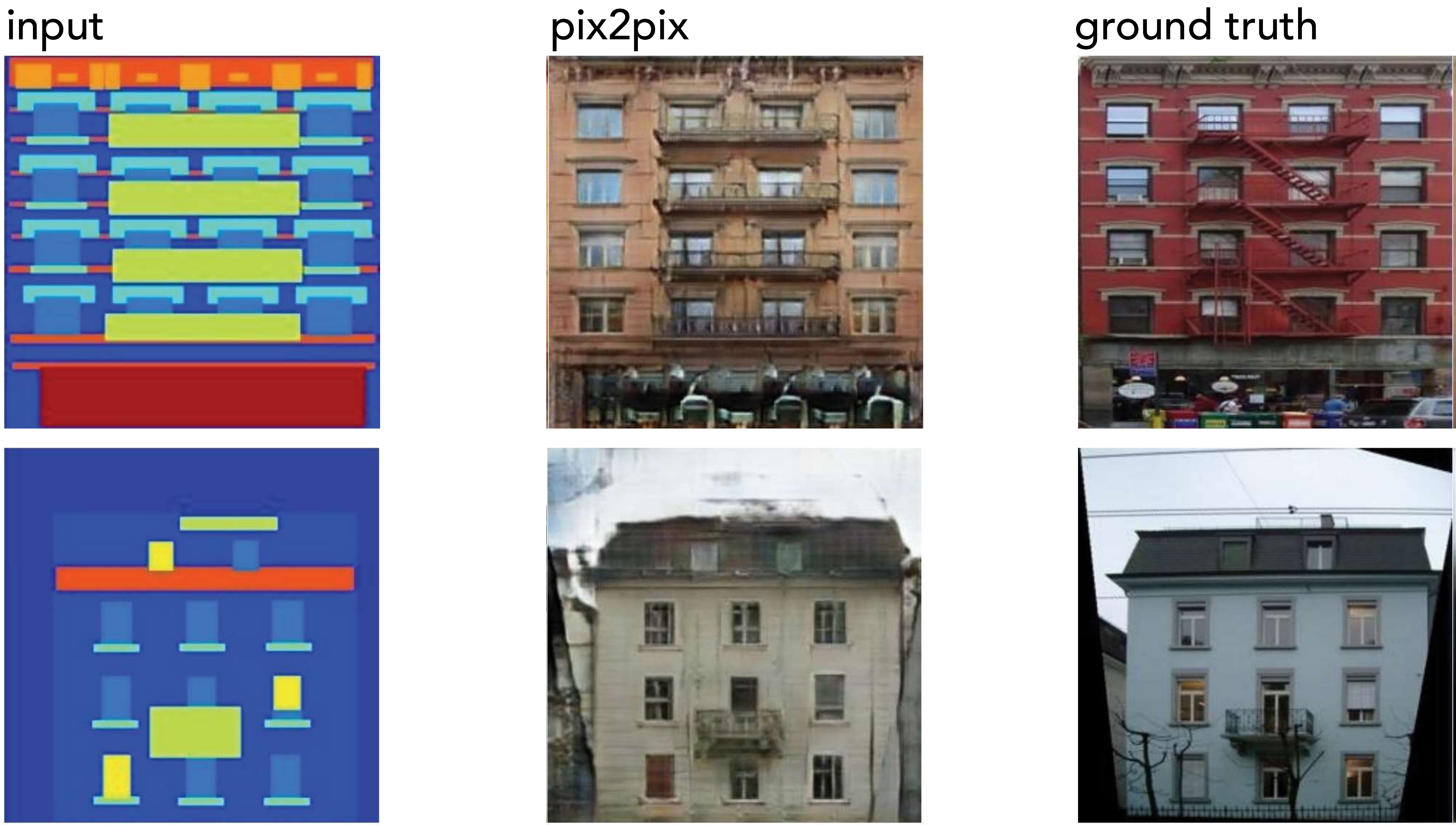
input

ground truth





Isola et al. CVPR 2017



Isola et al. CVPR 2017

The Cheating Problem

Horse

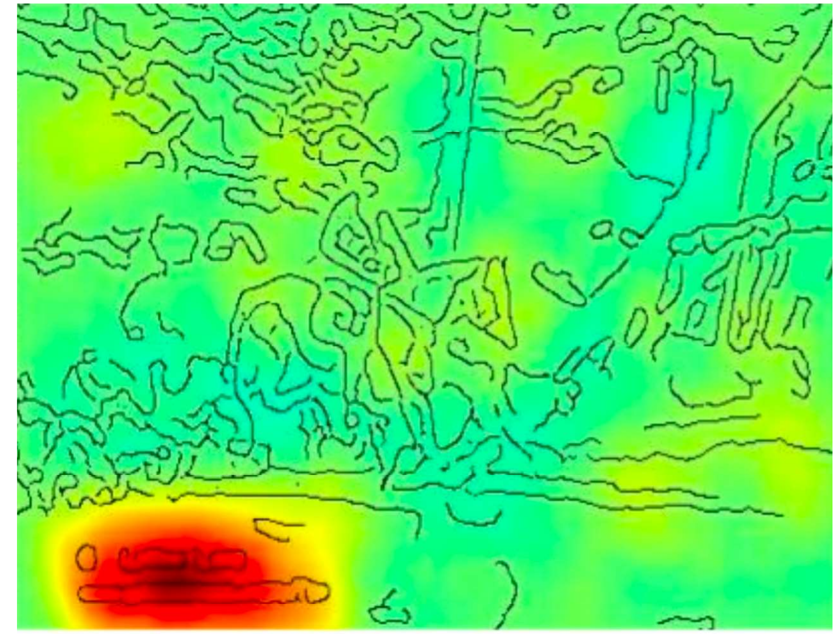
Lapuschkin *et al. PLoS One* (2019)

Horse

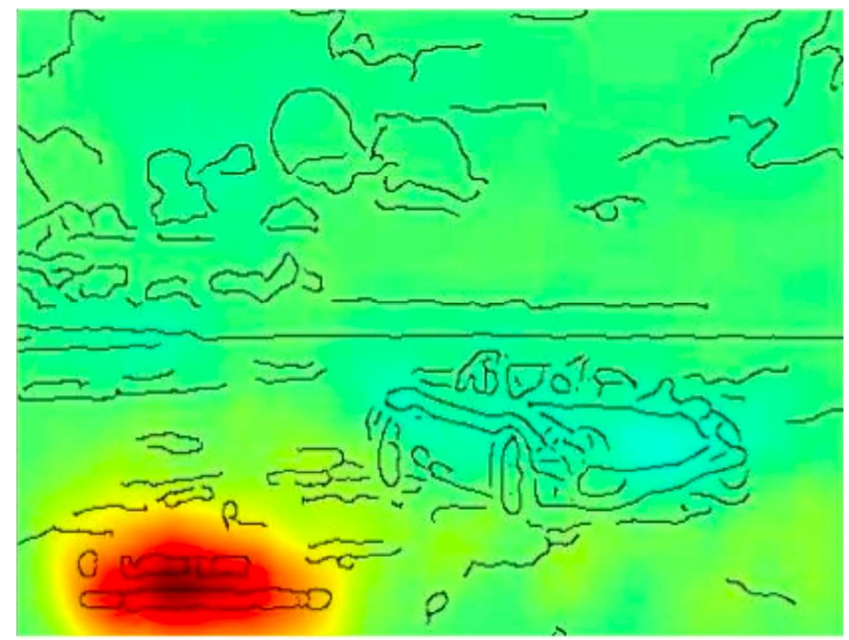
Horse

Lapuschkin *et al. PLoS One* (2019)

Horse



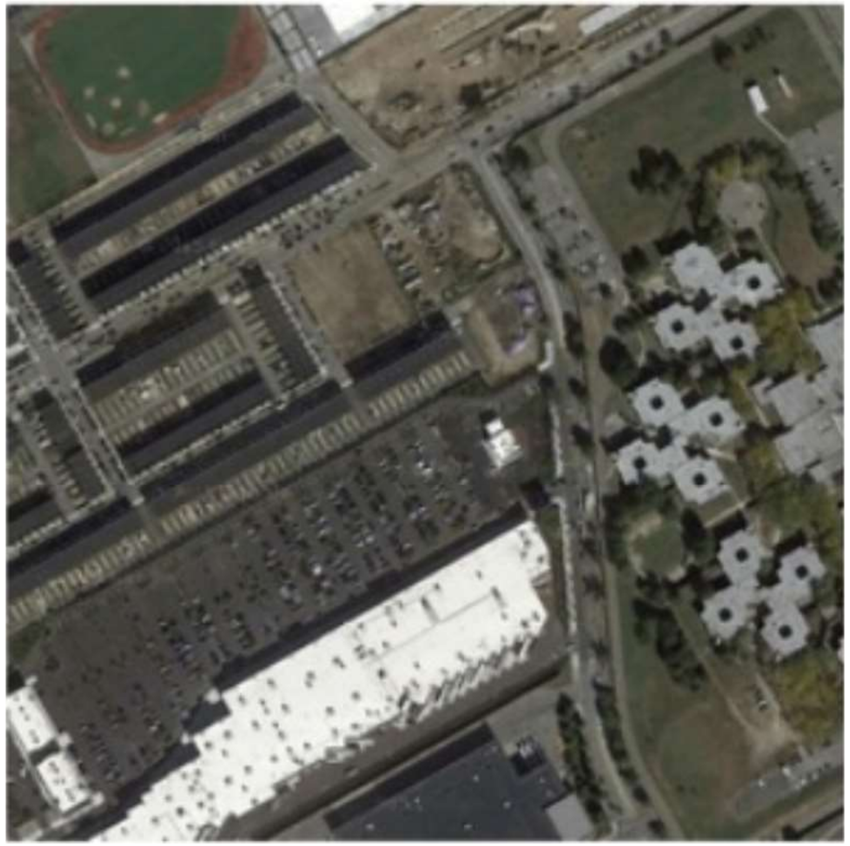
Horse



Lapuschkin *et al. PLoS One* (2019)

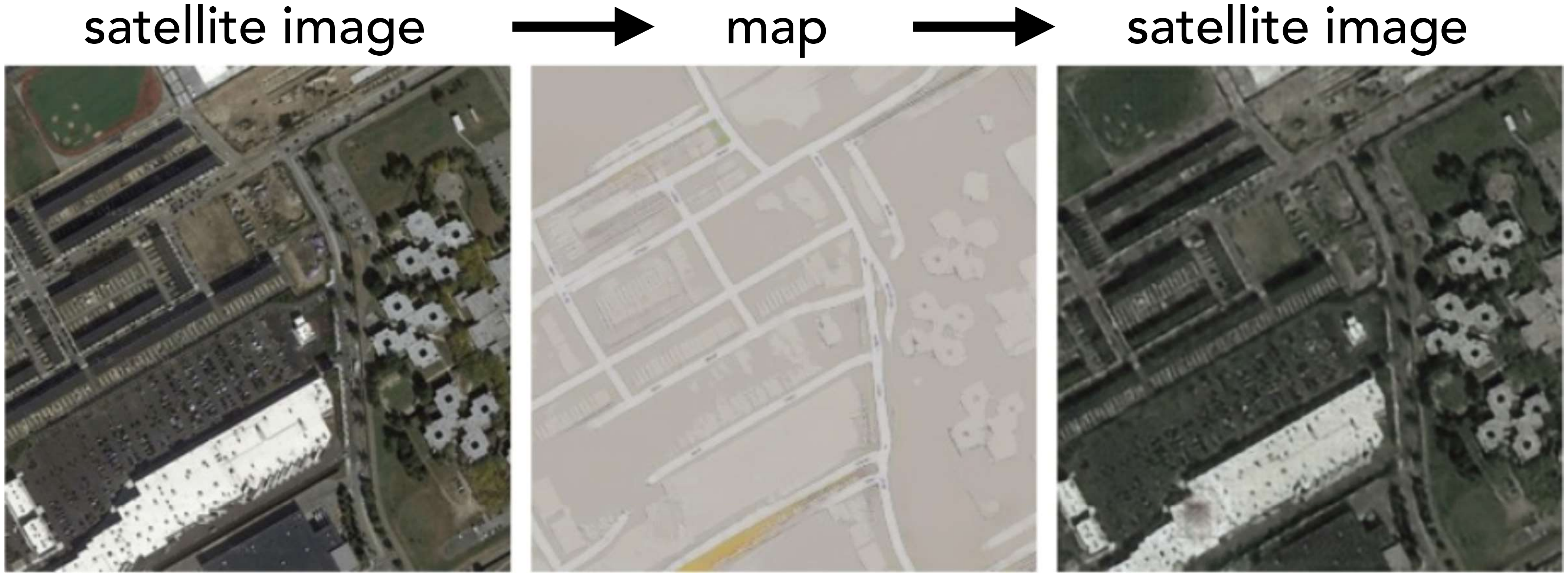
Layer-wise Relevance Propagation (LRP)
Bach et al. (2015)

satellite image

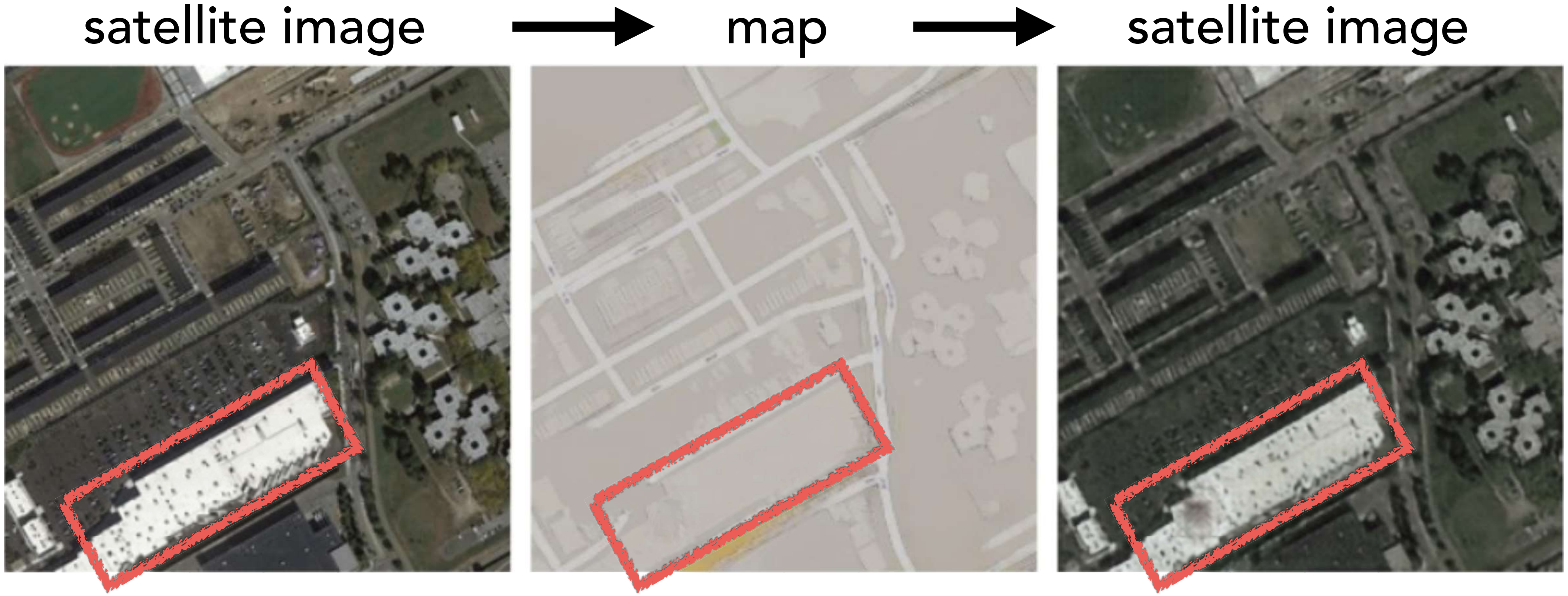


map

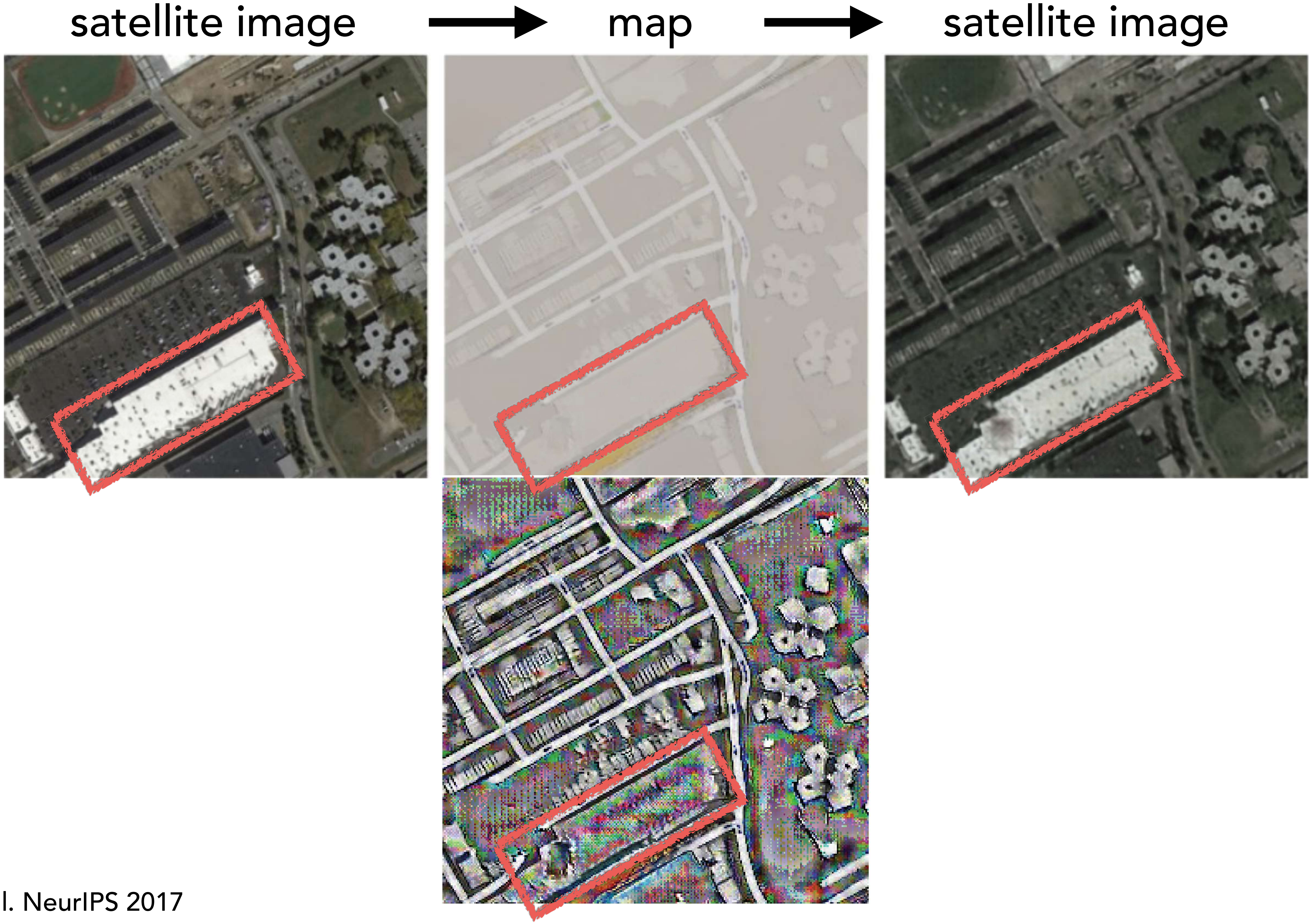
Chu et al. NeurIPS 2017



Chu et al. NeurIPS 2017



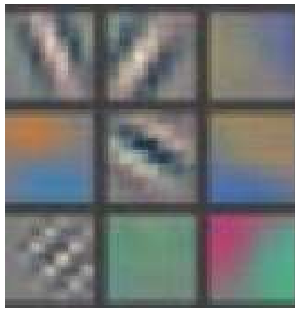
Chu et al. NeurIPS 2017

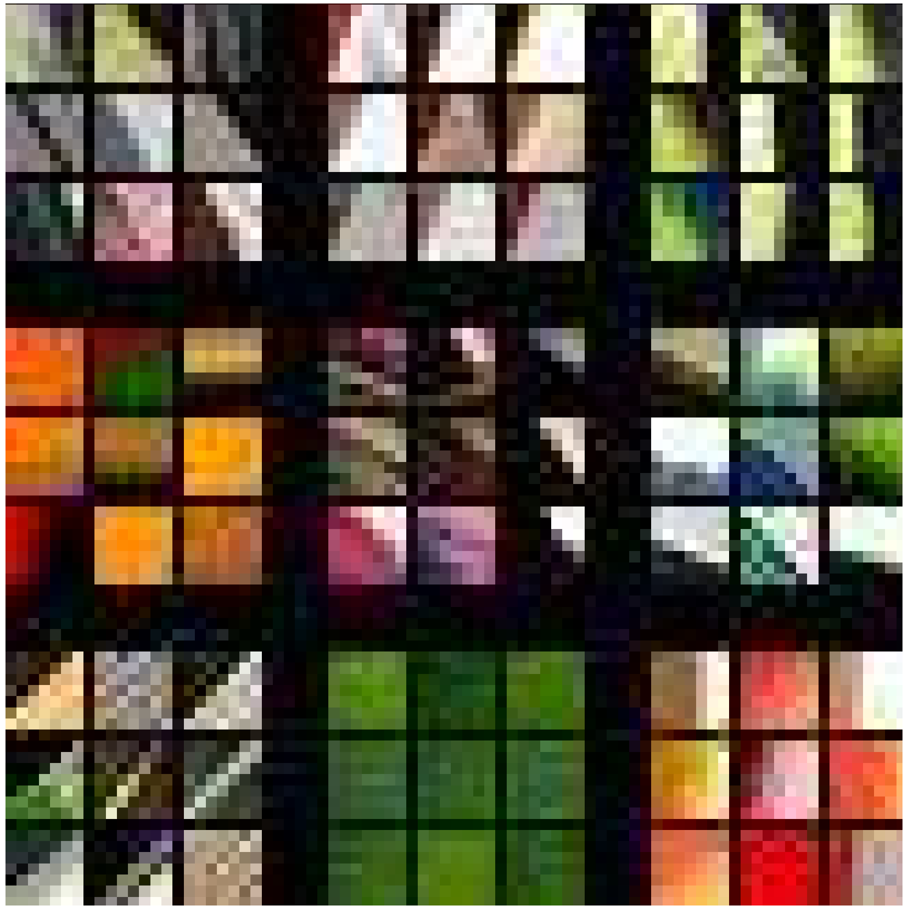


Chu et al. NeurIPS 2017

The Black Box Problem

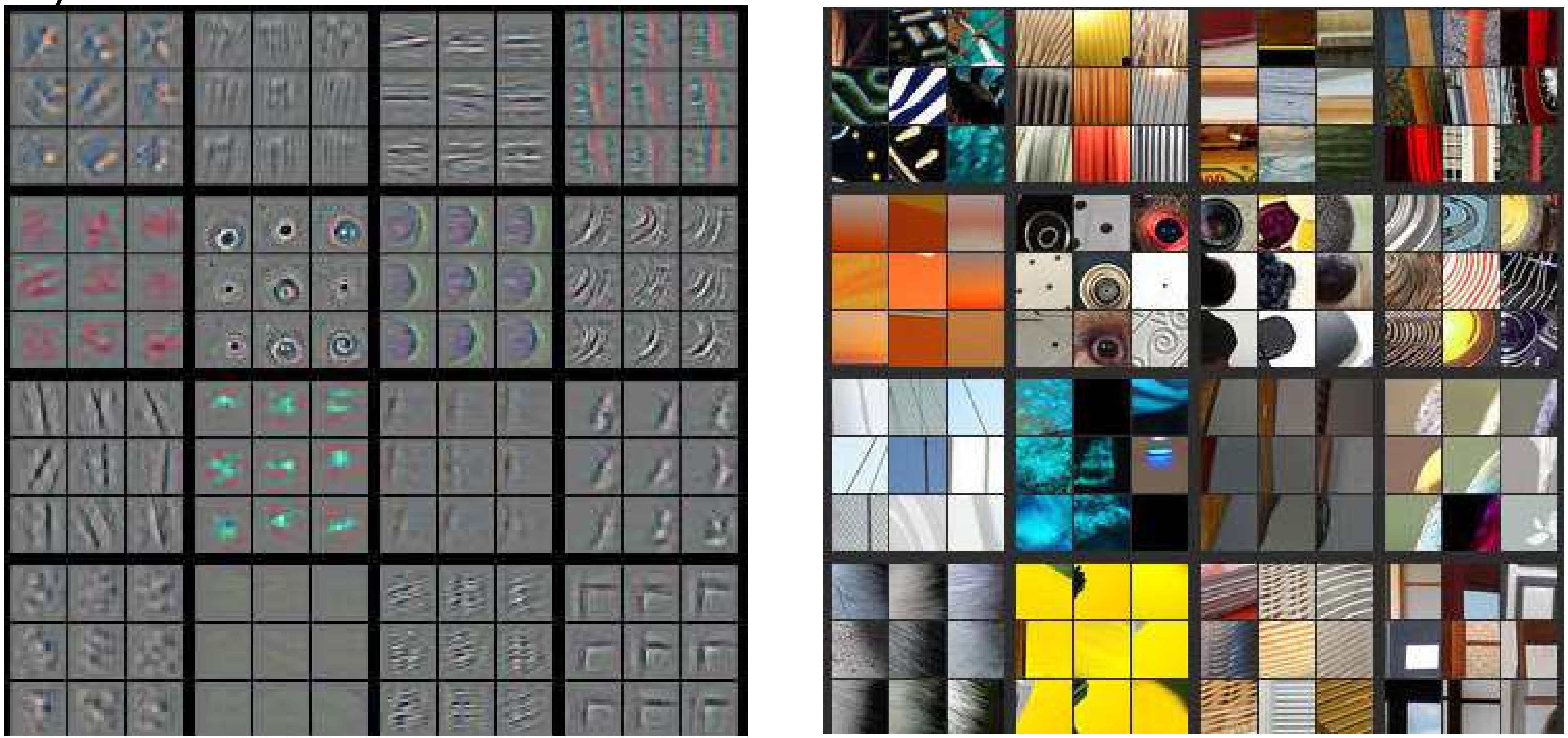
Layer 1





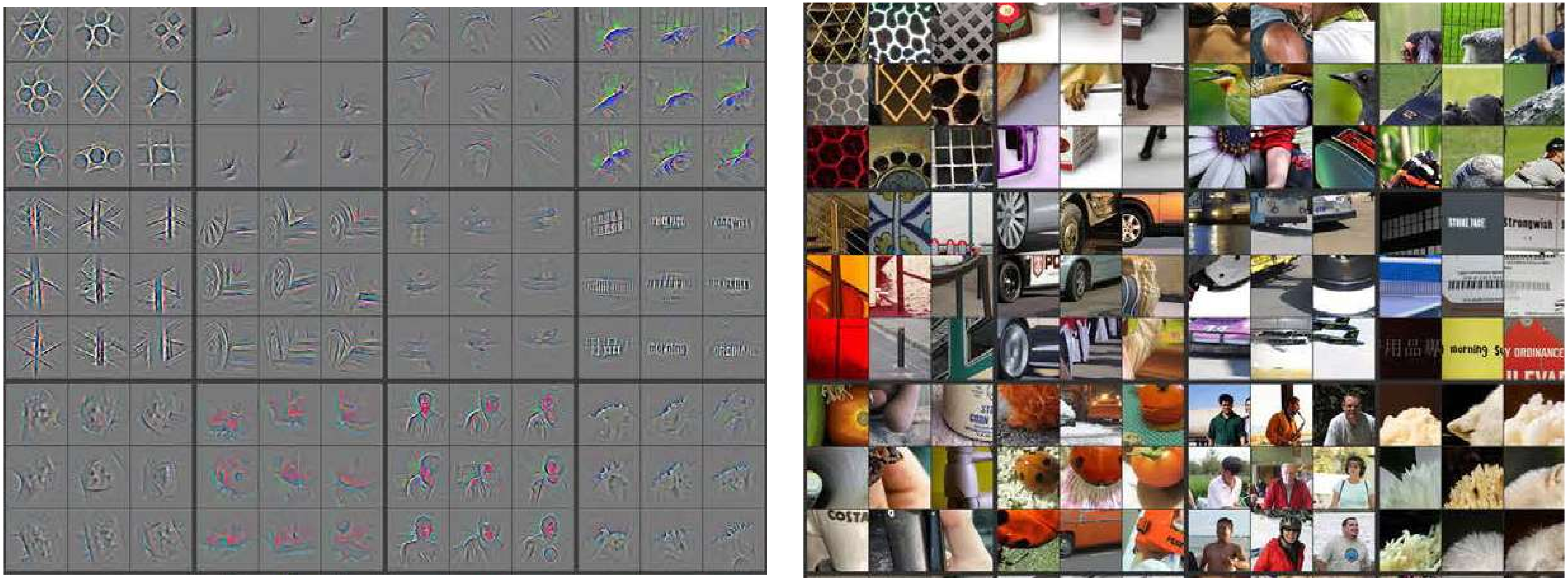
Zeiler et al. ECCV 2013

Layer 2



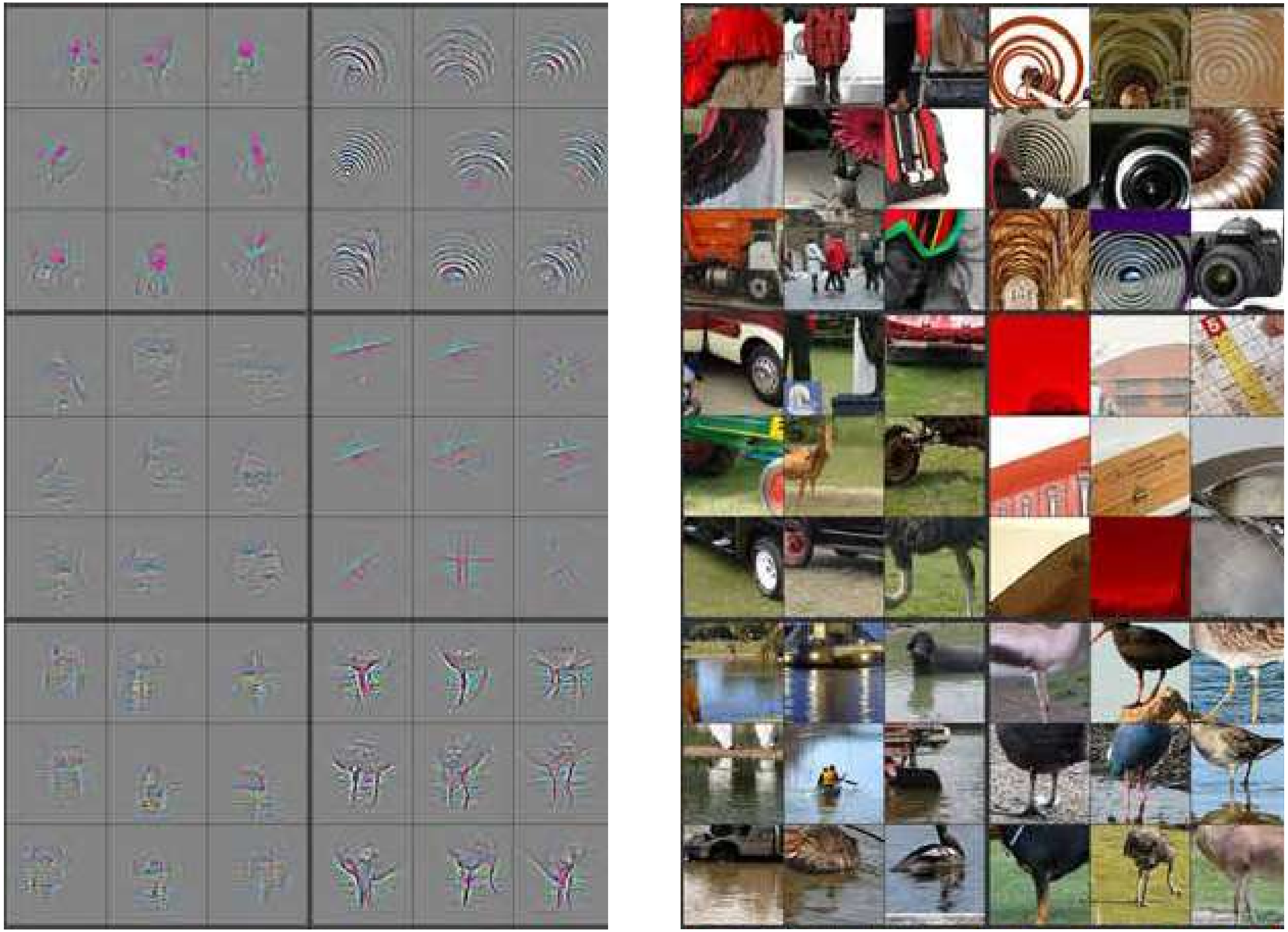
Zeiler et al. ECCV 2013

Layer 3



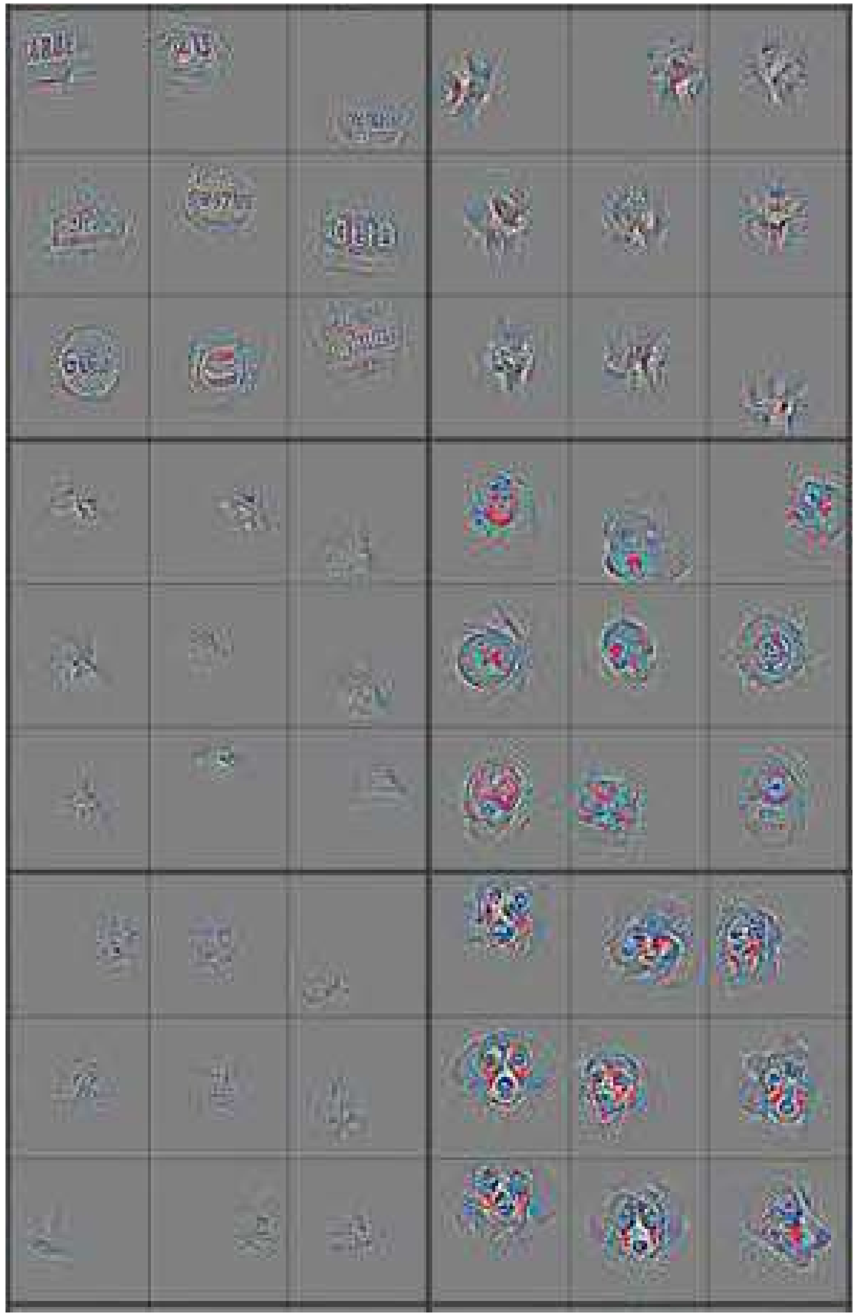
Zeiler et al. ECCV 2013

Layer 4

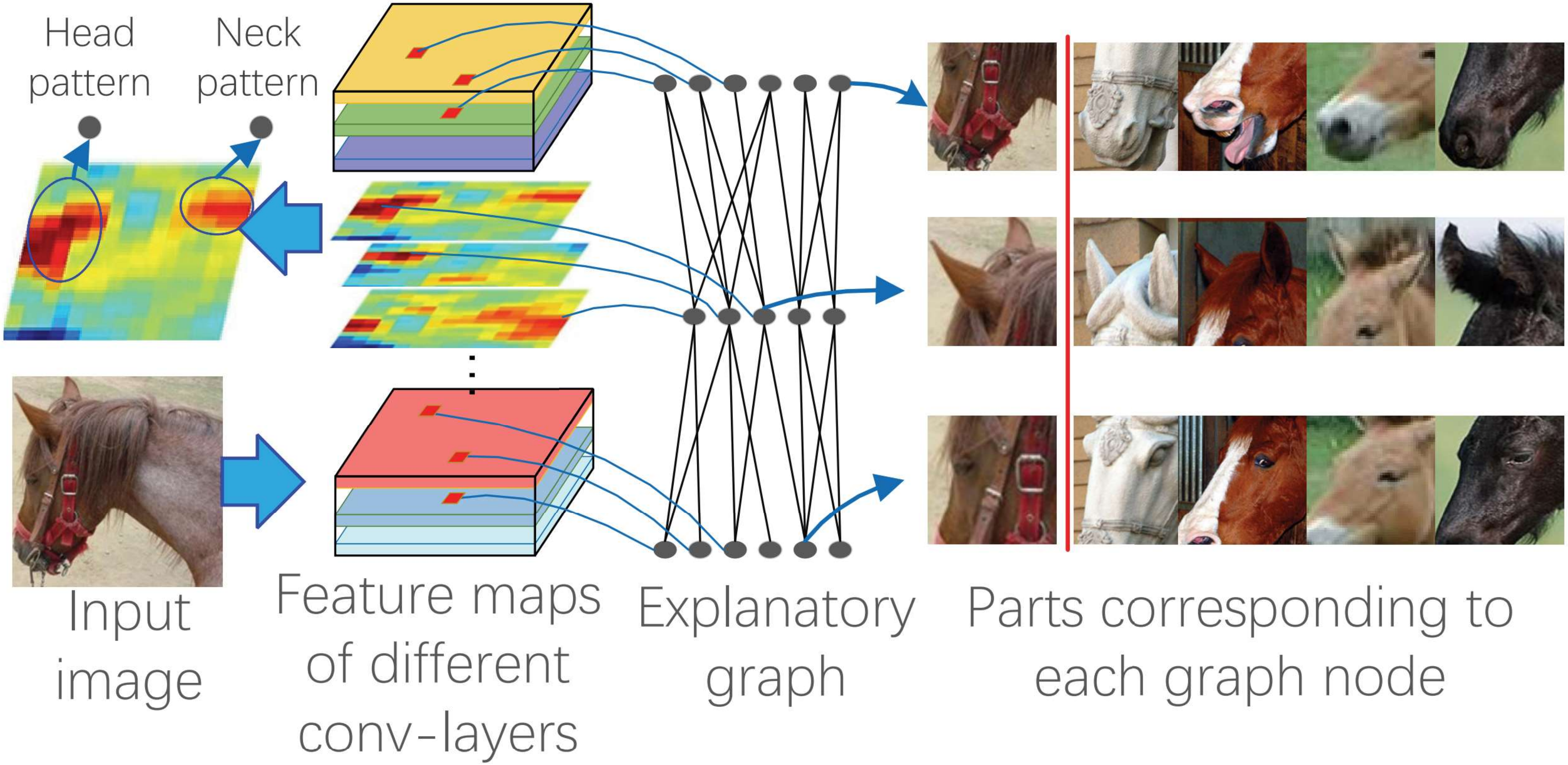


Zeiler et al. ECCV 2013

Layer 5

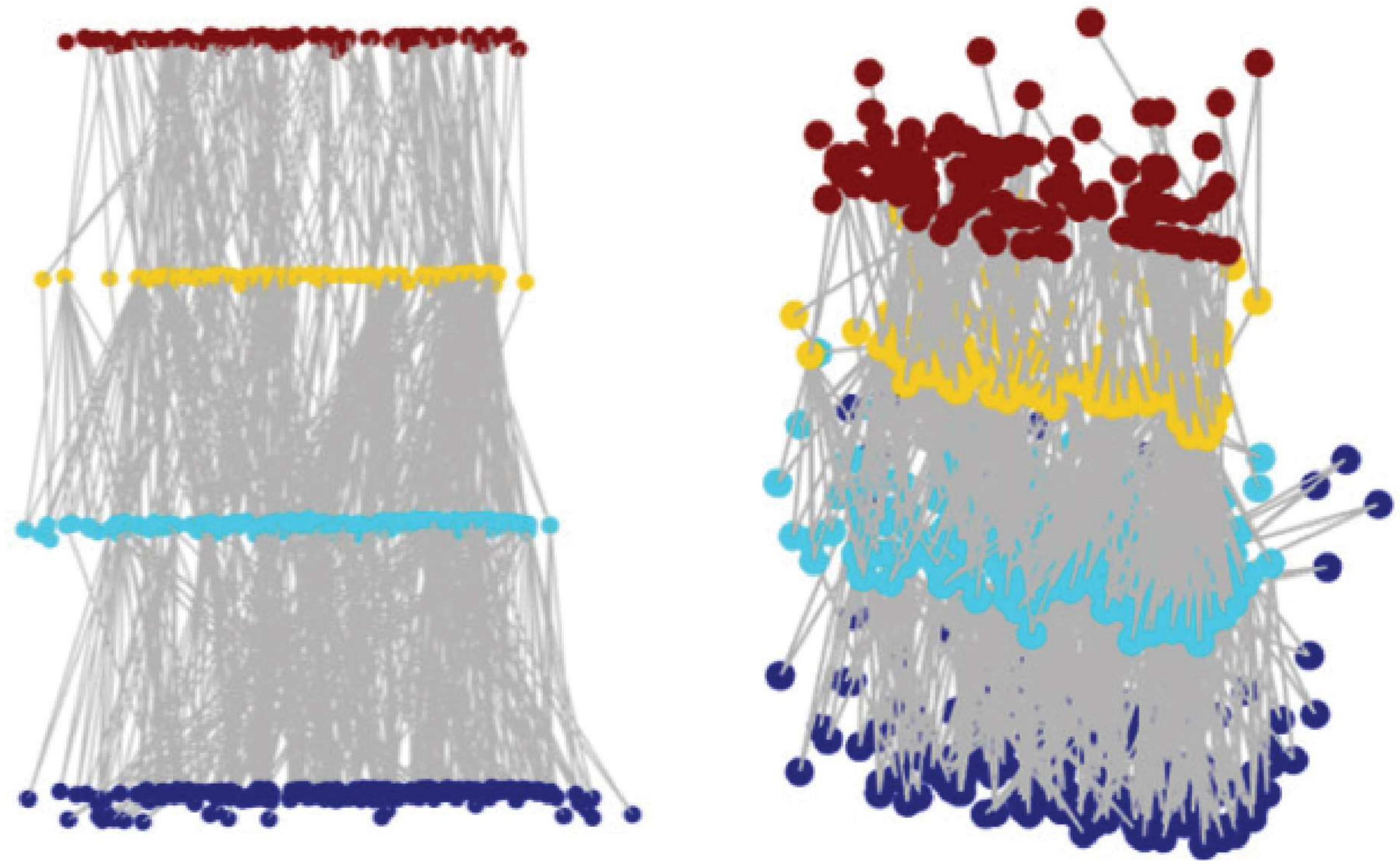


Zeiler et al. ECCV 2013



Zhang et al. AAAI 2018

Right…
'explained'

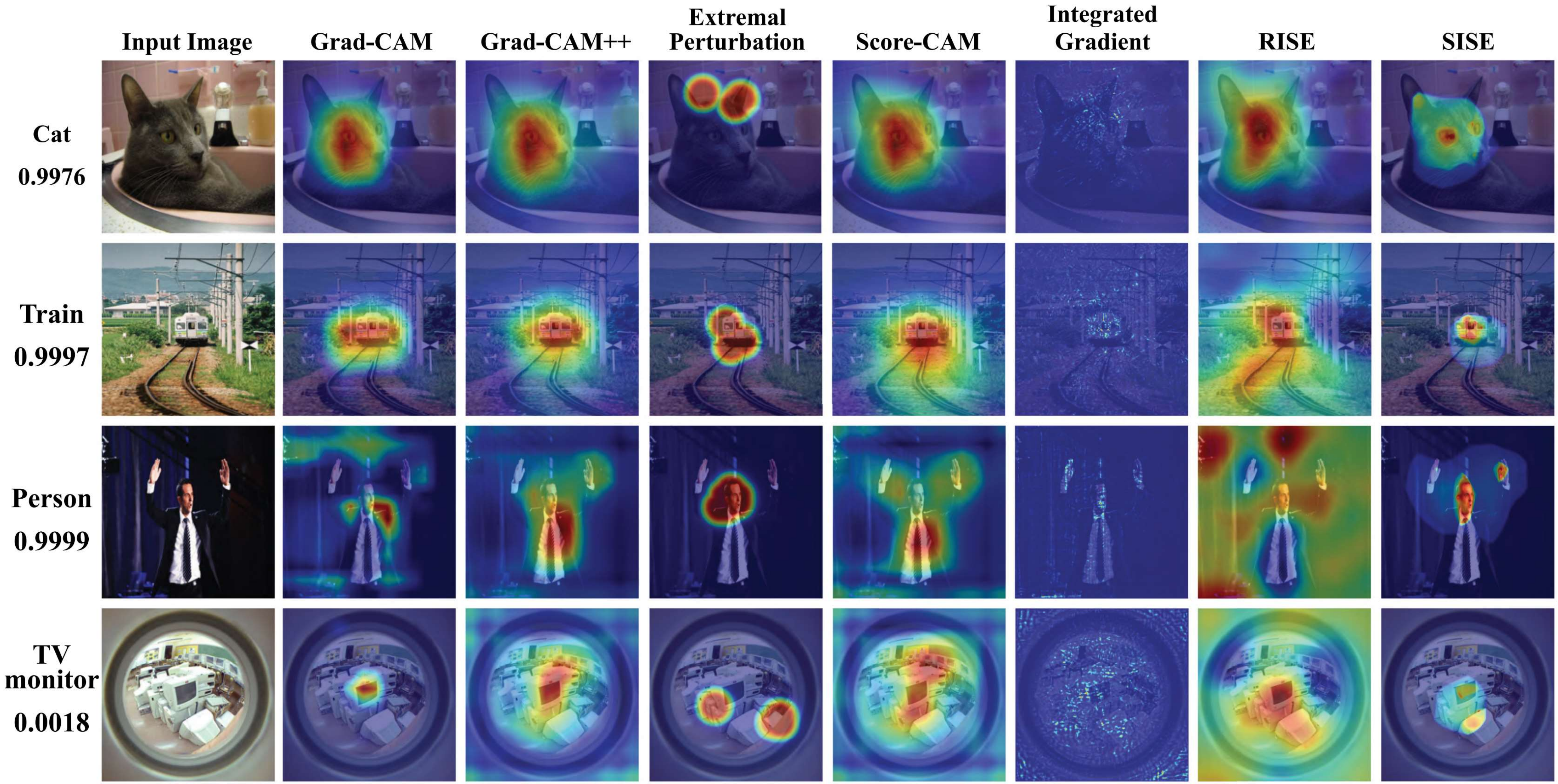


Zhang et al. AAAI 2018

Explaining by ‘Attribution’

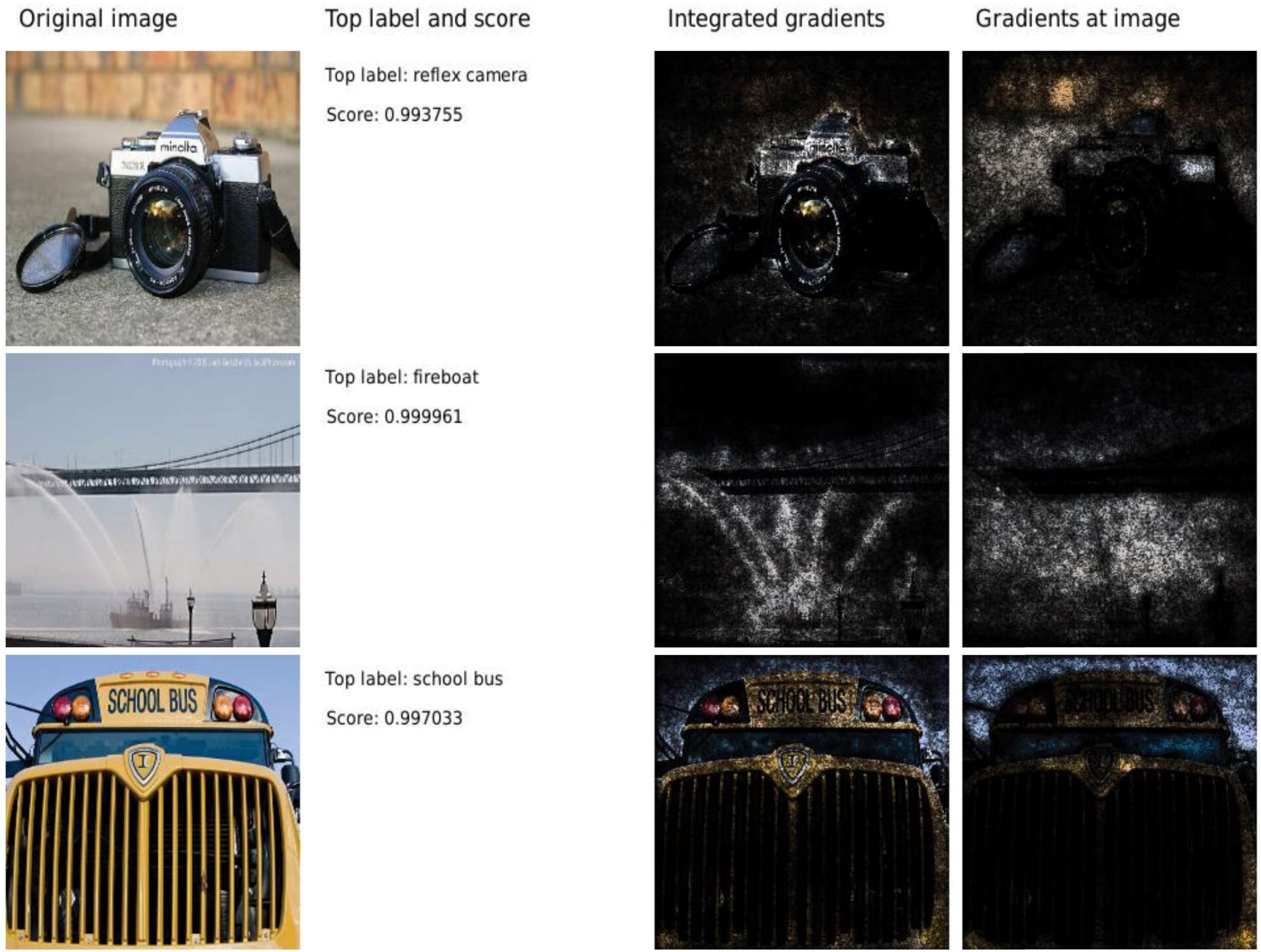
Saliency Maps
which pixels matter?

Simonyan arXiv 2013



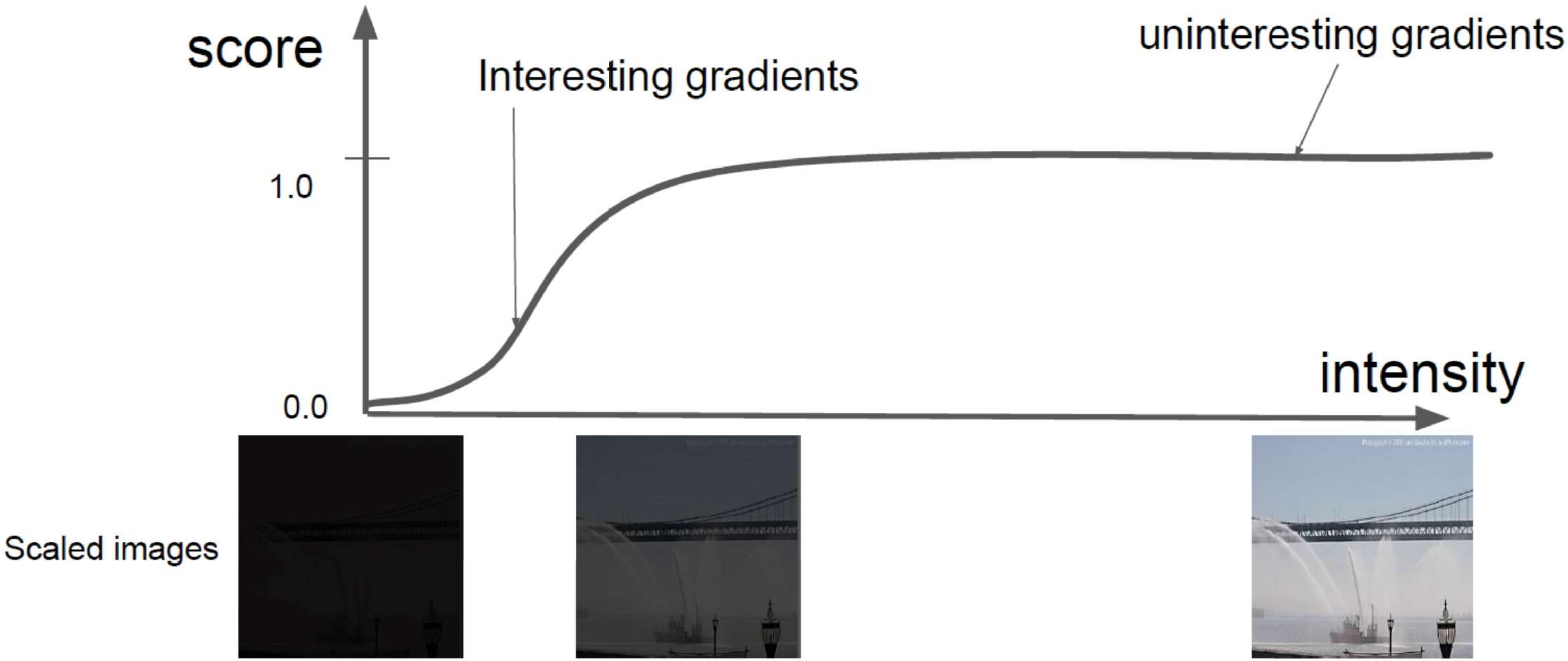
Sattarzadeh et al. arXiv 2020

Integrated gradients



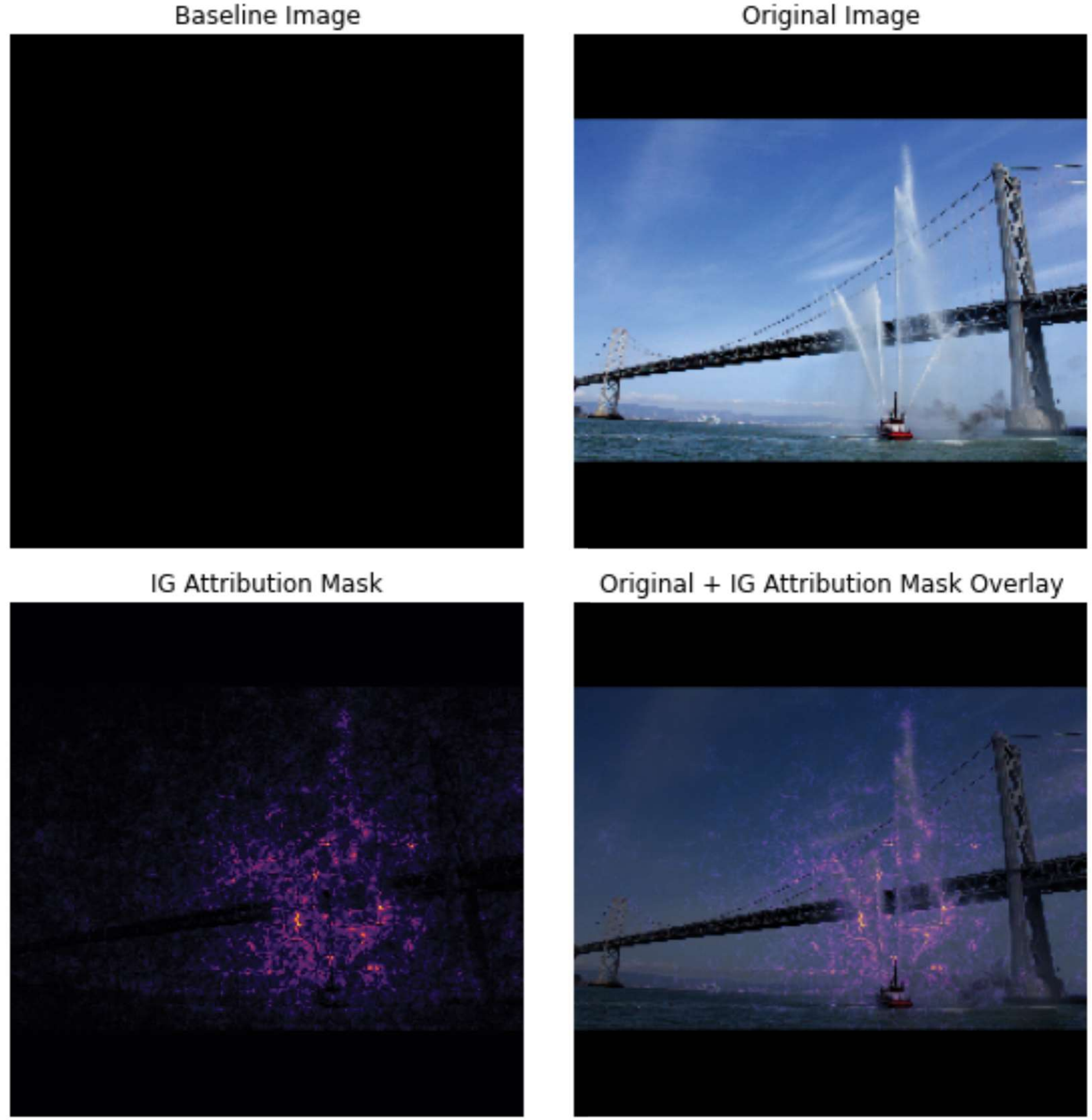
Sundararajan et al. arXiv 2017

Integrated gradients

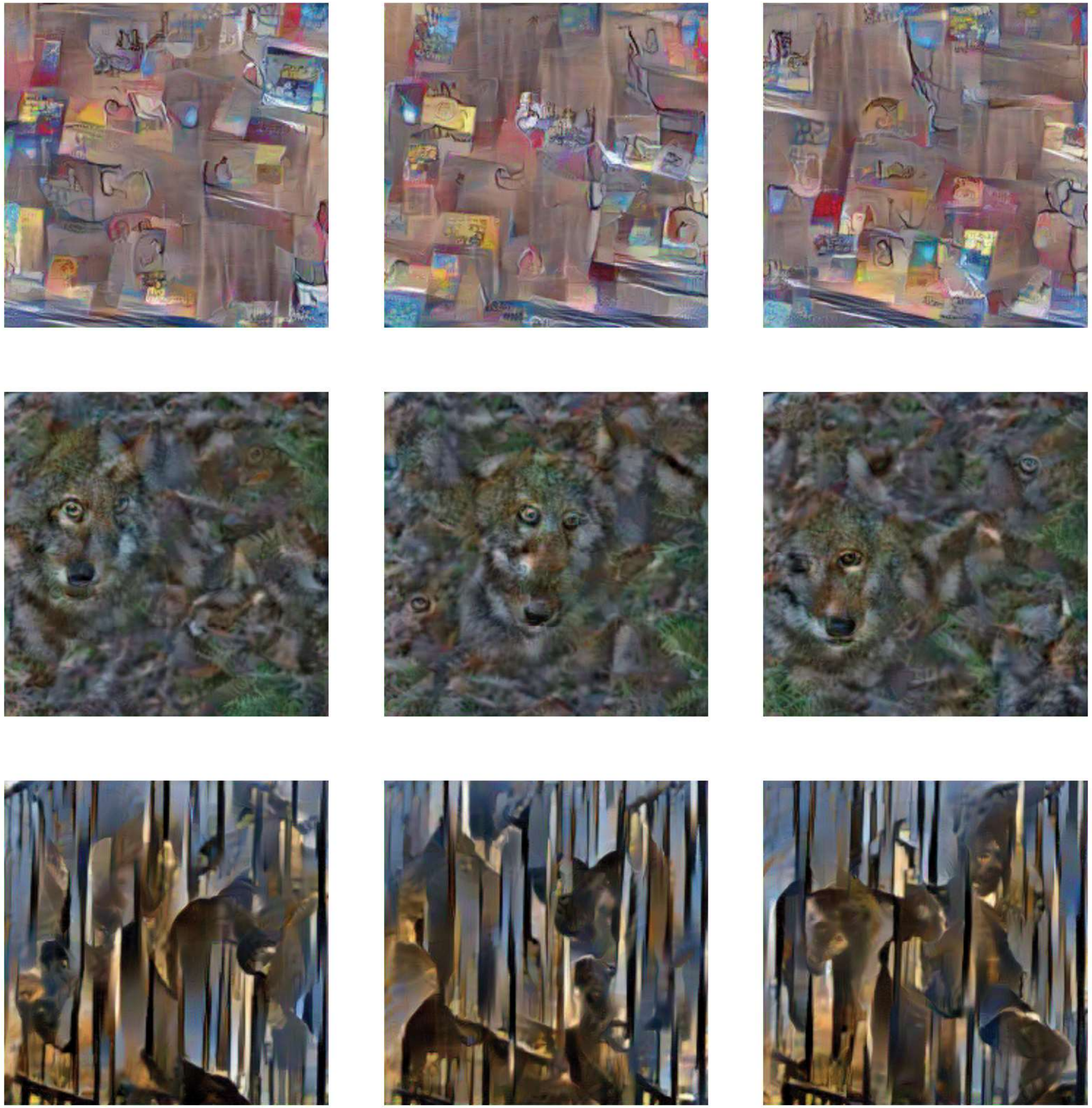


Sundararajan et al. arXiv 2017

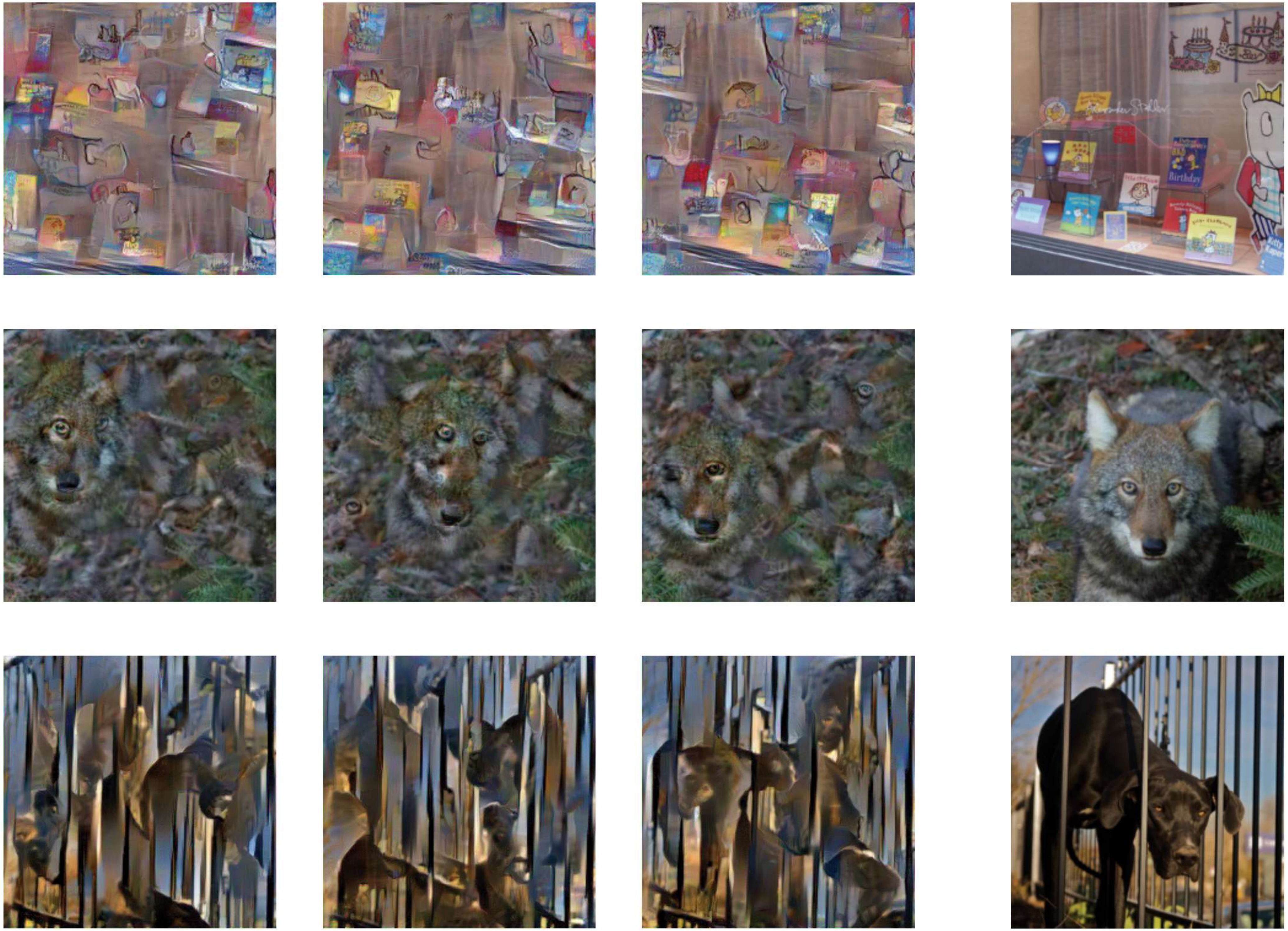
Integrated gradients



The ‘not that smart’ Problem

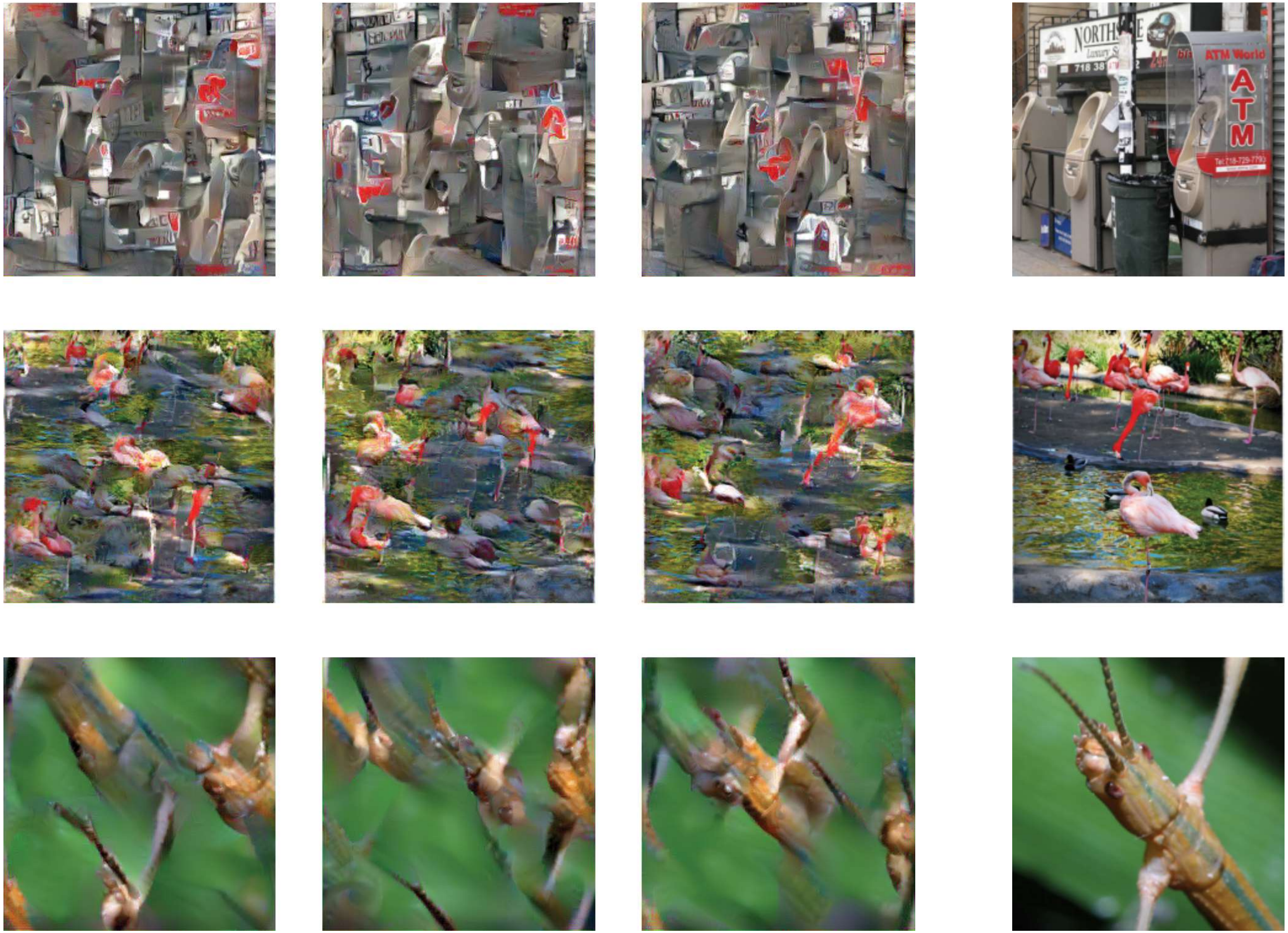


Brendel
& Bethge
ICLR 2019

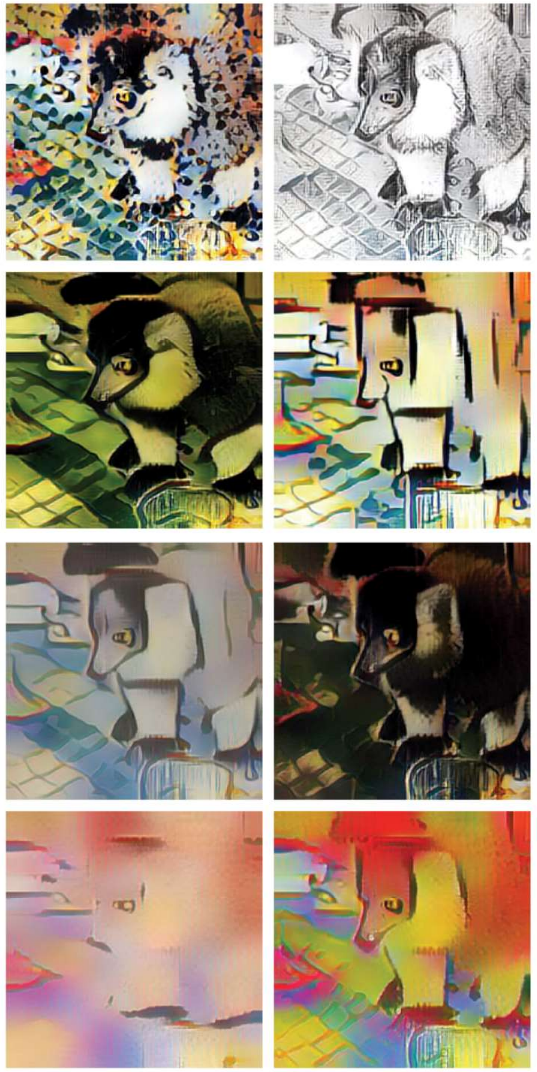


Brendel
& Bethge
ICLR 2019

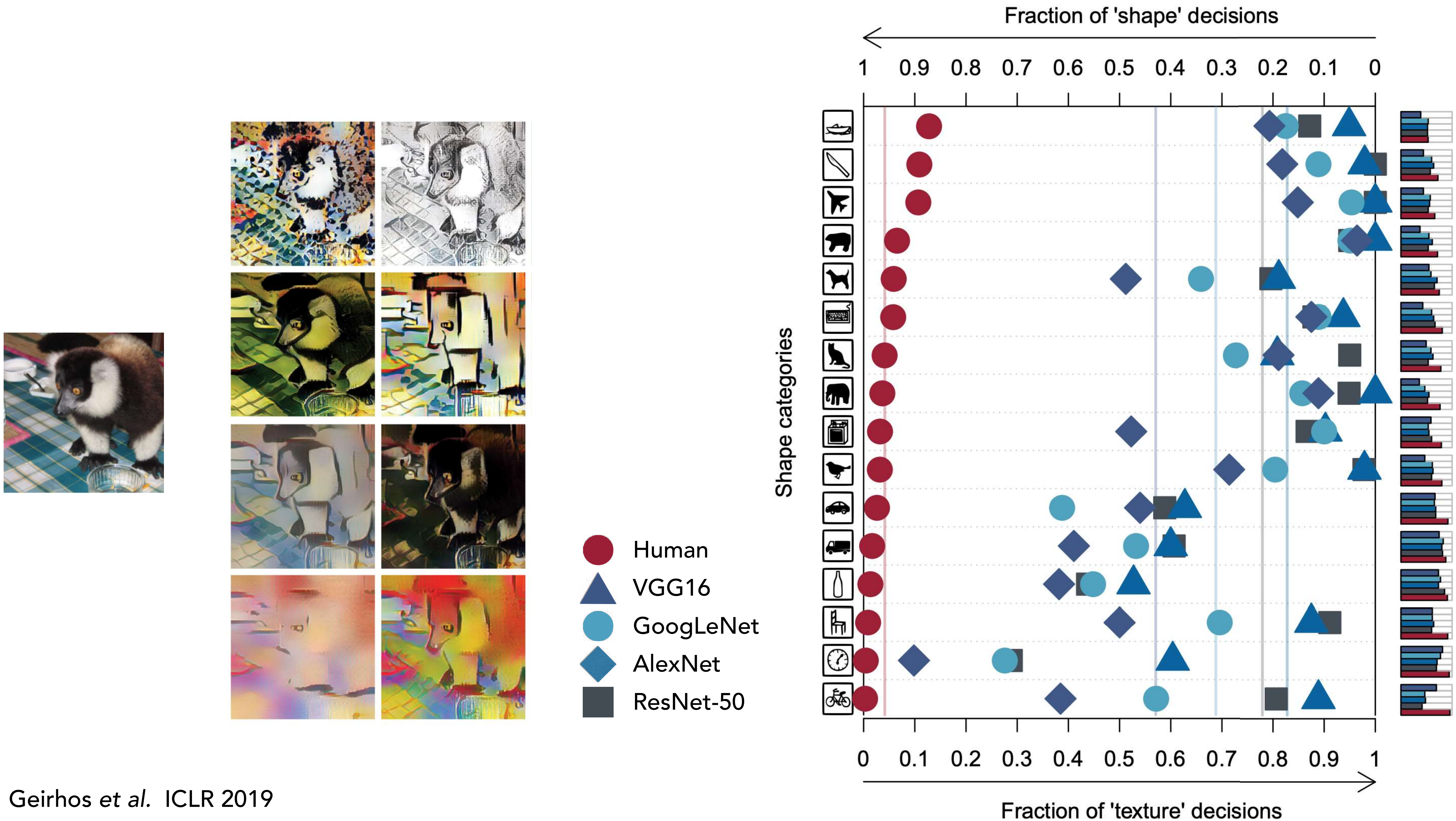
Brendel
& Bethge
ICLR 2019



Brendel
& Bethge
ICLR 2019



Geirhos *et al.* ICLR 2019



Geirhos *et al.* ICLR 2019

Panel Discussion