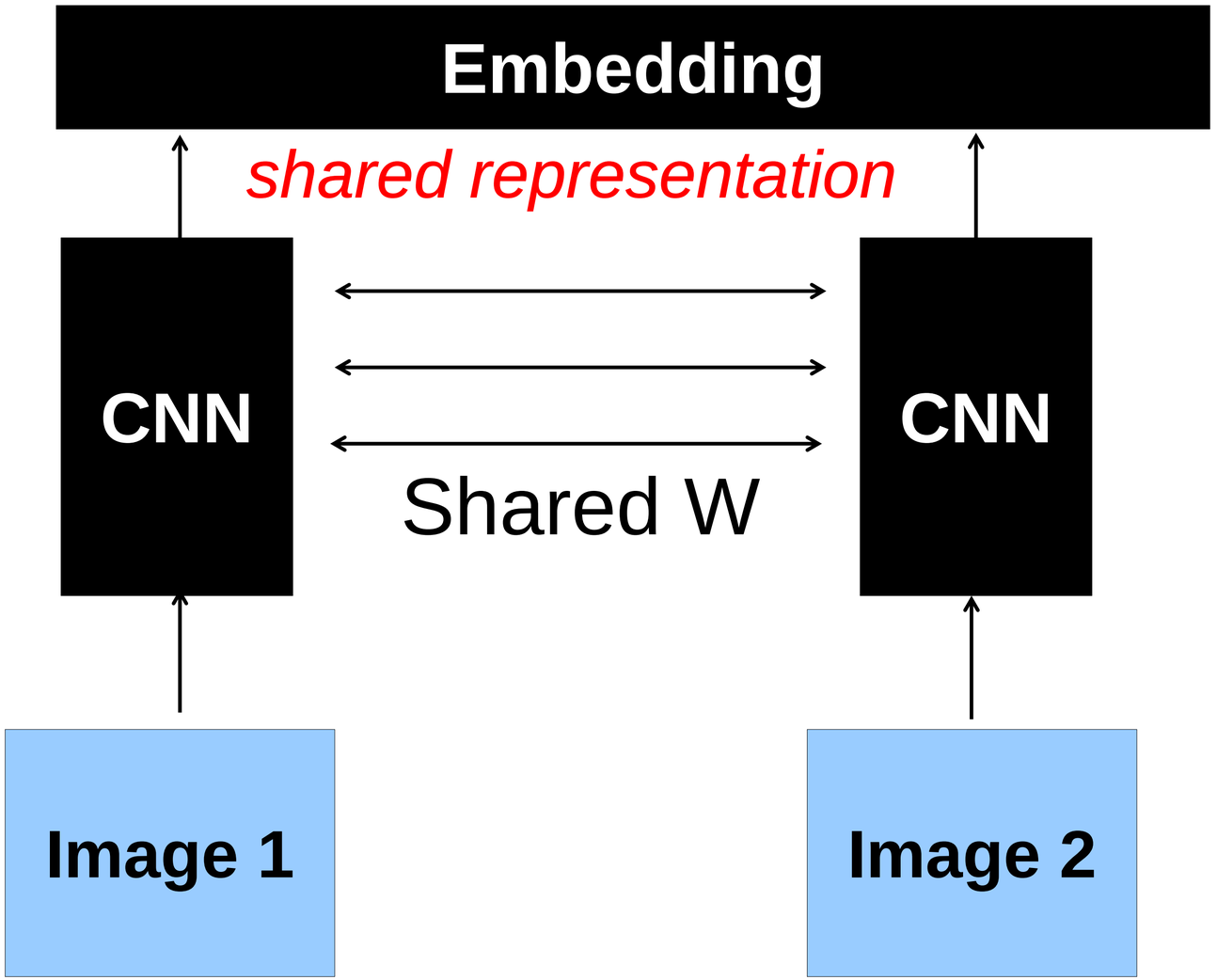
Fancier Neural Networks, fun applications, and scary cats

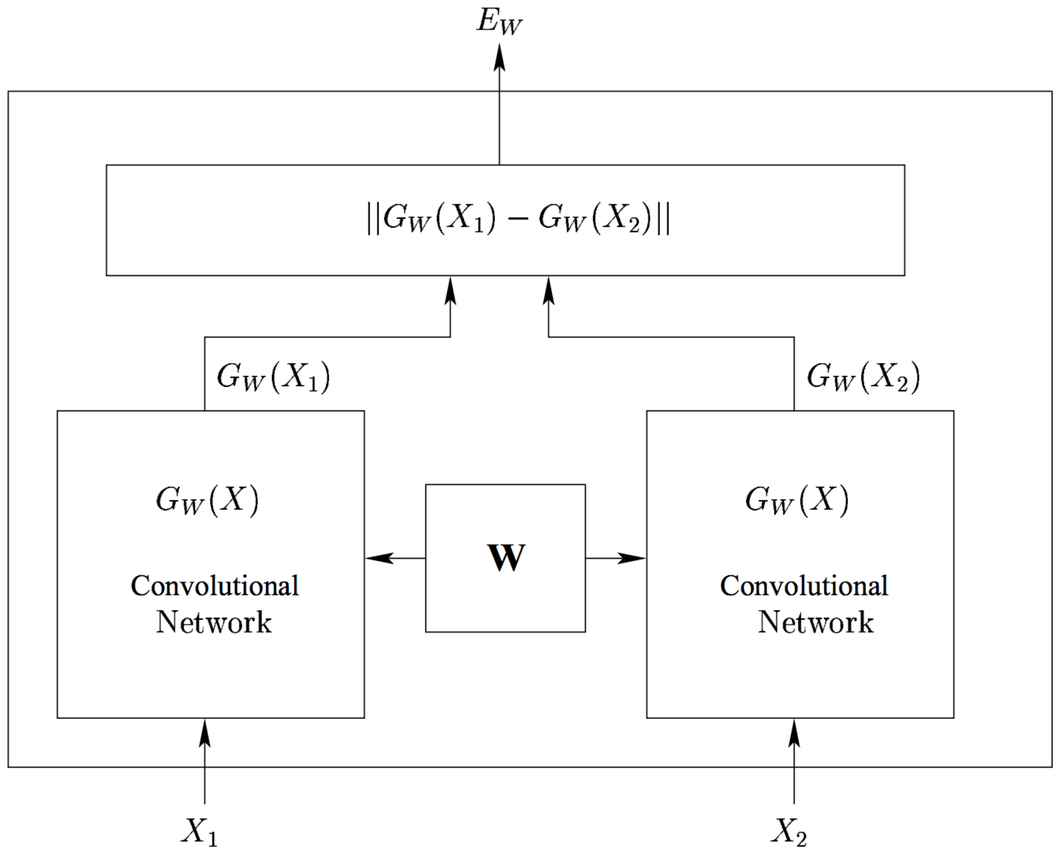
Many slides stolen from Sean Bell,
Jon Long, Carl Doersch, and Phillip Isola

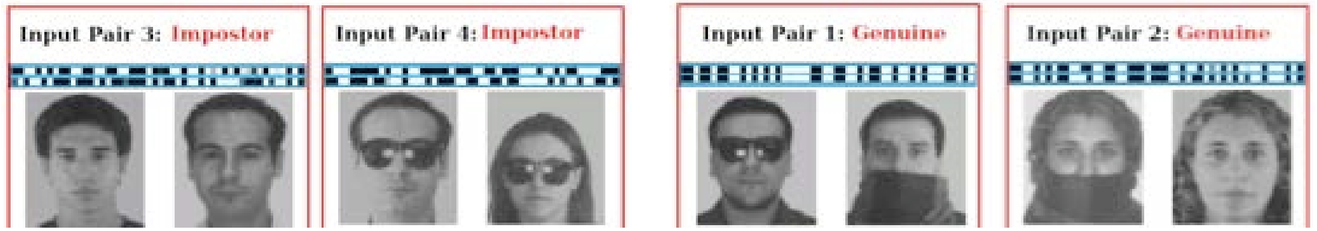
CS280, Spring 2017

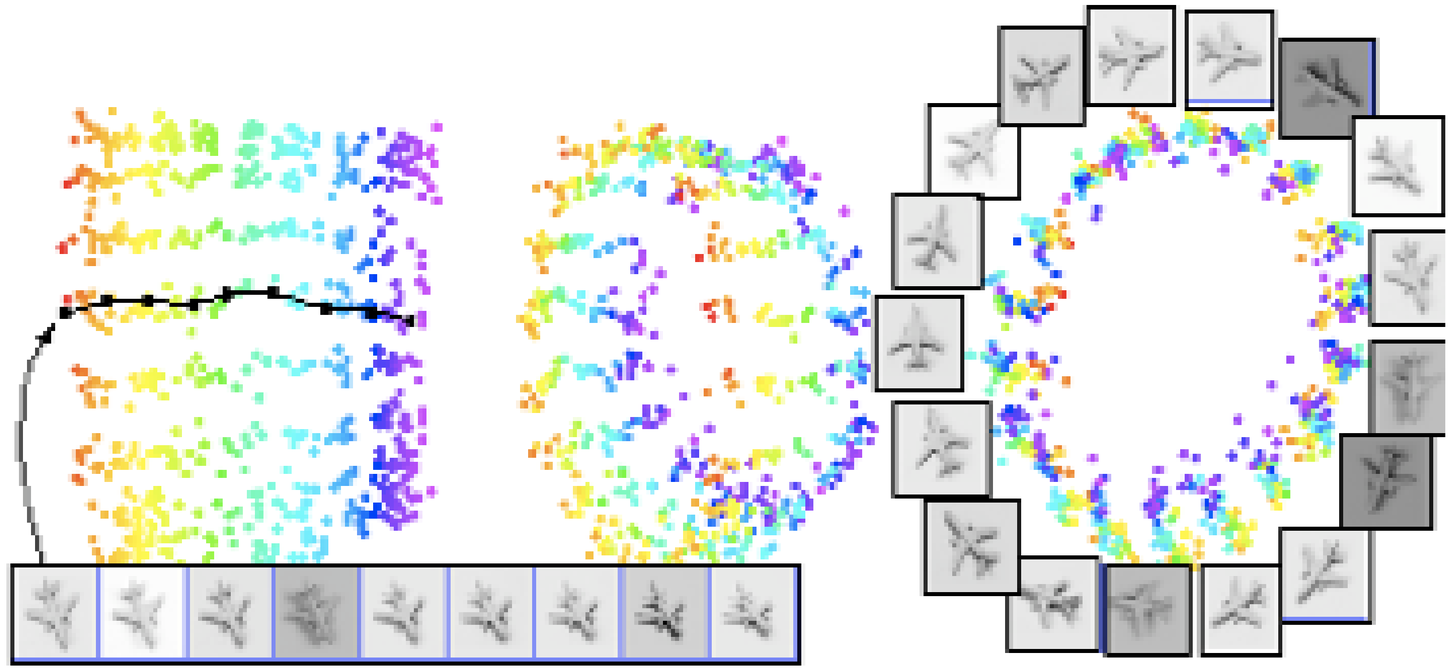
Fancier Architectures: “Siamese”



Siamese Network







Siamese Architecture
[Chopra 2005, Hadsell 2006]

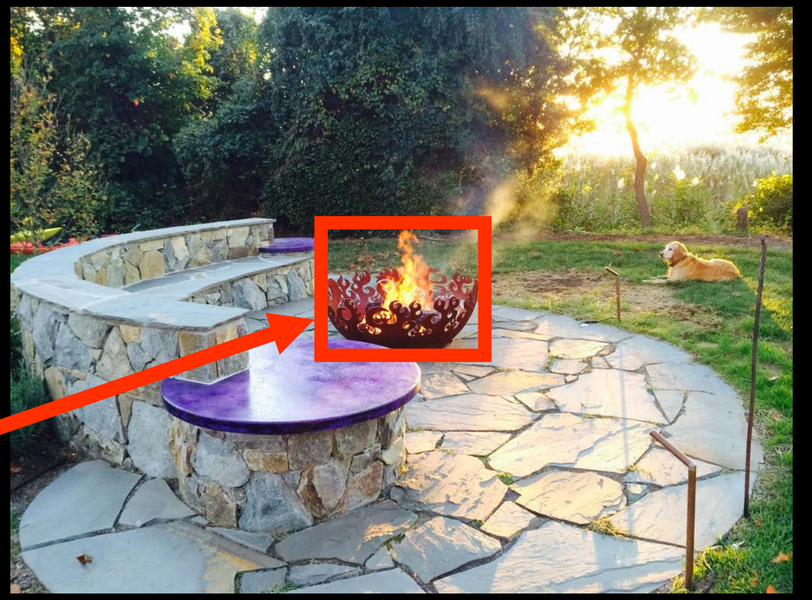
LEARNING VISUAL SIMILARITY FOR PRODUCT DESIGN WITH CONVOLUTIONAL NEURAL NETWORKS

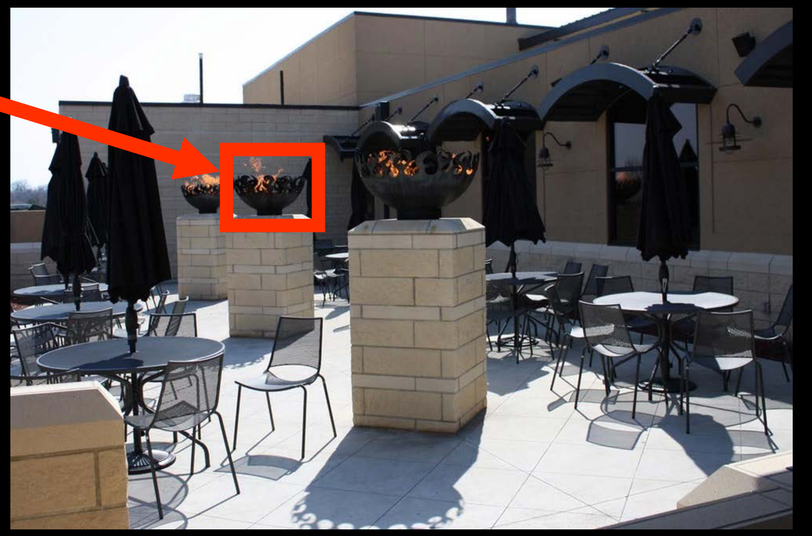
SEAN BELL AND KAVITA BALA
CORNELL UNIVERSITY

THE PROBLEM

(1) “What is this?”

(2) “Where is it used?”





Name: ”Great Bowl O’Fire Sculptural Fire Bowl”

Category: Fire pit

Sold by: John T. Unger, LLC

THE PROBLEM

(1) “What is this?”

(2) “Where is it used?”

Name: ”Great Bowl O’Fire Sculptural Fire Bowl”

Category: Fire pit

Sold by: John T. Unger, LLC

Challenge: determine whether these are the same product (different resolution, viewpoint, color, lighting, occlusions)

TWO KINDS OF IMAGES

Iconic

(From a product website)

In context

(Cropped from a scene photo)

PROJECTING INTO A JOINT EMBEDDING

Iconic

In context

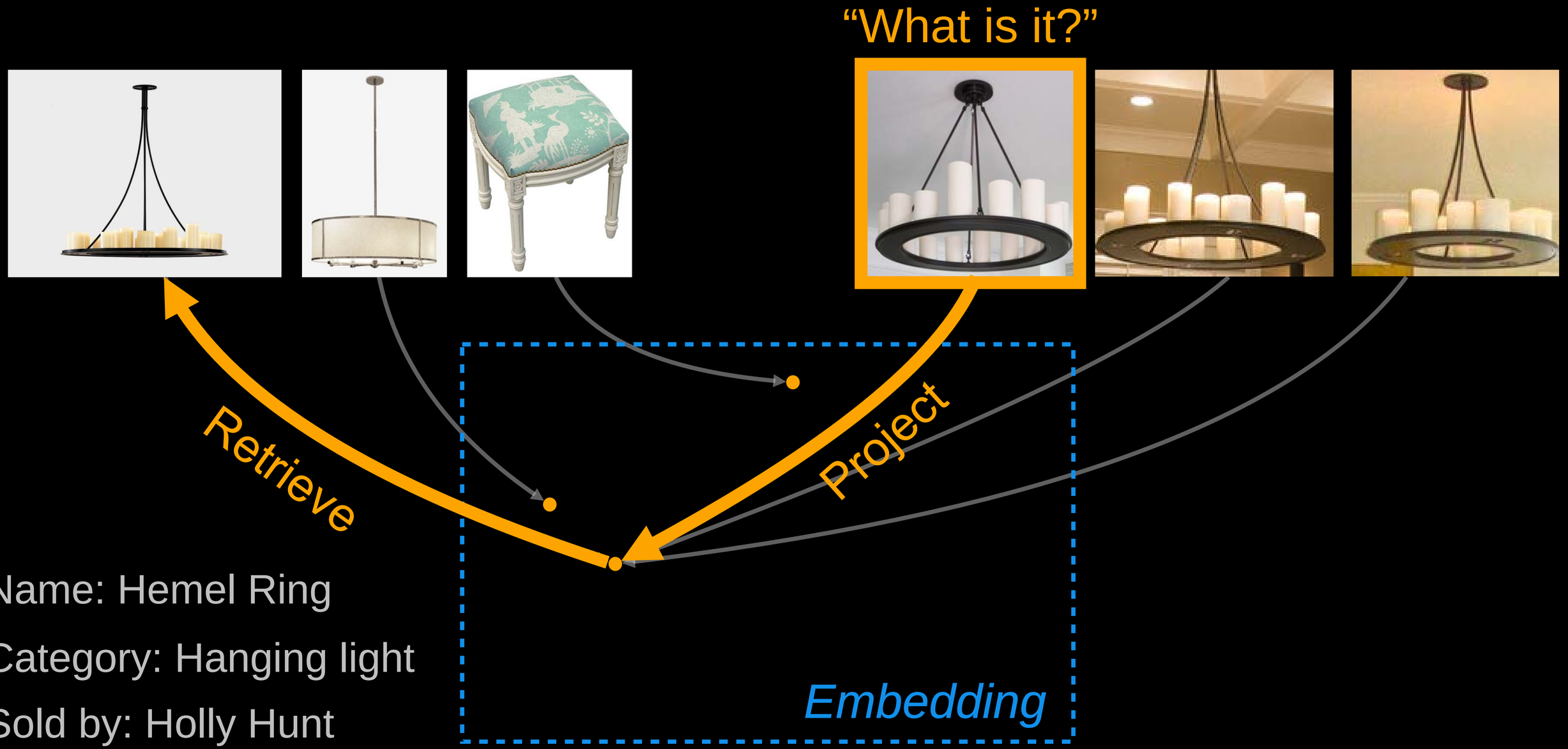





Project

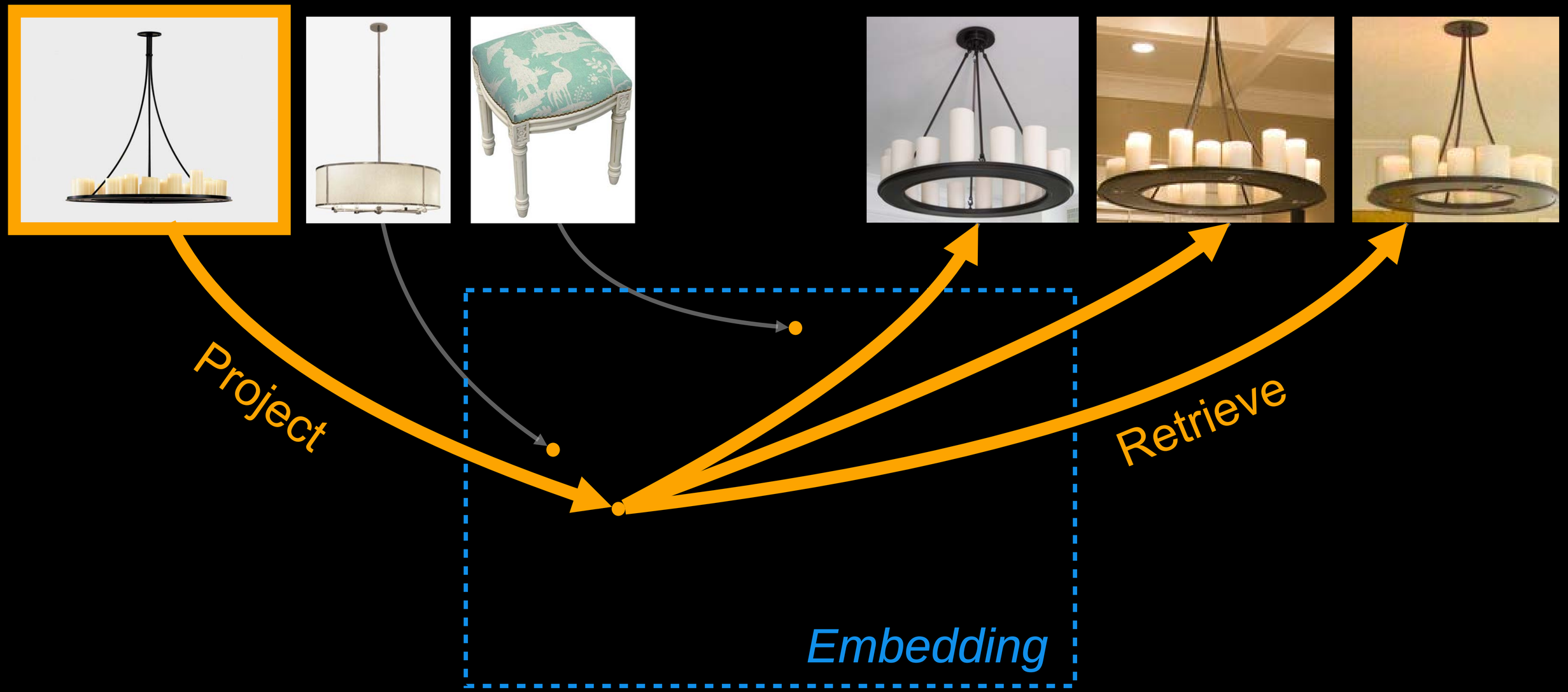
Embedding

SEARCH USING THE EMBEDDING

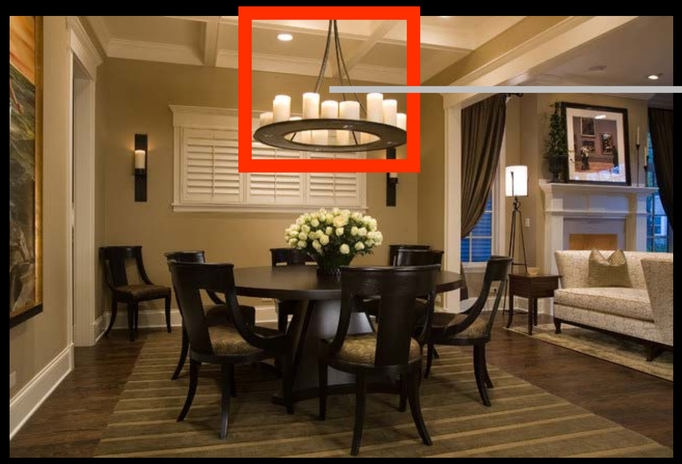


SEARCH USING THE EMBEDDING

"Where is it used?"



CONTRASTIVE LOSS: POSITIVE EXAMPLE



In context

Iconic (same)

CNN

Parameters θ

CNN

x_q

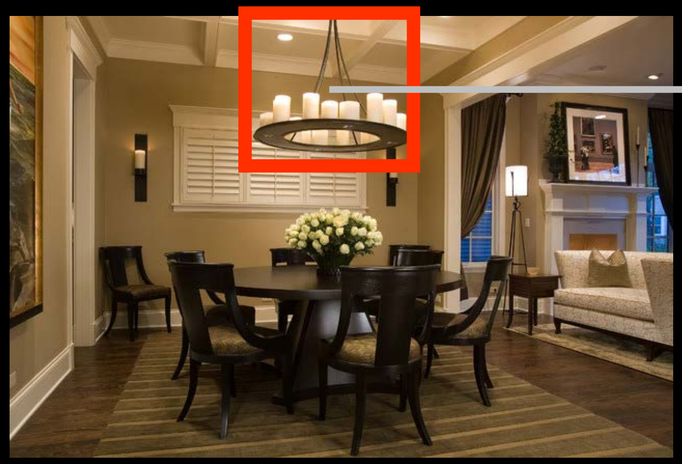
Loss L_p

x_p

Embedding

$$L_p(x_q, x_p) = ||x_q - x_p||_2^2$$

CONTRASTIVE LOSS: NEGATIVE EXAMPLE



In context

CNN

Parameters θ

Iconic (different)

CNN

Loss L_n

x_q

x_n

Margin *m*

Embedding

$$L_n(x_q, x_n) = \max\left(0, m^2 - ||x_q - x_n||_2^2\right)$$

CONTRASTIVE LOSS: ALL TOGETHER

$$L(\theta) = \underbrace{\sum_{(x_q, x_p)} L_p(x_q, x_p)}_{\text{Penalty for similar images that are far away}} + \underbrace{\sum_{(x_q, x_n)} L_n(x_q, x_n)}_{\text{Penalty for dissimilar images that are nearby}}$$

$$L_p(x_q, x_p) = ||x_q - x_p||_2^2$$

$$L_n(x_q, x_n) = \max\left(0, m^2 - ||x_q - x_n||_2^2\right)$$

Margin

Minimize L(θ) with stochastic gradient descent and momentum

[Chopra 2005, Hadsell 2006]

TRAINING PIPELINE

Image pairs

Stochastic Gradient Descent

θ

CNN Parameters

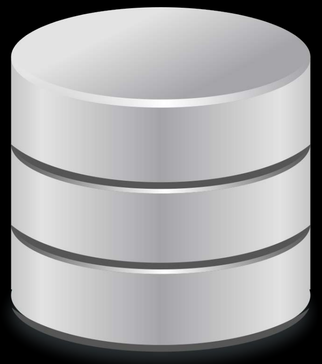


Image database

CNN

RESULTS: “WHAT IS IT?”

In context

RESULTS: “WHAT IS IT?”

In context

Iconic

Louis Ghost Chair with Arms

Living Room · Contemporary · 1 · View jpg · Outdoor Lounge Chairs · Eclectic · 44 · View jpg

2 boxes

Top 4 results:

Victoria Ghost Chair by Philippe Starck for Kartell | Design Within Reach

Dining Chairs · Modern · 15 · View jpg · View detail

0.019

Kartell | Victoria Ghost Chair, Set of 4

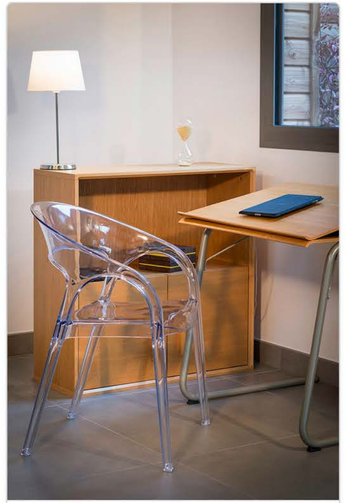
Dining Chairs · Modern · View jpg · View detail

0.023

Kartell | Louis Ghost Chair

Dining Chairs · Contemporary · View jpg · View detail

0.024



Chaises

Office Chairs · Unknown style · View jpg · View detail

0.025

RESULTS: “WHAT IS IT?”

In context

RESULTS: “WHAT IS IT?”

In context

Iconic

Tolix Stool “Tabouret”

Kitchen | Contemporary | 1 | View jpg | Bar Stools And Counter Stools | Industrial | 305 | View jpg

2 boxes

Top 4 results:

East Berlin District Metal Barstool

Serving Dishes And Platters | Modern | View jpg | View detail

0.046

Amelia Metal Cafe Barstool in Orange - Set of

Bar Stools And Counter Stools | Contemporary | 3 | View jpg | View detail

0.049

Amelia Metal Cafe Barstool (Set of 2), Red

Bar Stools And Counter Stools | Traditional | 2 | View jpg | View detail

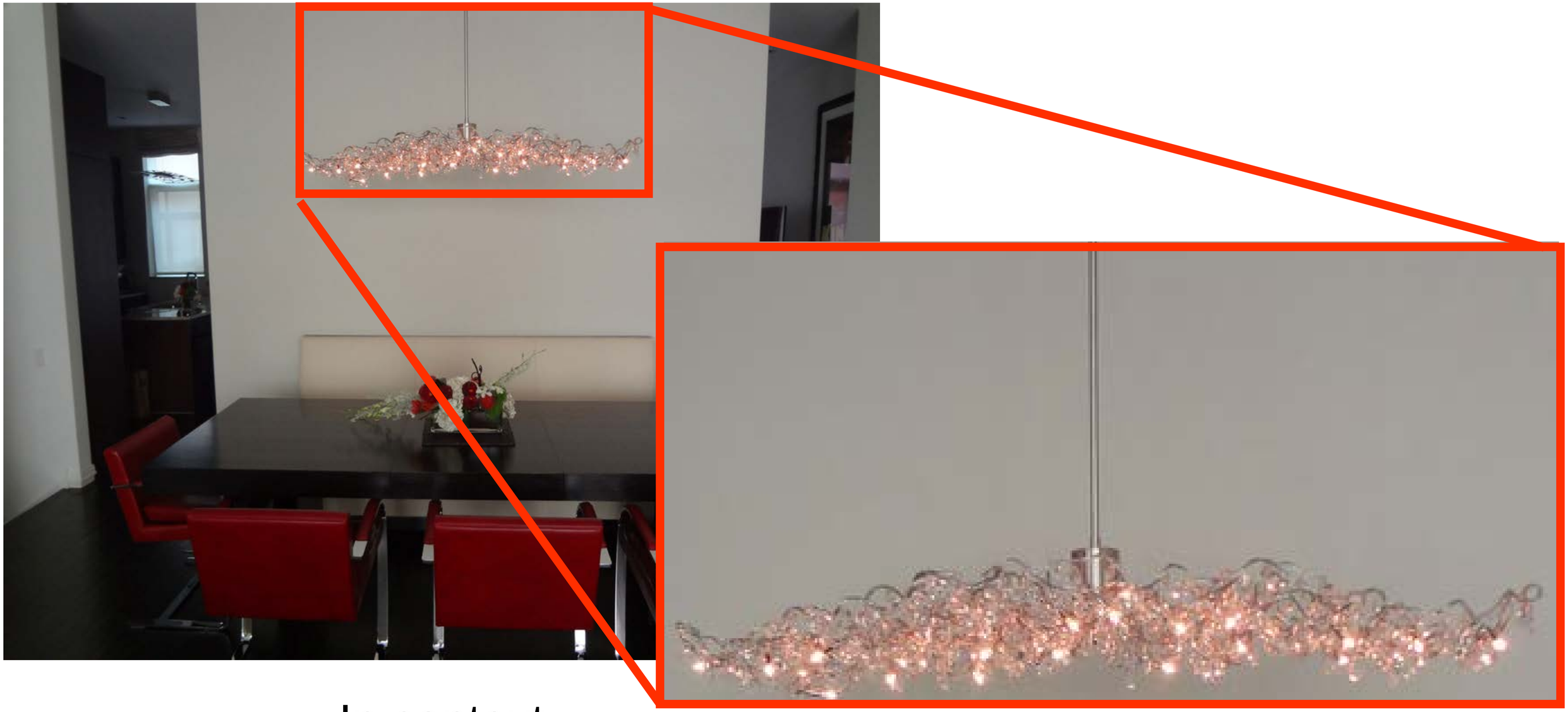
0.050

Winsome Dubliner 30 in. Bar Stool - Set of 2

Bar Stools And Counter Stools | Modern | View jpg | View detail

0.050

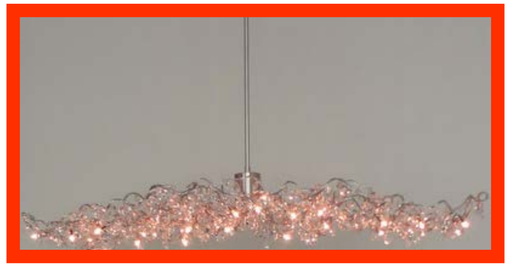
RESULTS: “WHAT IS IT?”



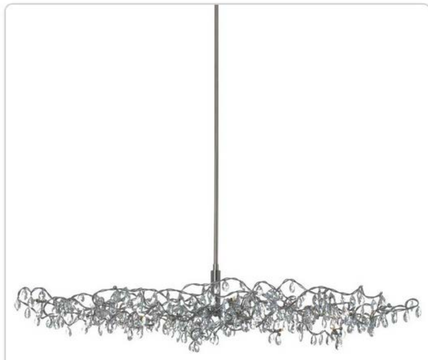
In context

RESULTS: “WHAT IS IT?”

In context



Iconic



Tiara Oval Suspension by Harco Loor

Dining Room | Contemporary | 1 | View jpg | Wall Sconces | Unknown style | 1 | View jpg

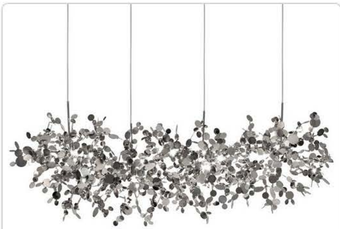
2 boxes

Top 4 results:

HARCO LOOR Tiara Chandelier

Chandeliers | Contemporary | 1 | View jpg | View detail

0.035



Argent N92S Suspension Light

Bathroom Vanity Lighting | Modern | View jpg | View detail

0.038

Eurofase 25620-016 Divo 9 Light Pendant in Nickel 25620-016

Pendant Lighting | Modern | View jpg | View detail

0.041

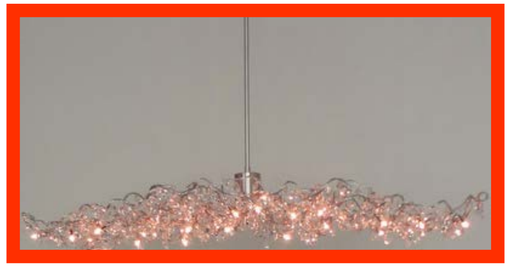
Eurofase 25618-013 Divo 6 Light Pendant in Nickel 25618-013

Pendant Lighting | Modern | View jpg | View detail

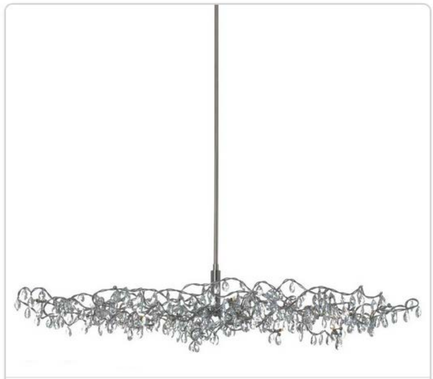
0.047

COMPARISON: TRAINED ONLY ON CATEGORIES

In context



Iconic



Tiara Oval Suspension by Harco Loor

Dining Room | Contemporary | 1 | View jpg | Wall Sconces | Unknown style | 1 | View jpg

2 boxes

Top 4 results:

'Apollon' Black Shaded 6-light Crystal Chandelier

Chandeliers | Contemporary | View jpg | View detail

1792.880

Authentic Deer and Elk Antler Banquet Table Chandelier

Chandeliers | Unknown style | View jpg | View detail

1895.330

Harco Loor | Tiara Oval HL 15 Suspension Light

Pendant Lighting | Modern | View jpg | View detail

1911.520

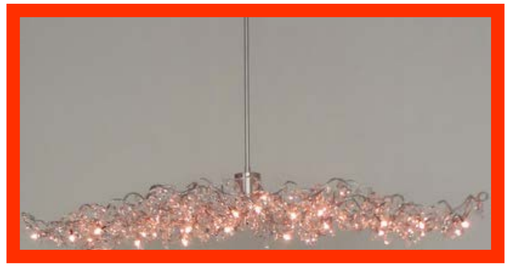
Vermeer Hexagonal Pendant

Pendant Lighting | Modern | View jpg | View detail

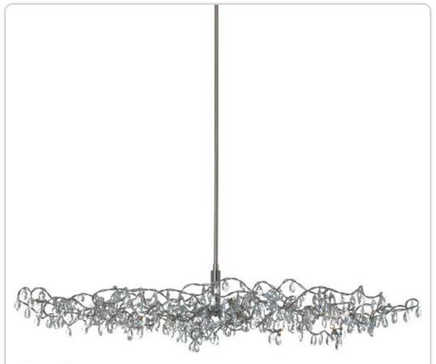
2035.000

COMPARISON: TRAINED ONLY ON IMAGENET

In context



Iconic



Tiara Oval Suspension by Harco Loor

Dining Room | Contemporary | 1 | View jpg | Wall Sconces | Unknown style | 1 | View jpg

2 boxes

Top 4 results:

MERGING TOGETHER

Artwork | Contemporary | View jpg | View detail

498.493

"Elephant View" Artwork

Paintings | Contemporary | View jpg | View detail

499.135

Terzani Argent N92S Chandelier

Chandeliers | Modern | 2 | View jpg | View detail

503.232

Mensa Hanging Light Fixture

Chandeliers | Contemporary | View jpg | View detail

514.867

RESULTS: FAILURE CASE

In context

RESULTS: FAILURE CASE

In context

Iconic

Porta Romana Bianca Pendant Light

Staircase | Transitional | 1 | View jpg | Pendant Lighting | Contemporary | 2 | View jpg

2 boxes

Top 4 results:

Murray Feiss Khloe Modern Contempo Mini Pendant

Pendant Lighting | Modern | View jpg | View detail

0.066

Uttermost Metauro Wood Desk Lamp

Desk Lamps | Contemporary | View jpg | View detail

0.067

Cycling Blur Wall Mural - 42 Inches W x 40 Inches H

Prints And Posters | Contemporary | 3 | View jpg | View detail

0.069

Metauro Wood Desk Lamp

Desk Lamps | Contemporary | View jpg | View detail

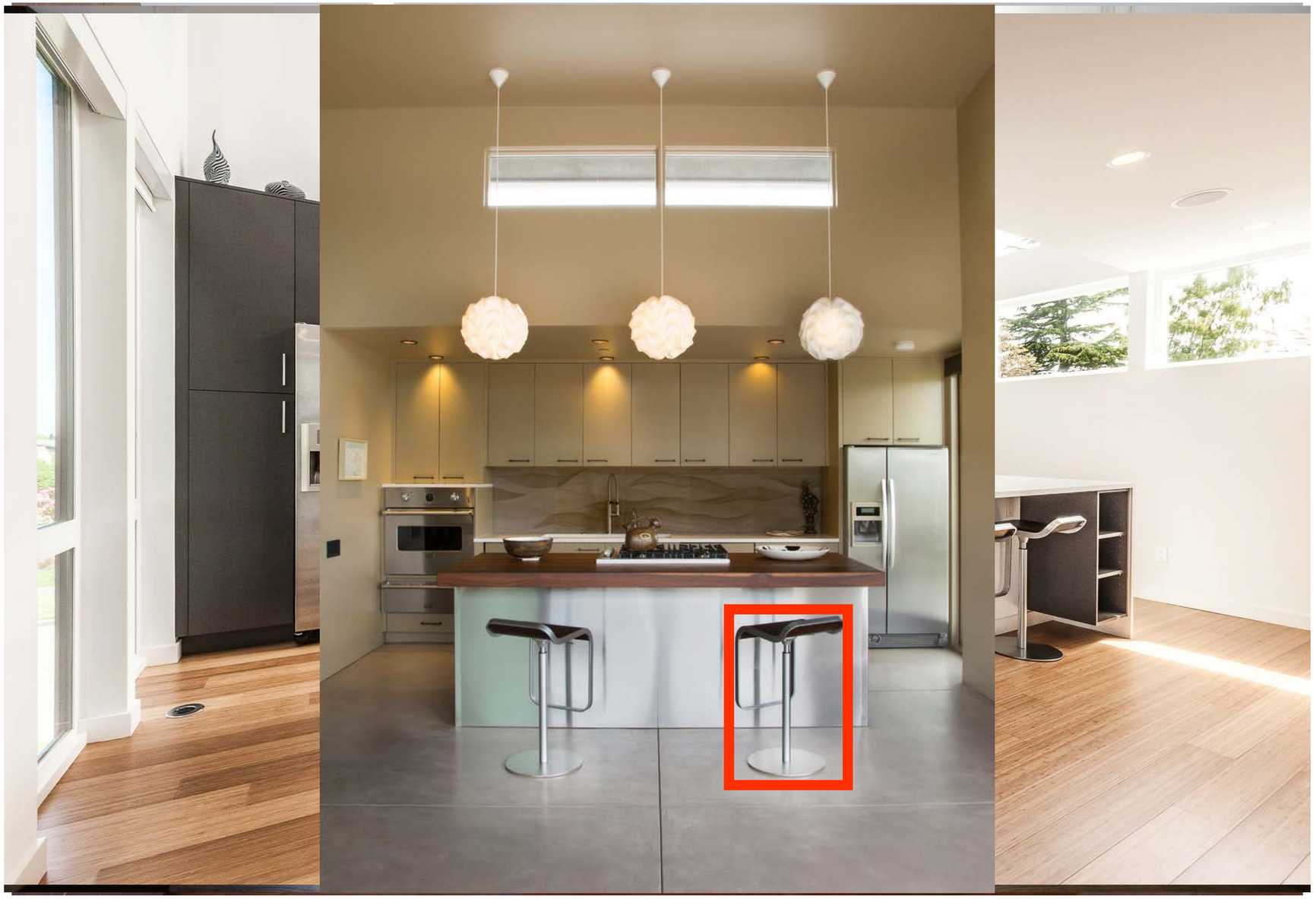
0.069

RESULTS: “WHERE IS IT USED?”

“Maskros Pendant Lamp”

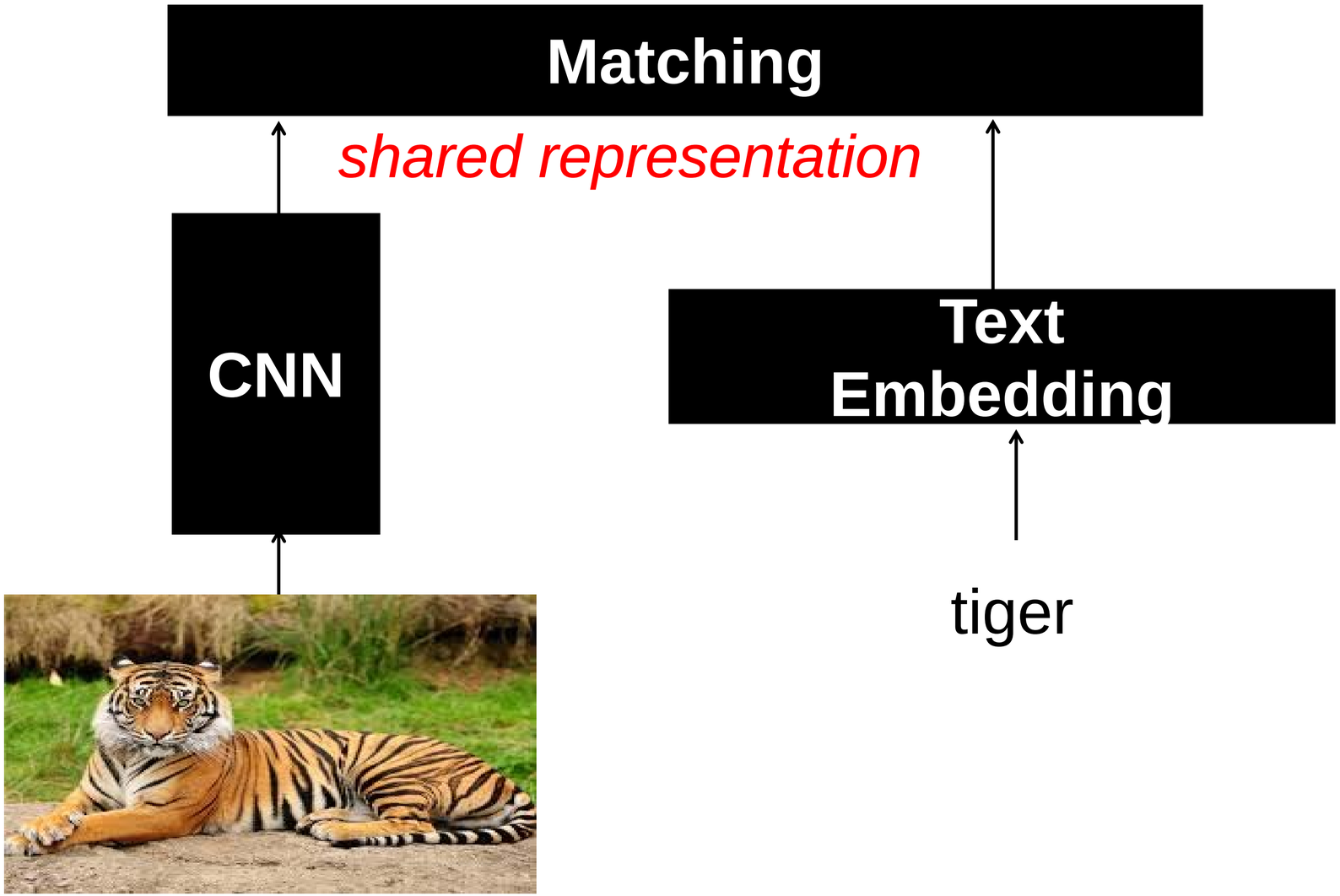
RESULTS: “WHERE IS IT USED?”

"LEM Piston Stool | Design Within Reach”



SEARCHING ACROSS CATEGORIES

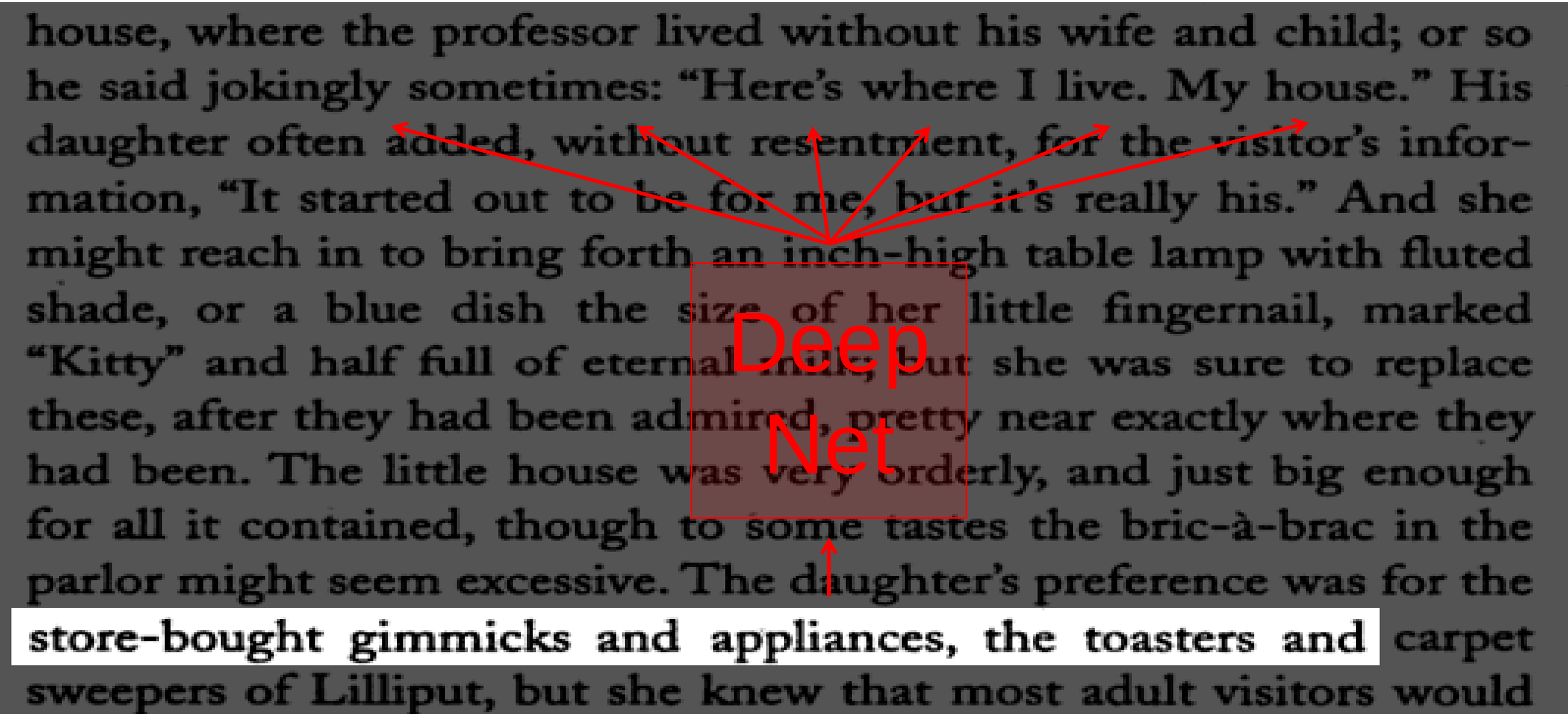
Fancier Architectures: Multi-Modal



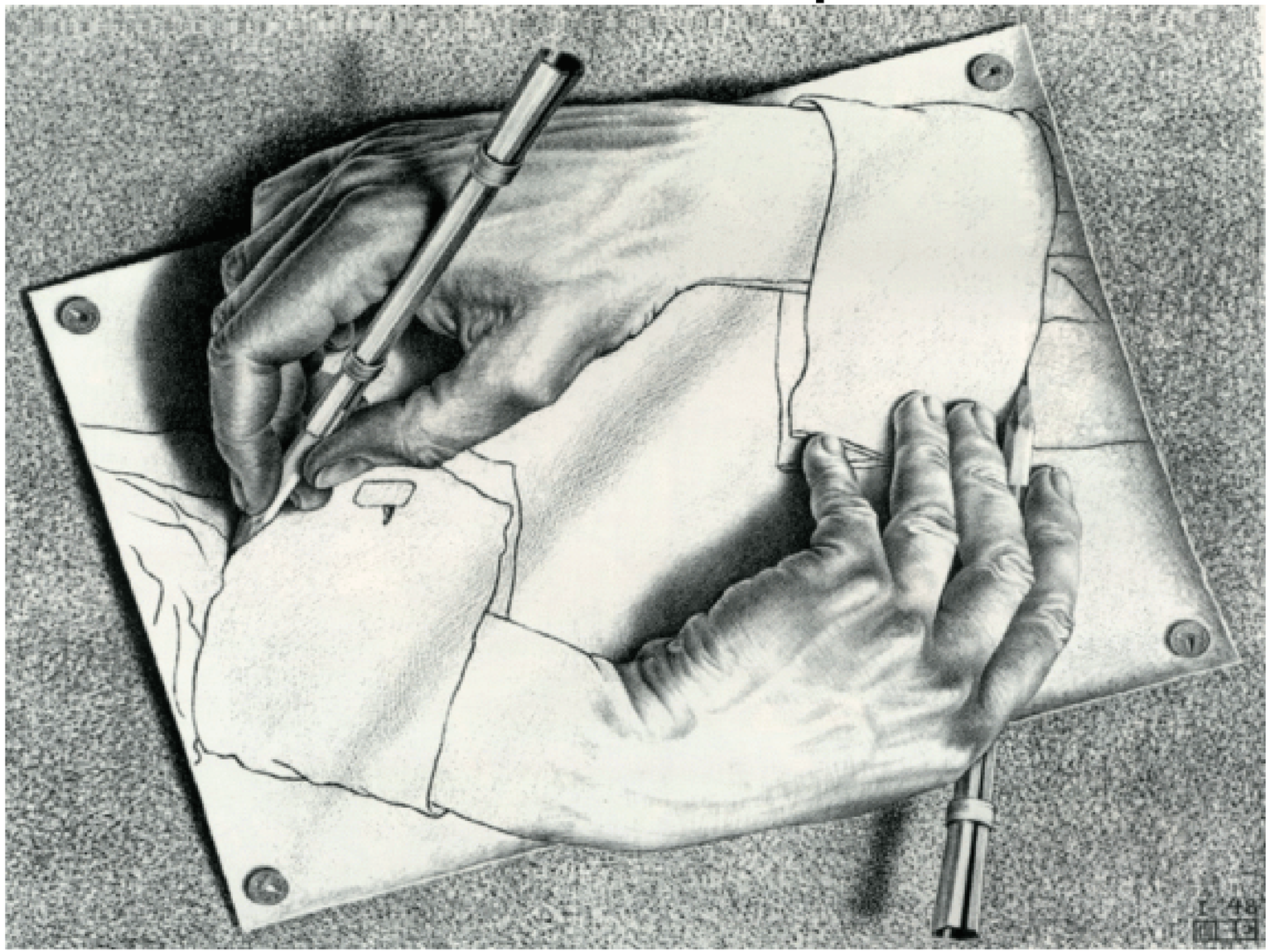
Frome et al. “Devise: a deep visual semantic embedding model” NIPS 2013

Context as Supervision

[Collobert & Weston 2008; Mikolov et al. 2013]



Self-supervised Learning: data as its own supervision



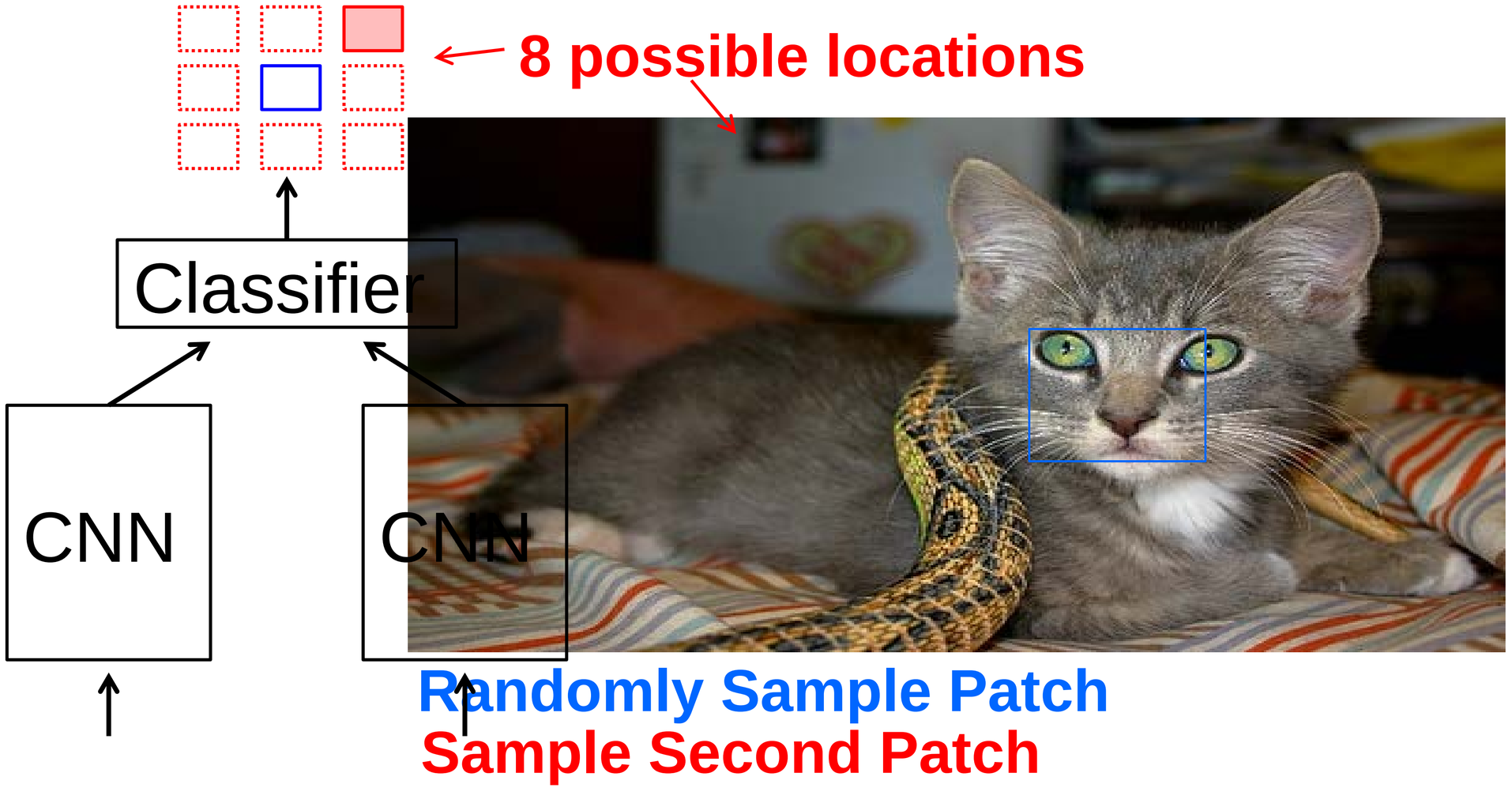
Context Prediction for Images

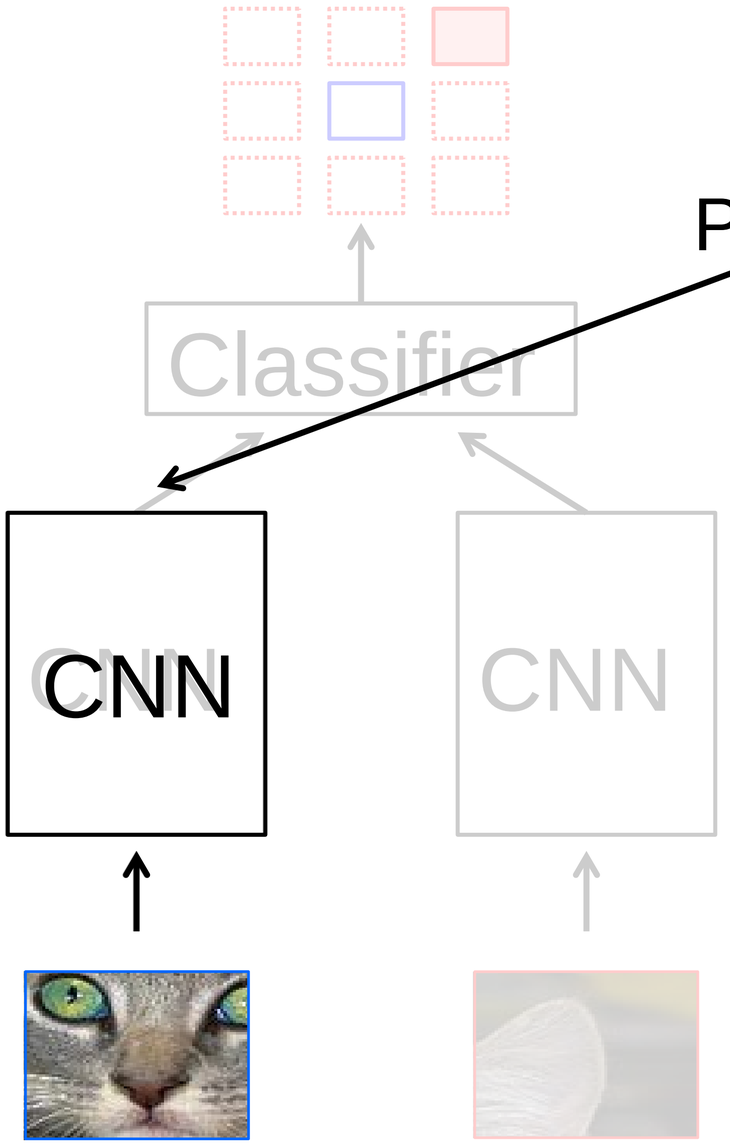
A

B

Semantics from a non-semantic task

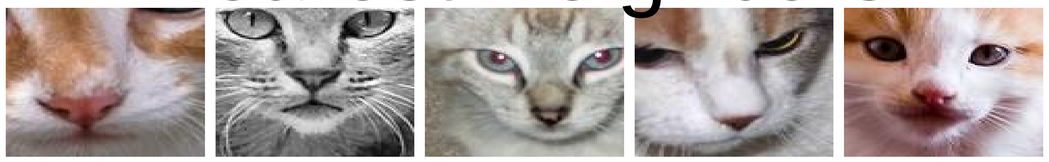
Relative Position Task





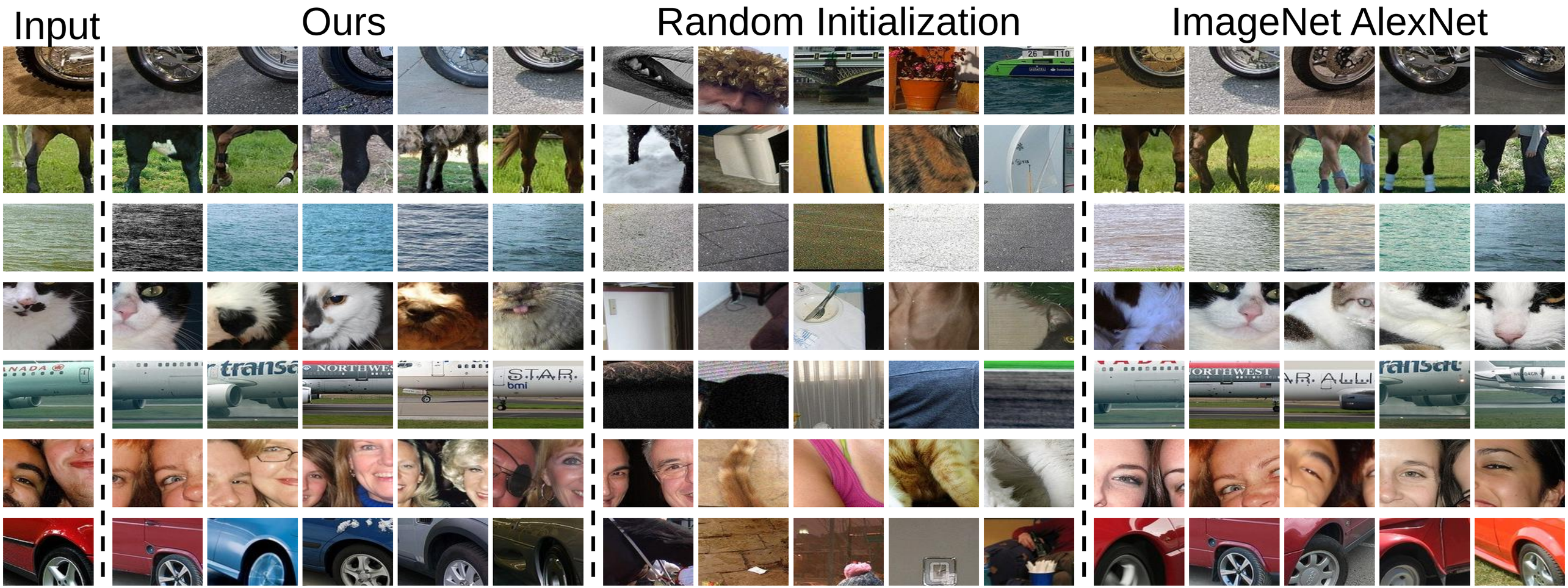
Input

Nearest Neighbors



Correspondence ***across*** instances!

What is learned?

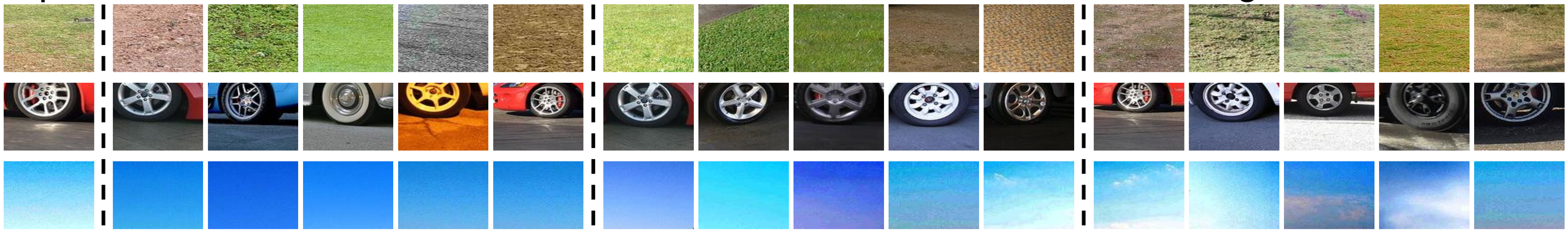


Still don’t capture everything

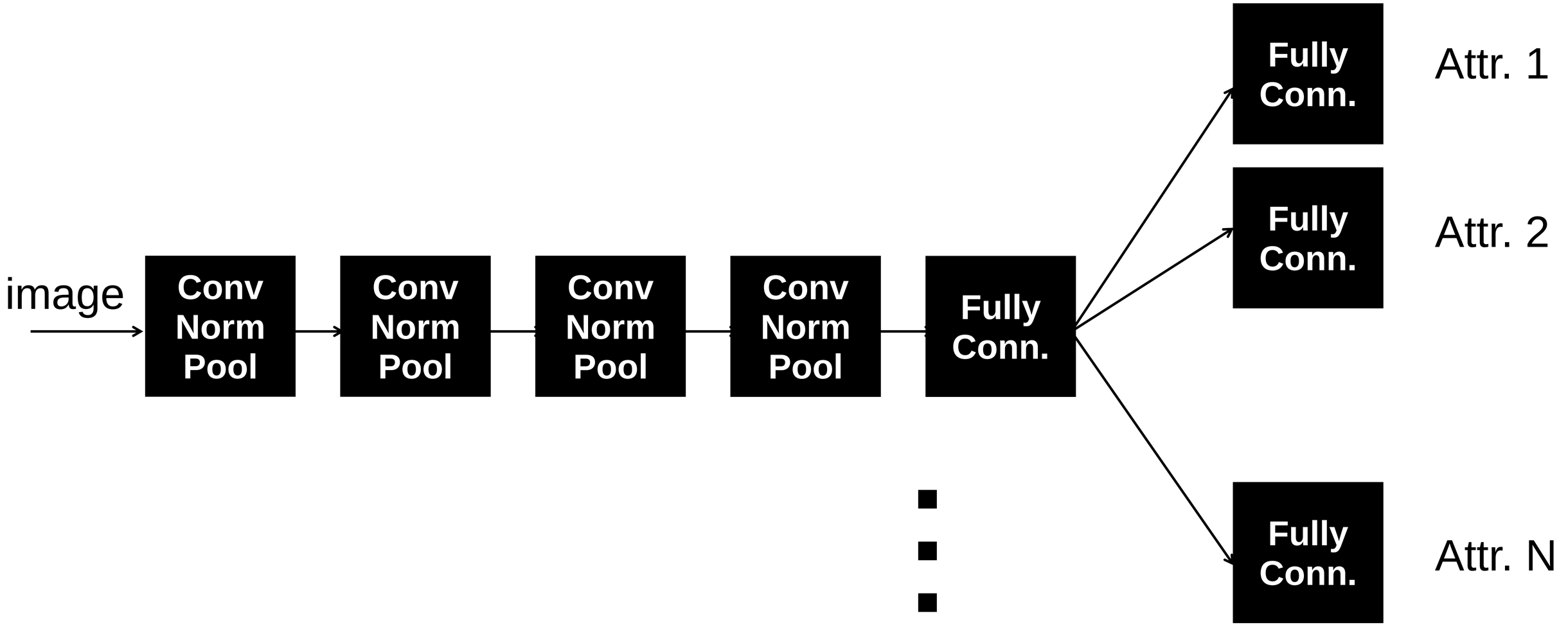
Input | Ours | Random Initialization | ImageNet AlexNet

You don’t always need to learn!

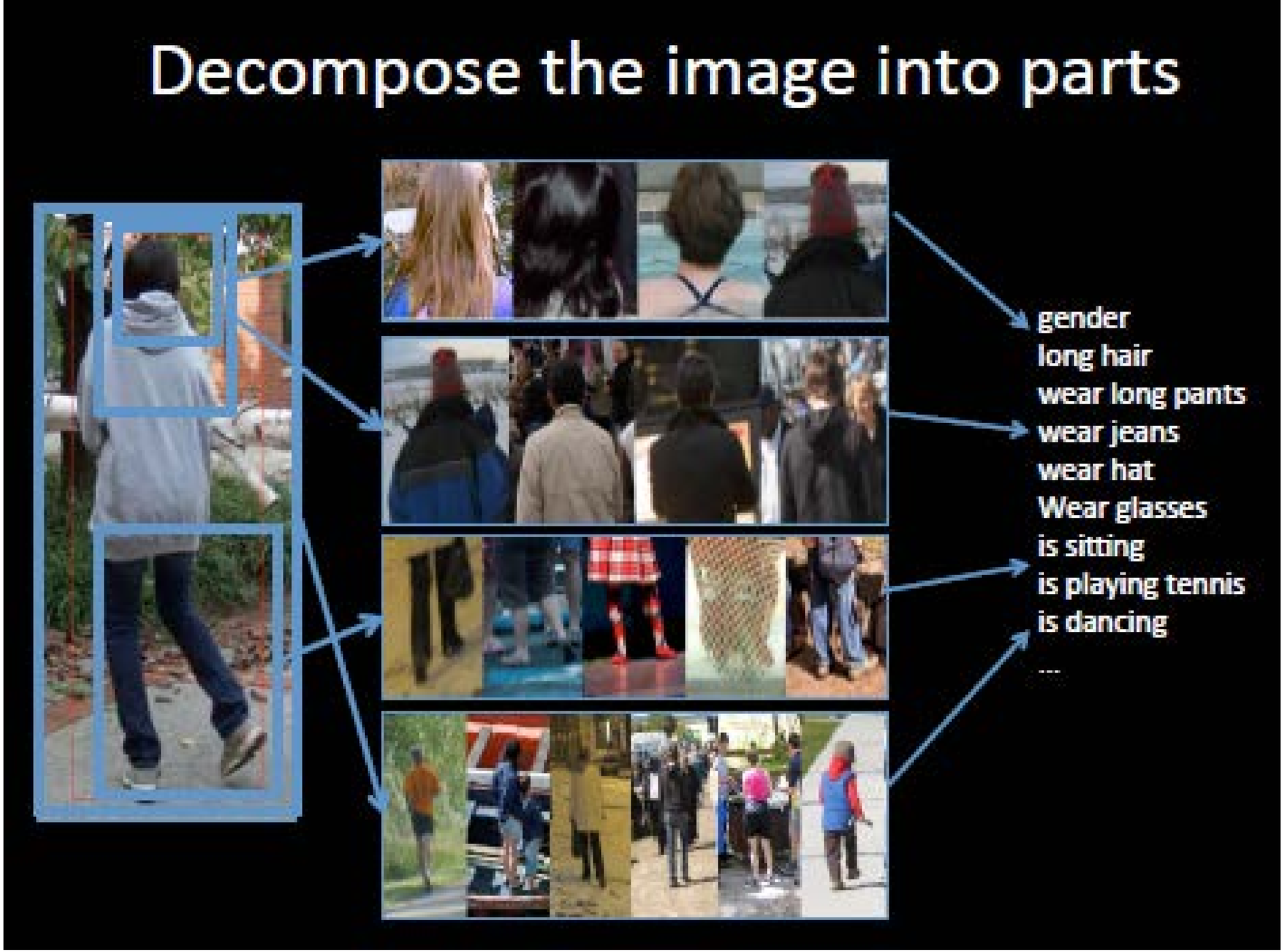
Input | Ours | Random Initialization | ImageNet AlexNet



Fancier Architectures: Multi-Task

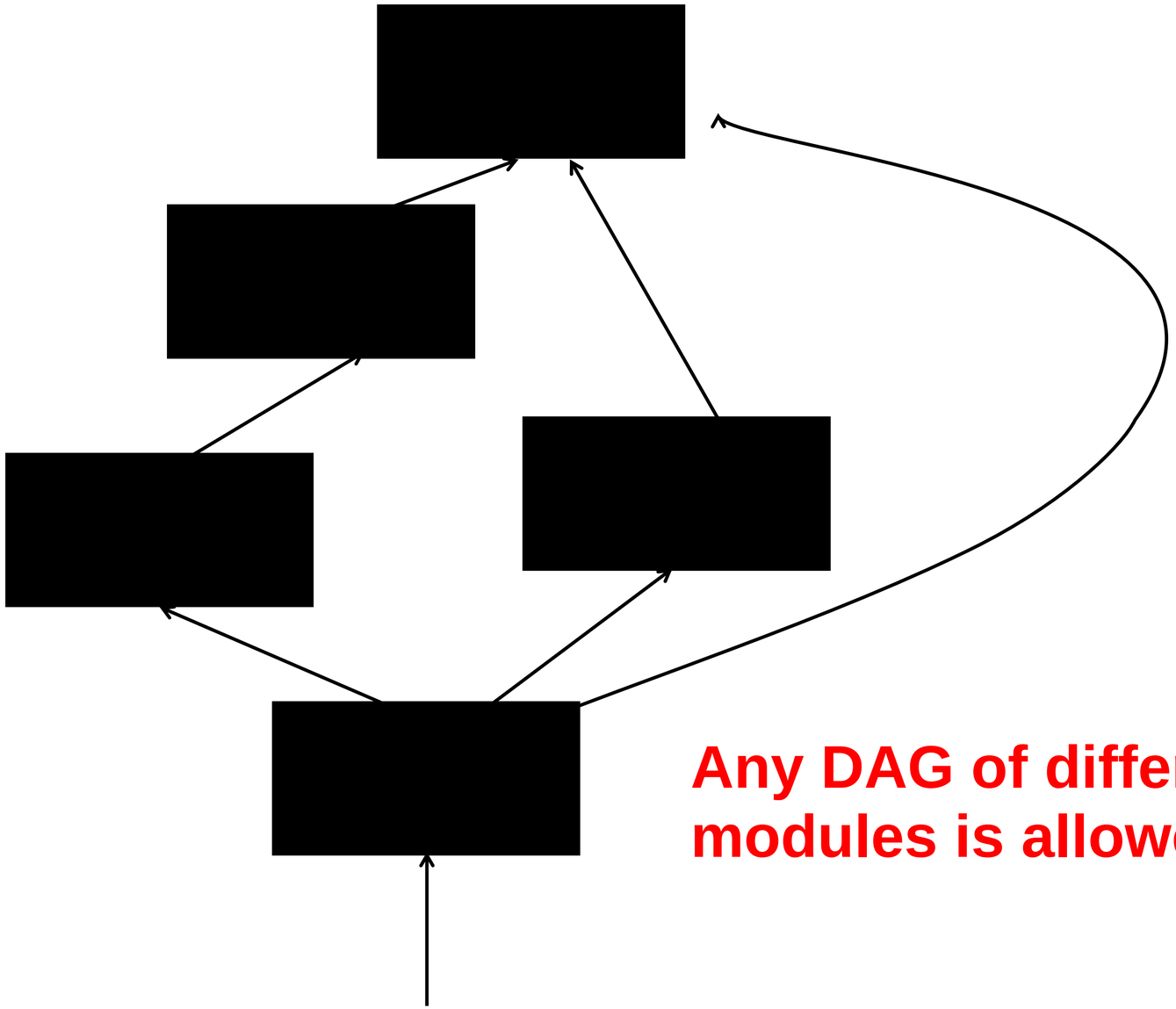


Zhang et al. “PANDA..” CVPR 2014



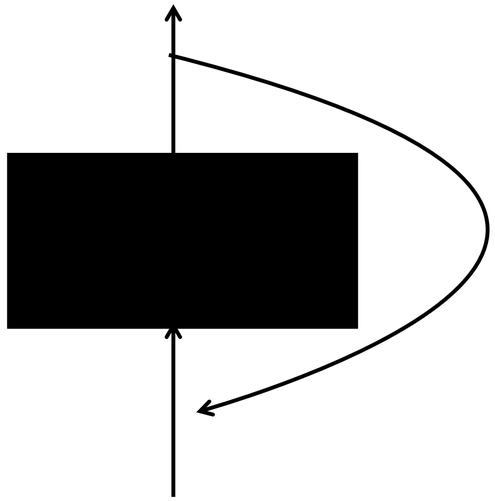
Zhang et al. “PANDA..” CVPR 2014

Fancier Architectures: Generic DAG



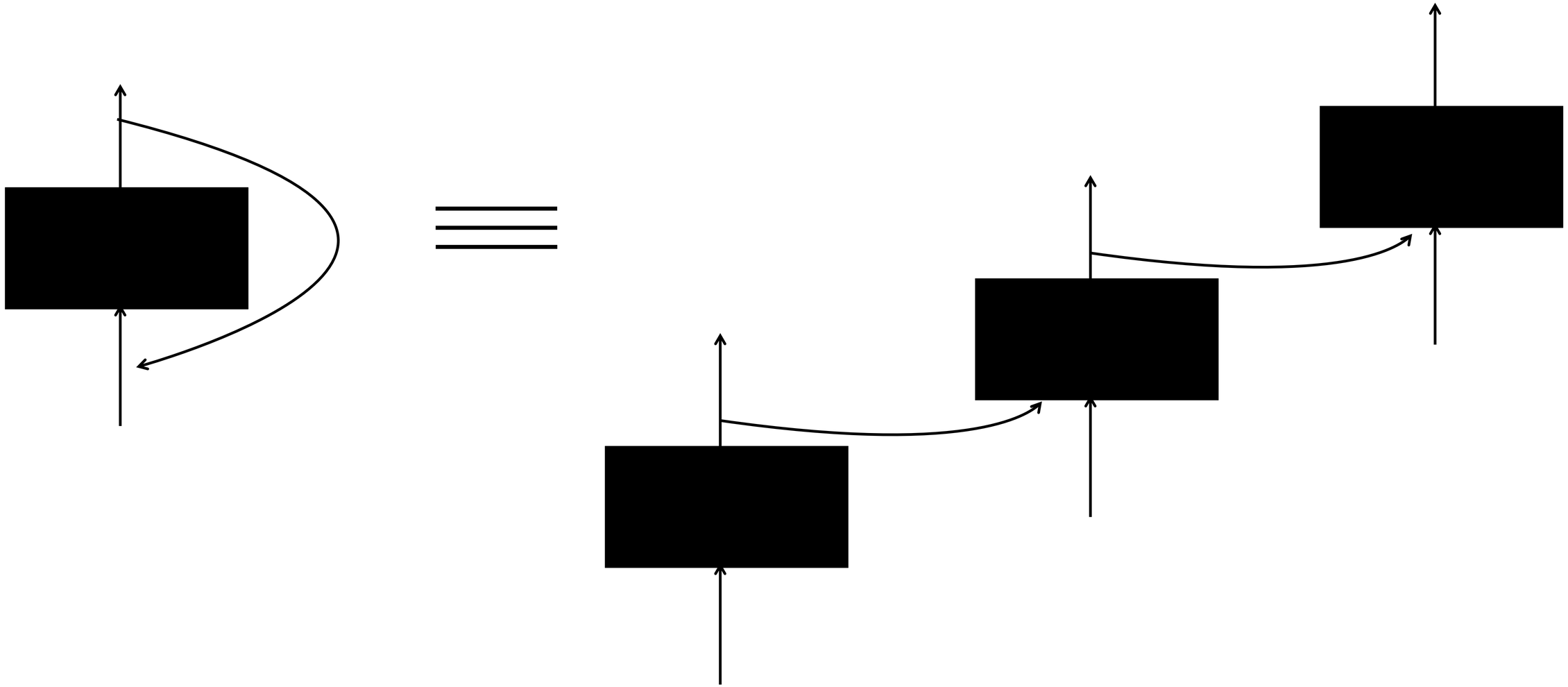
Any DAG of differentialble modules is allowed!

Fancier Architectures: Generic DAG



Fancier Architectures: Recurrent NN

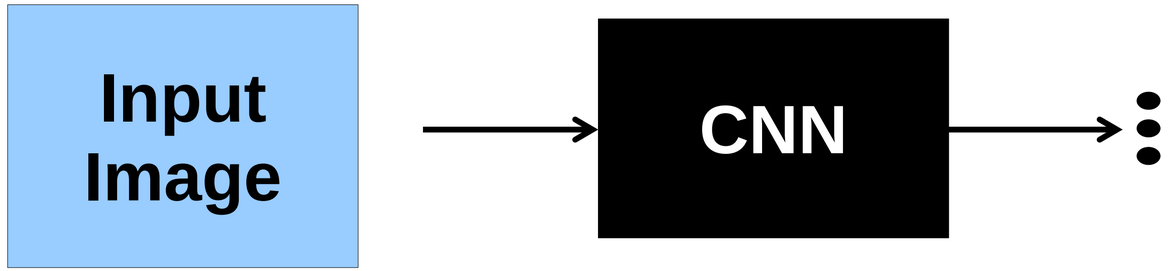
If there are cycles, one needs to un-roll it.



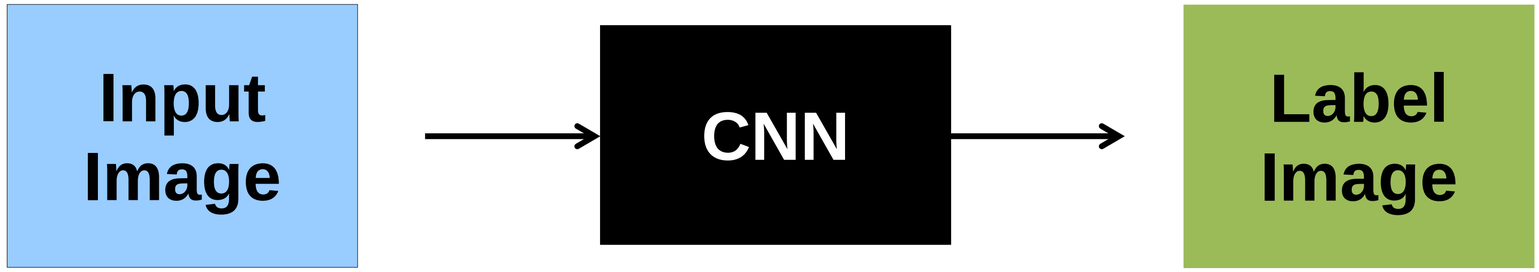
Pinheiro, Collobert “Recurrent CNN for scene labeling” ICML 2014
Graves “Offline Arabic handwriting recognition..” Springer 2012

Fancier Architectures: Fully-Convolutional

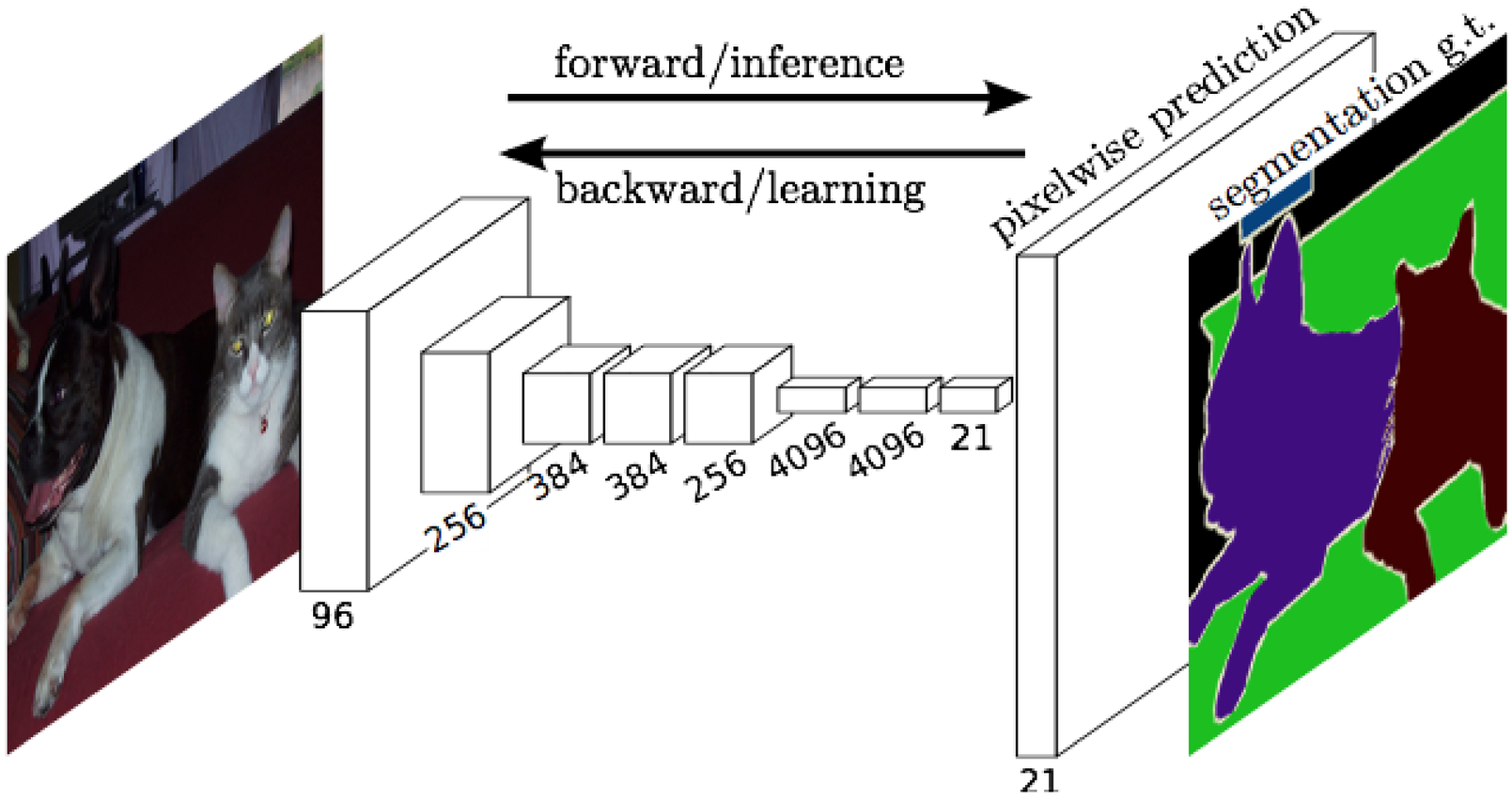
Classification ConvNet (e.g. ImageNet)



Per-pixel Labeling ConvNet



Fully Convolutional Networks for Semantic Segmentation

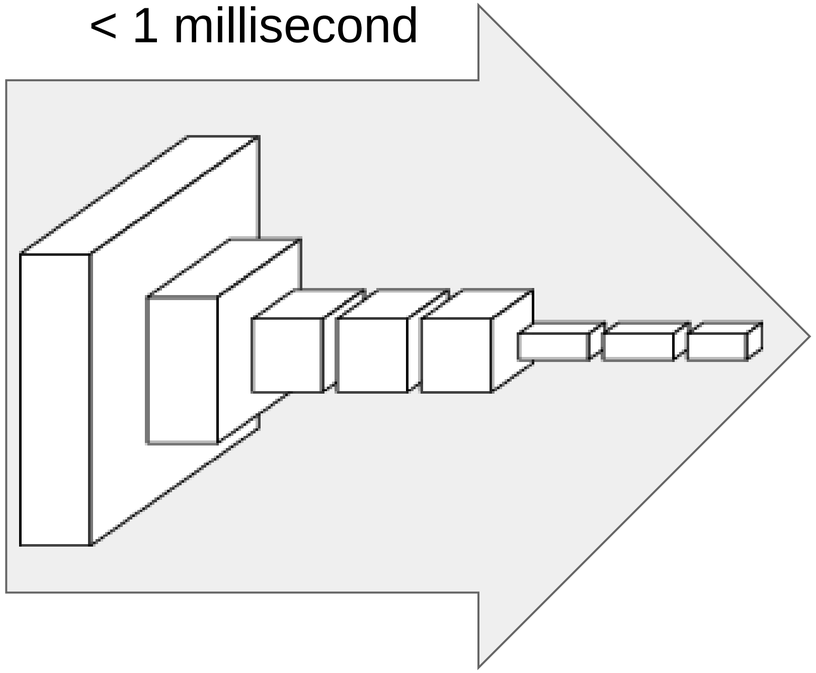


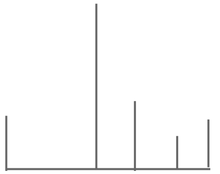
Evan Shelhamer* Jonathan Long* Trevor Darrell

UC Berkeley in CVPR'15, PAMI'16

convnets perform classification

< 1 millisecond



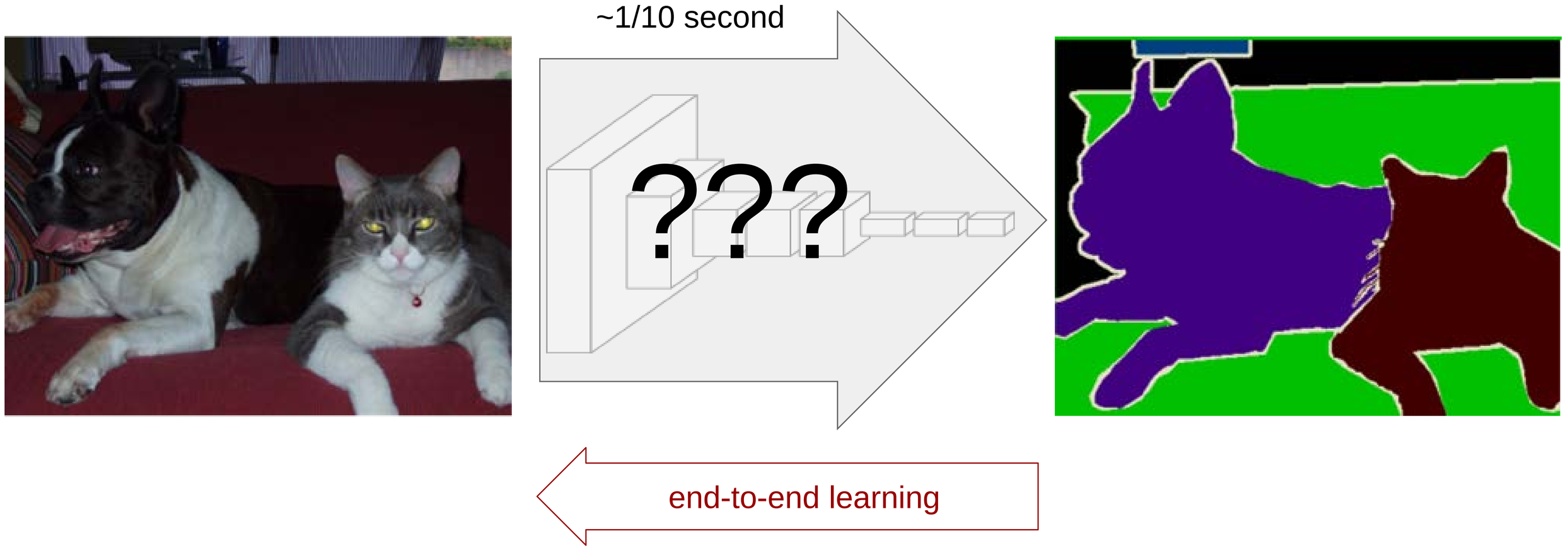


“tabby cat”

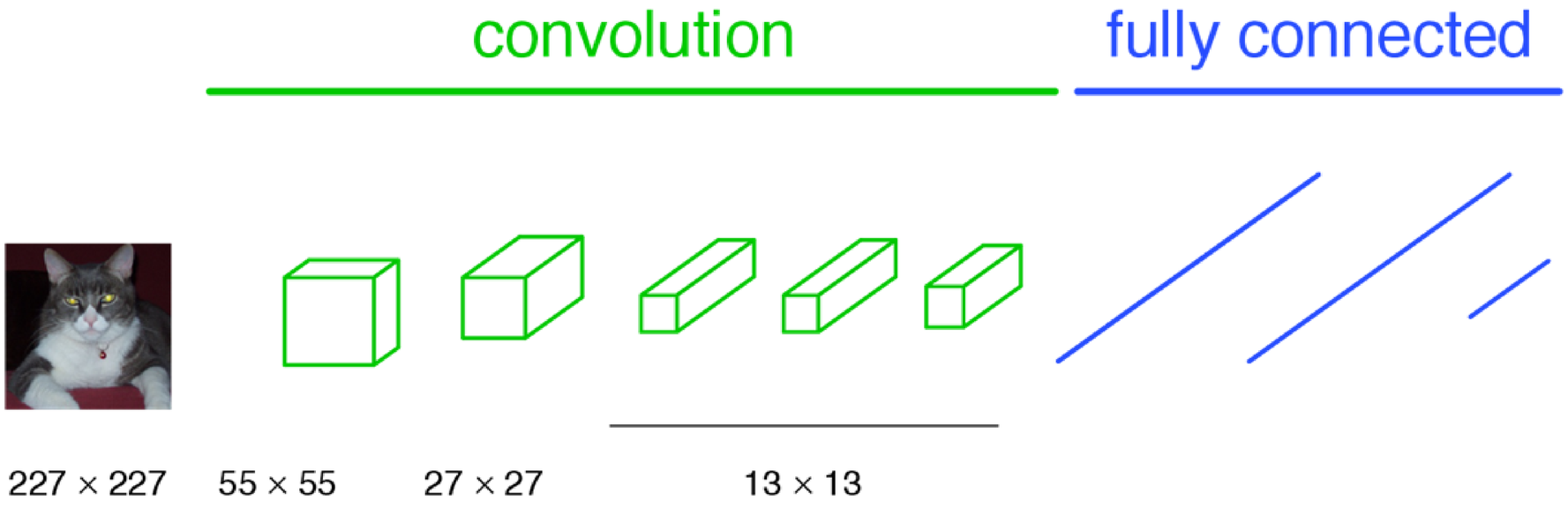
1000-dim vector

end-to-end learning

lots of pixels, little time?

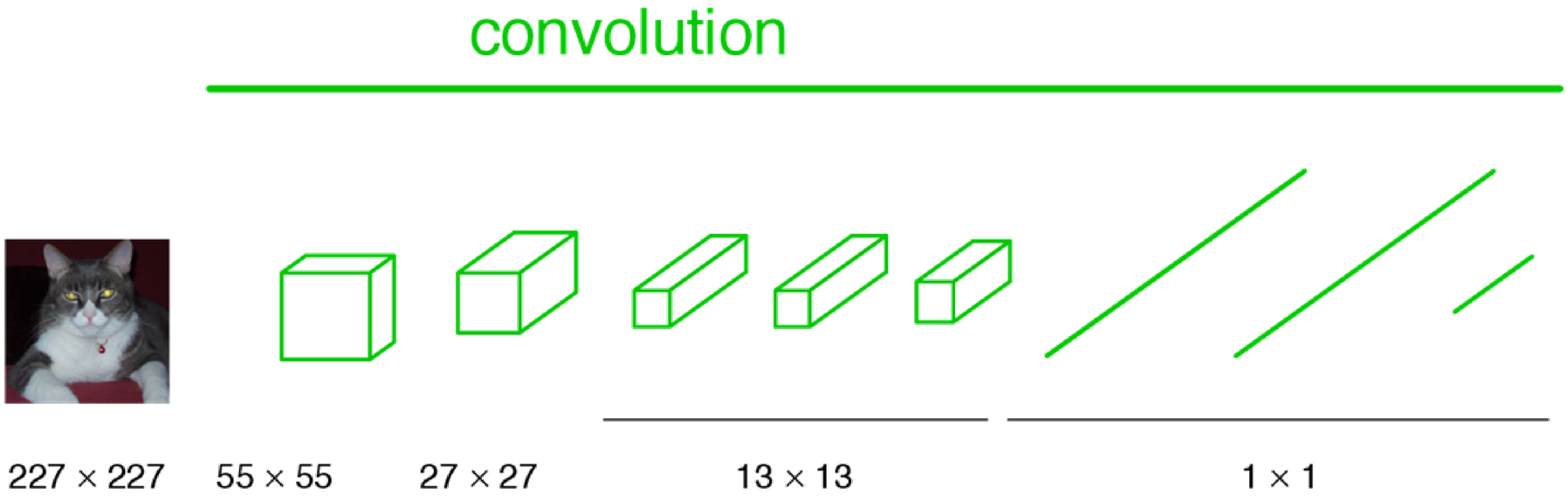


a classification network

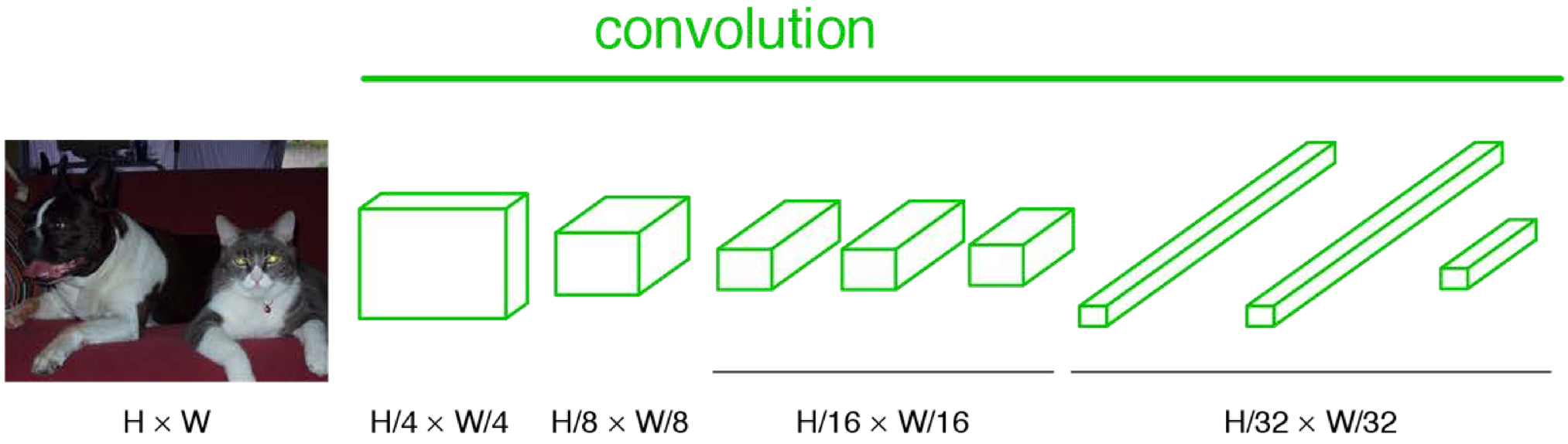


“tabby cat”

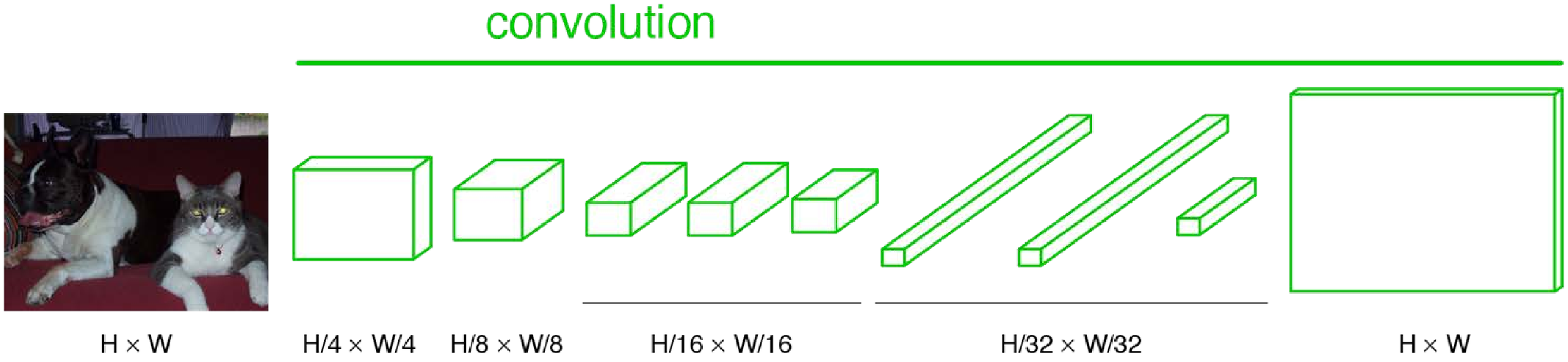
becoming fully convolutional



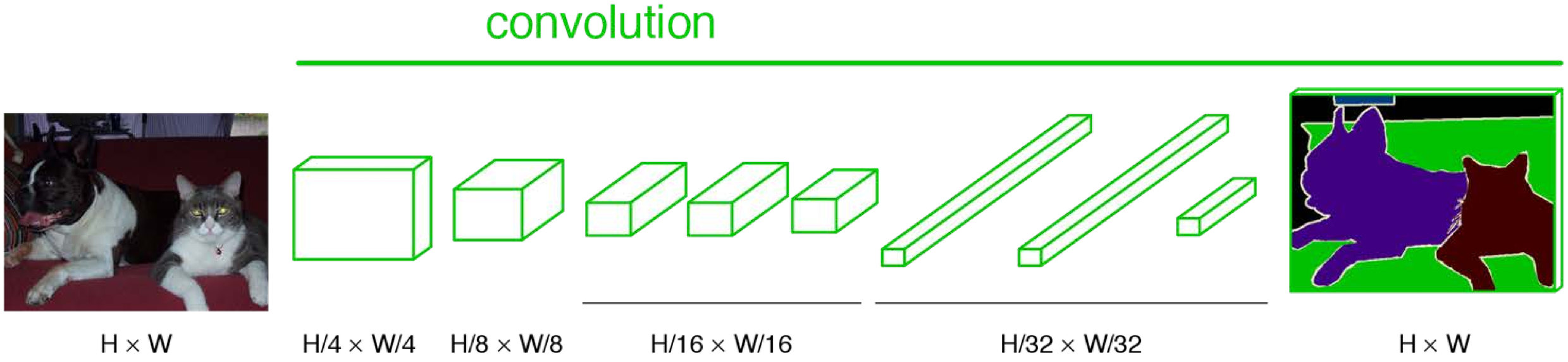
becoming fully convolutional



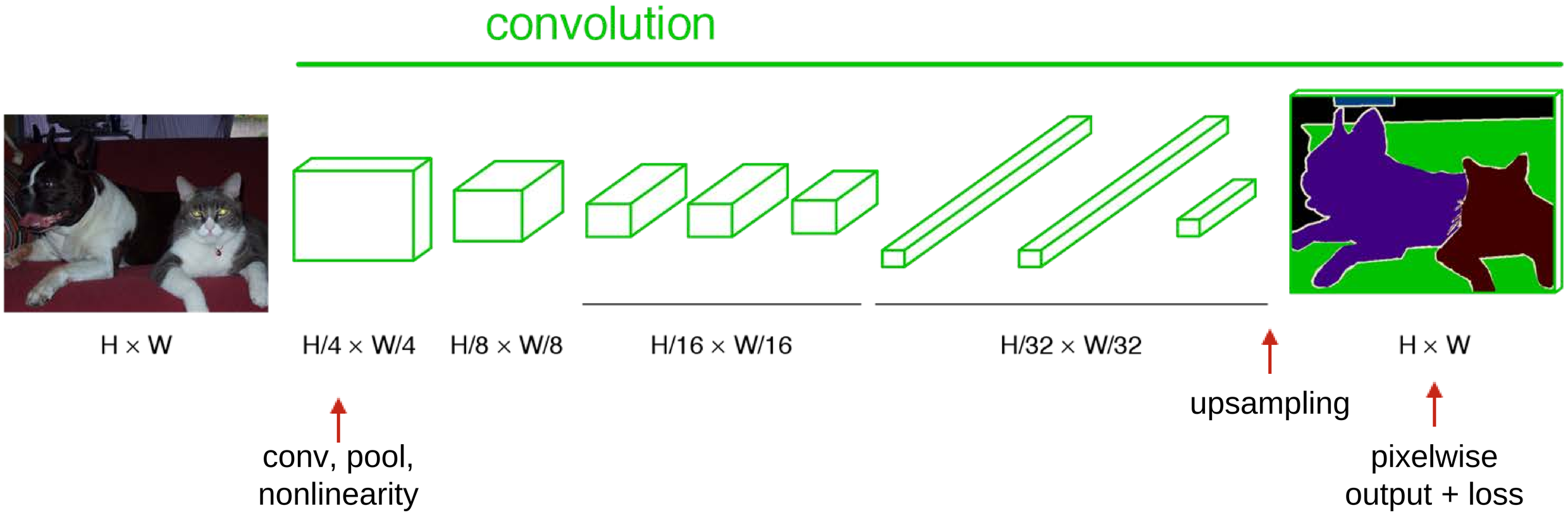
upsampling output



end-to-end, pixels-to-pixels network

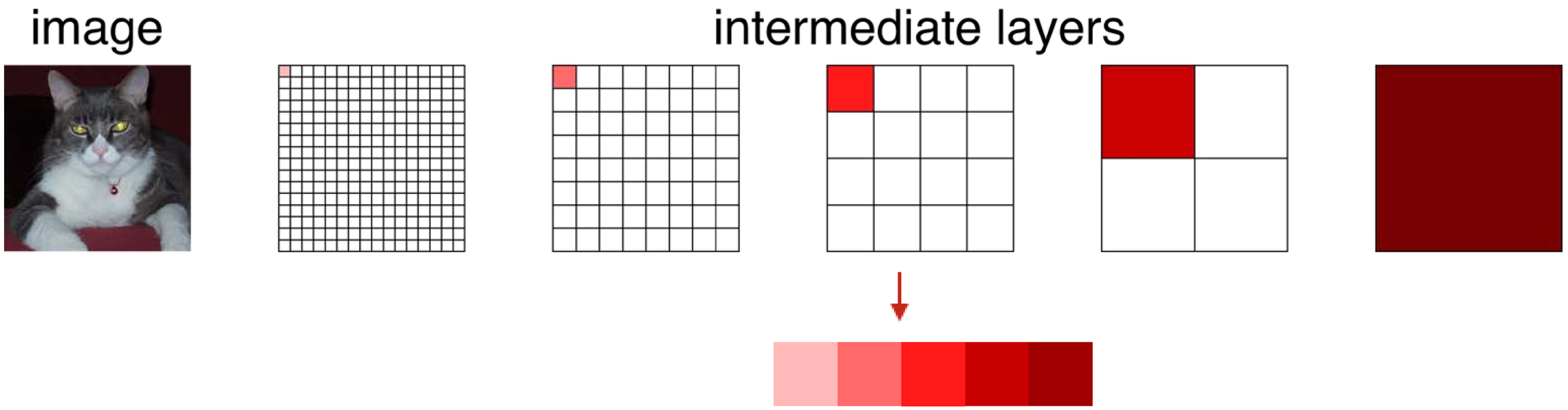


end-to-end, pixels-to-pixels network



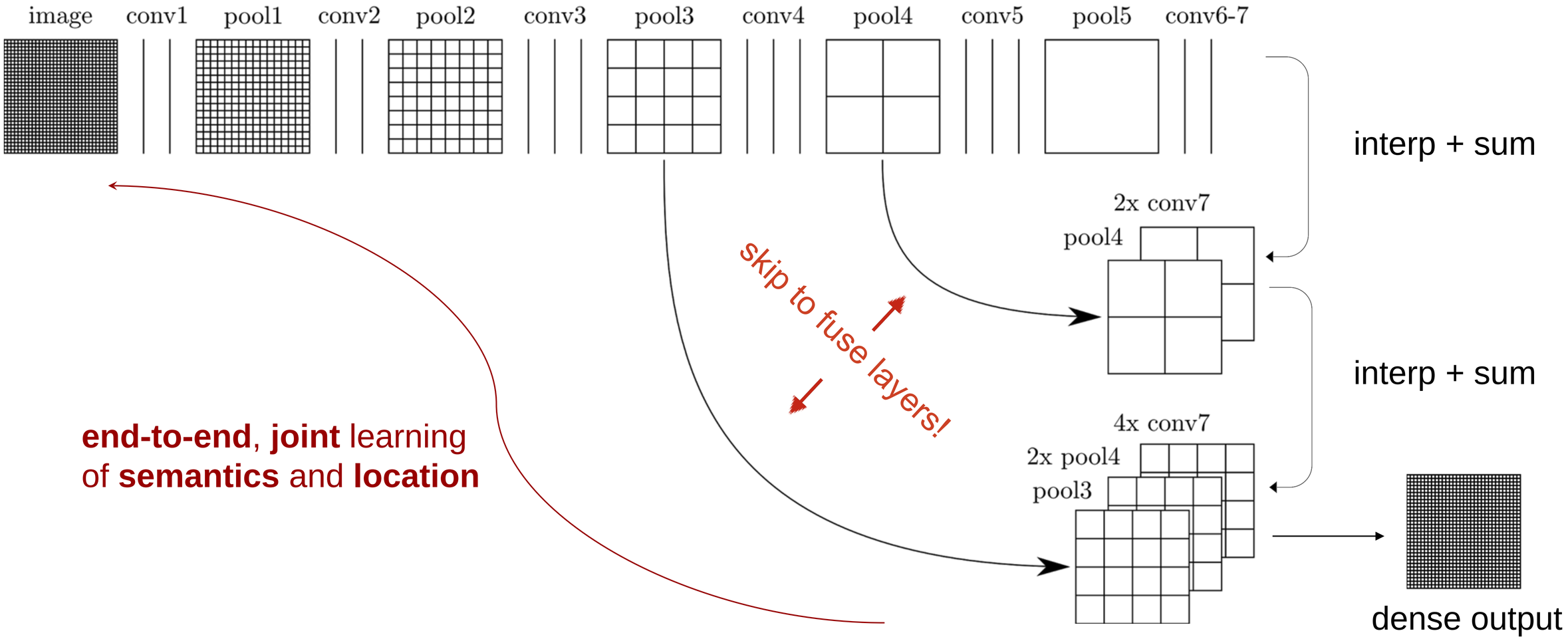
spectrum of deep features

combine *where* (local, shallow) with *what* (global, deep)



(cf. Hariharan et al. CVPR15 “hypercolumn”)

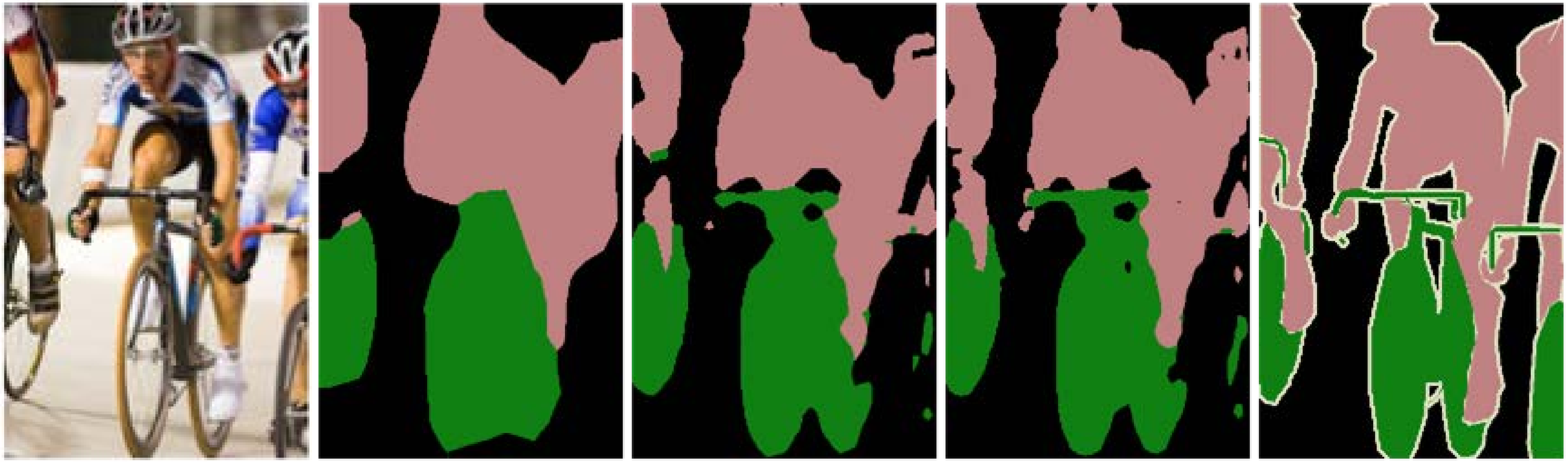
skip layers



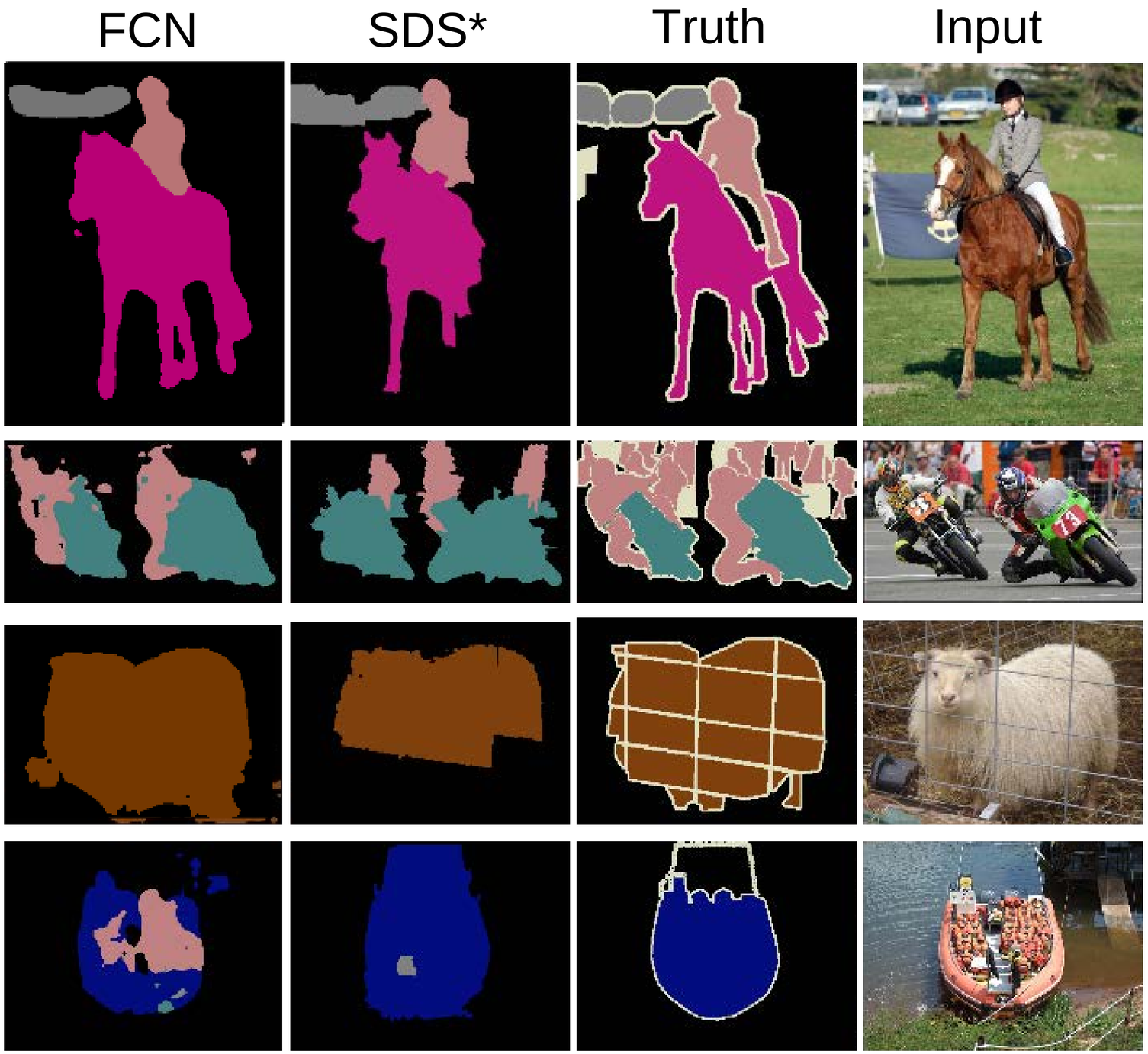
end-to-end, **joint** learning
of **semantics** and **location**

skip layer refinement

input image | stride 32 | stride 16 | stride 8 | ground truth



no skips | 1 skip | 2 skips



Relative to prior state-of-the-art SDS:

- 30% relative improvement for mean IoU
- 286× faster

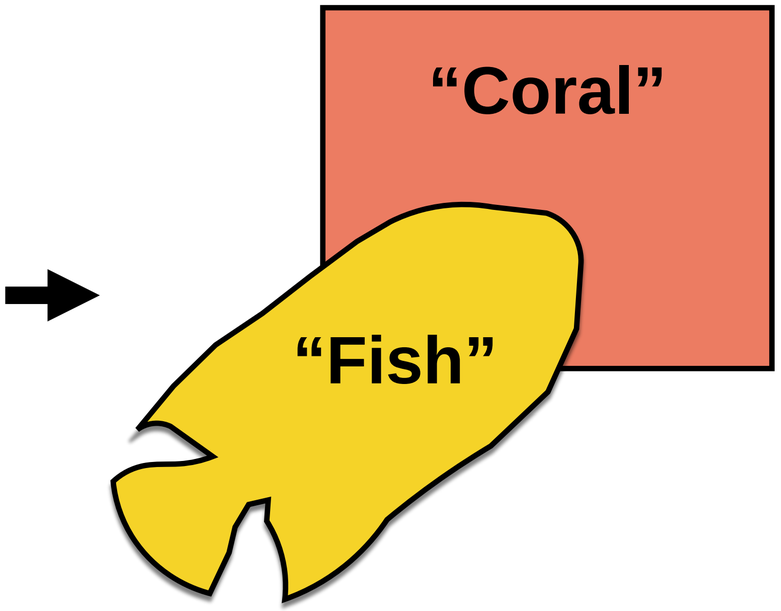
*Simultaneous Detection and Segmentation Hariharan et al. ECCV14

Fancier Architectures: Auto-encoders

Representation Learning

$\mathbf{x}$

Image



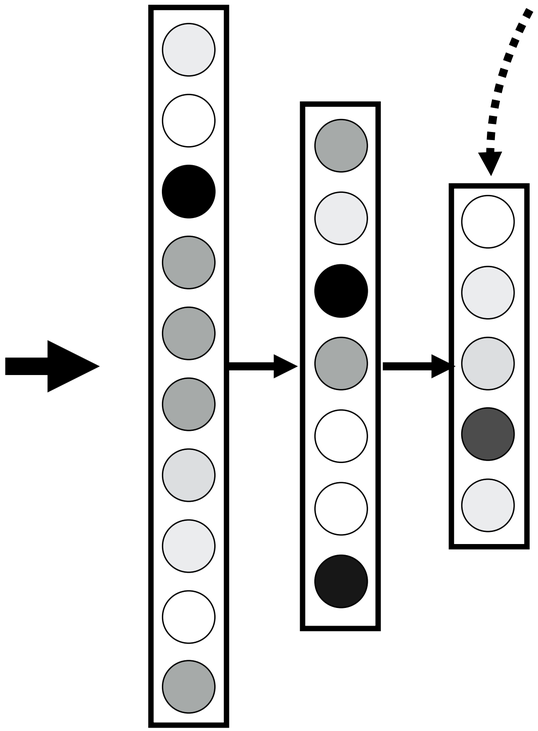
Compact mental representation

Representation Learning

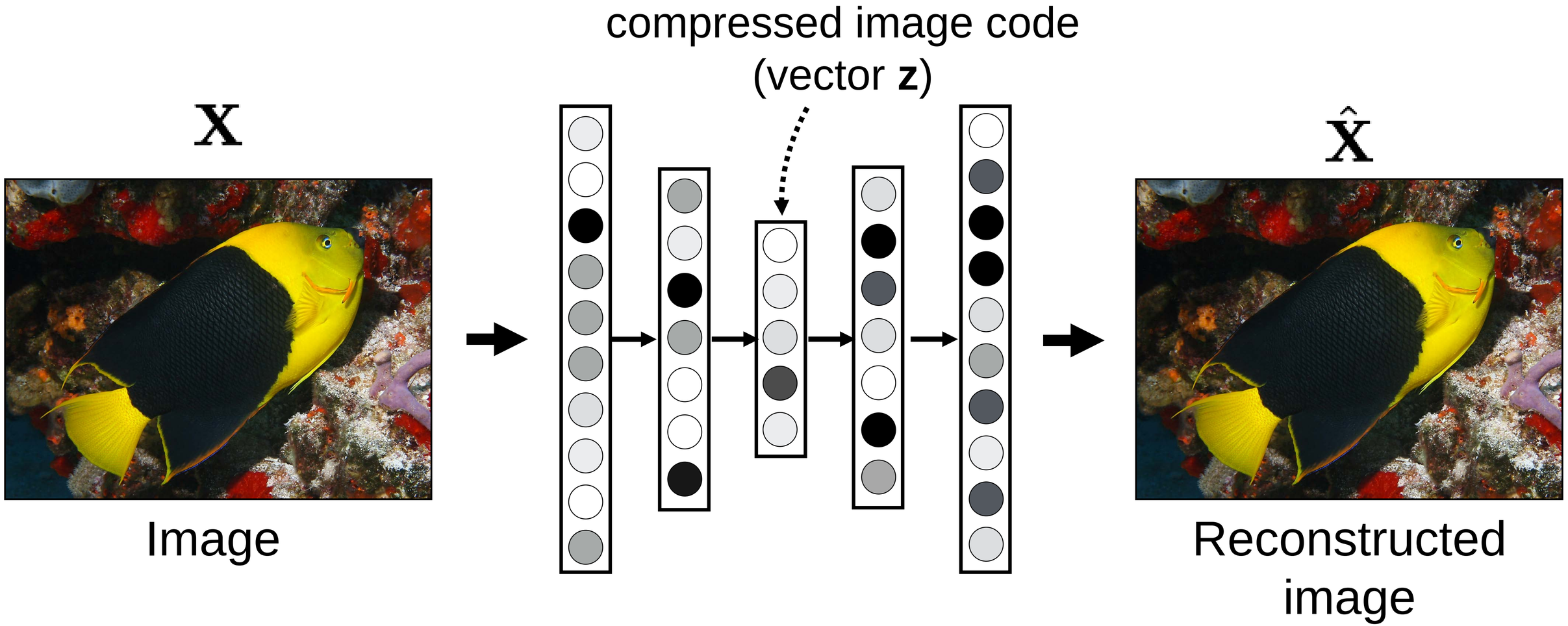
compressed image code
(vector **z**)

x

Image



Representation Learning



“Autoencoder”

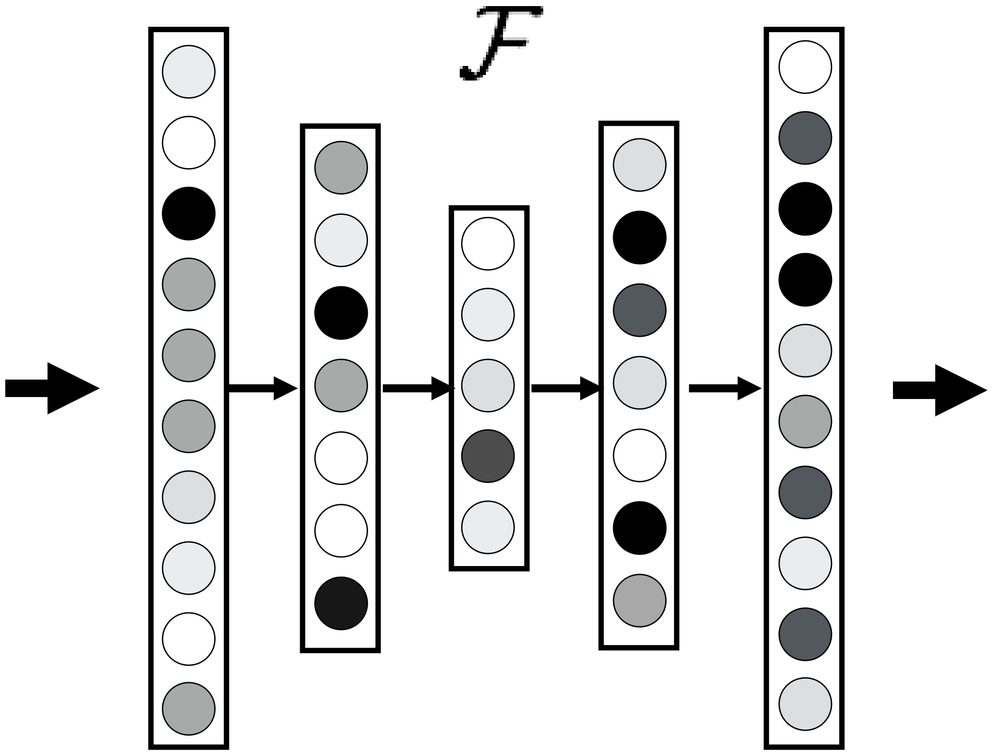
[e.g., Hinton & Salakhutdinov, Science 2006]

Autoencoder

$\mathbf{X}$

Image

$\mathcal{F}$

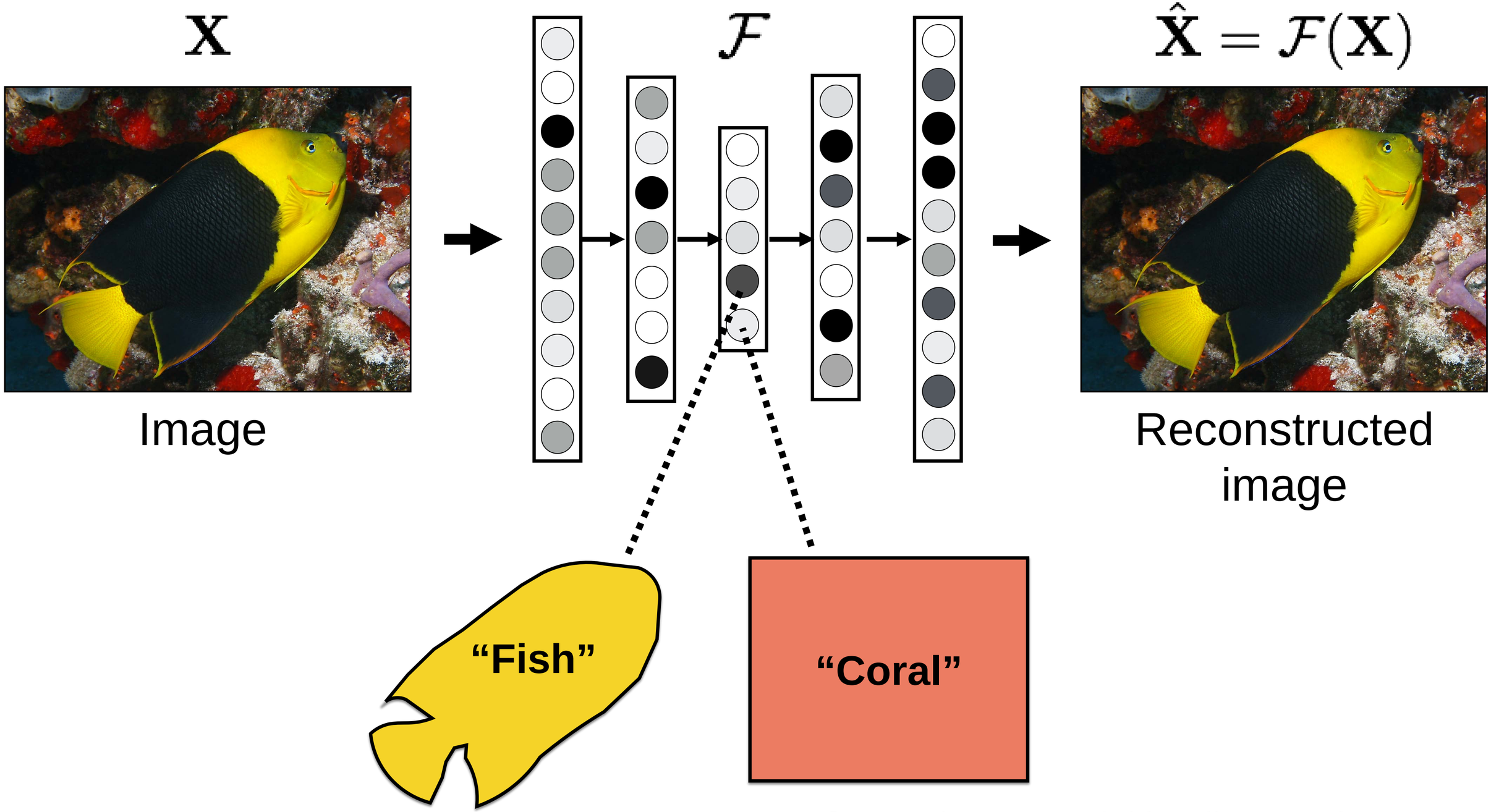


$\hat{\mathbf{X}} = \mathcal{F}(\mathbf{X})$

Reconstructed image

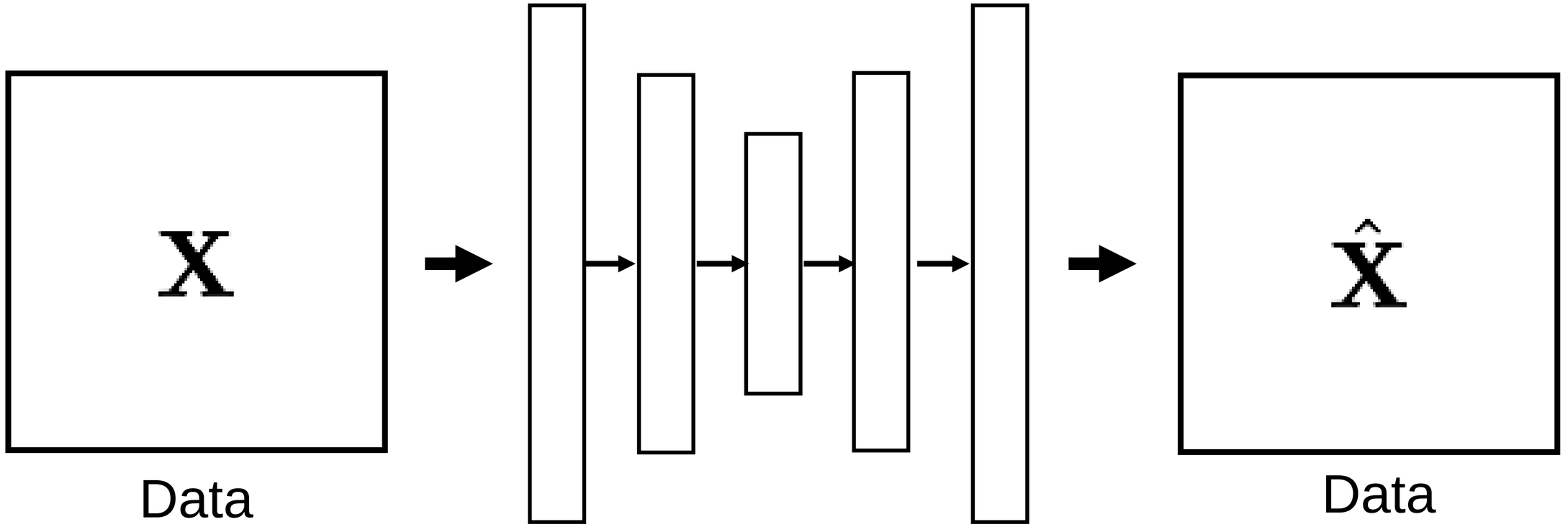
$$\arg\min_{\mathcal{F}} \mathbb{E}_{\mathbf{X}}[||\mathcal{F}(\mathbf{X}) - \mathbf{X}||]$$

[e.g., Hinton & Salakhutdinov, Science 2006]



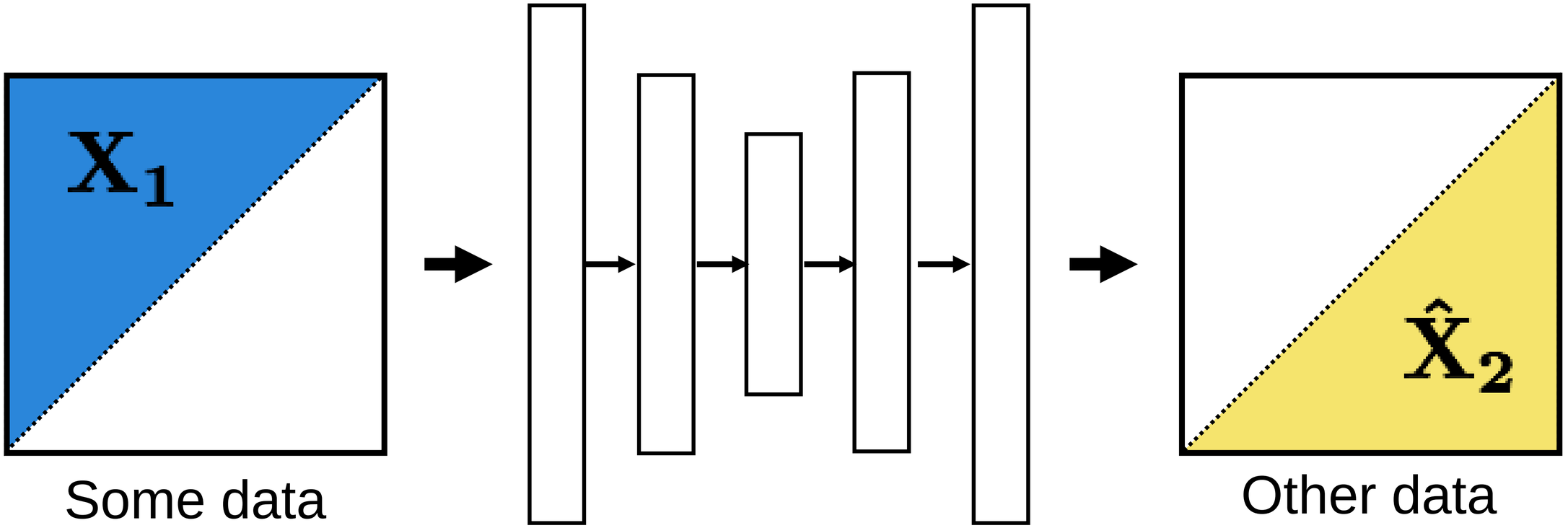
[e.g., Hinton & Salakhutdinov, Science 2006]

Data compression

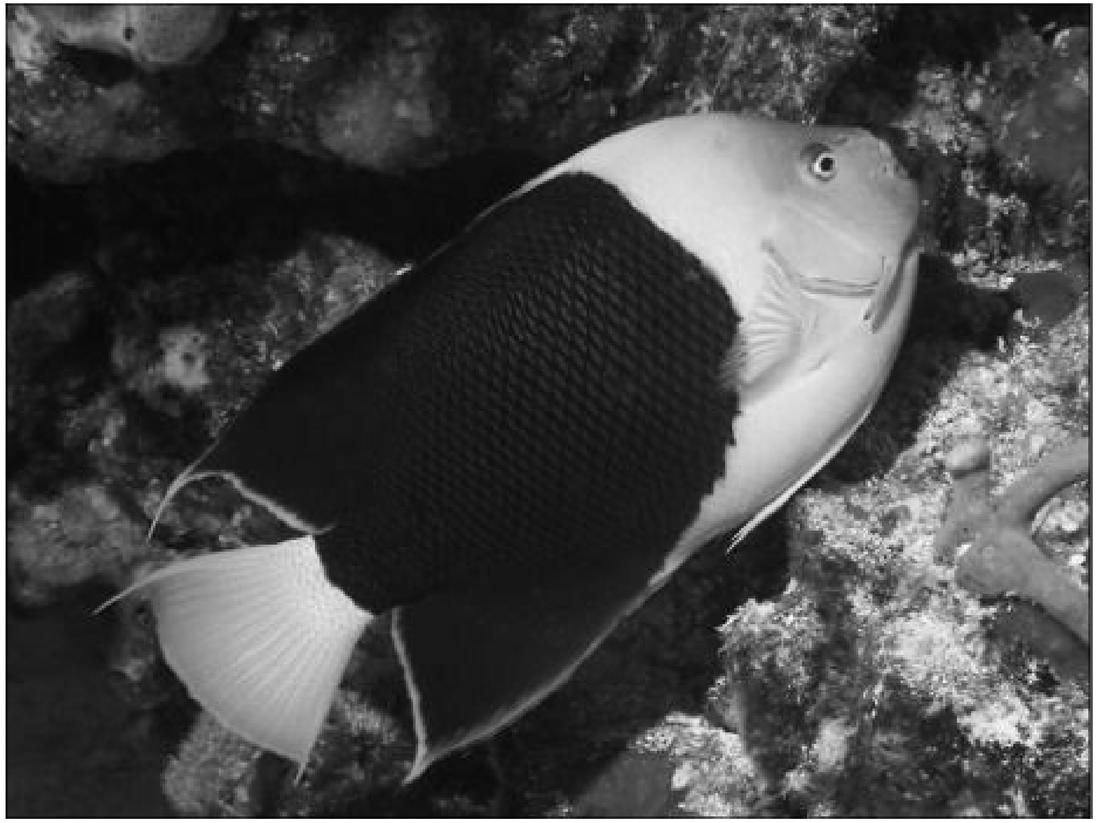


[Hinton & Salakhutdinov, Science 2009]

Data *prediction*



see also [Vincent et al., 2008]



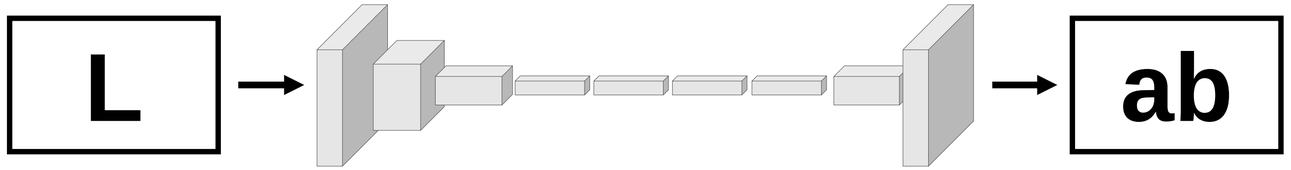
$\mathcal{F}$

Grayscale image: L channel

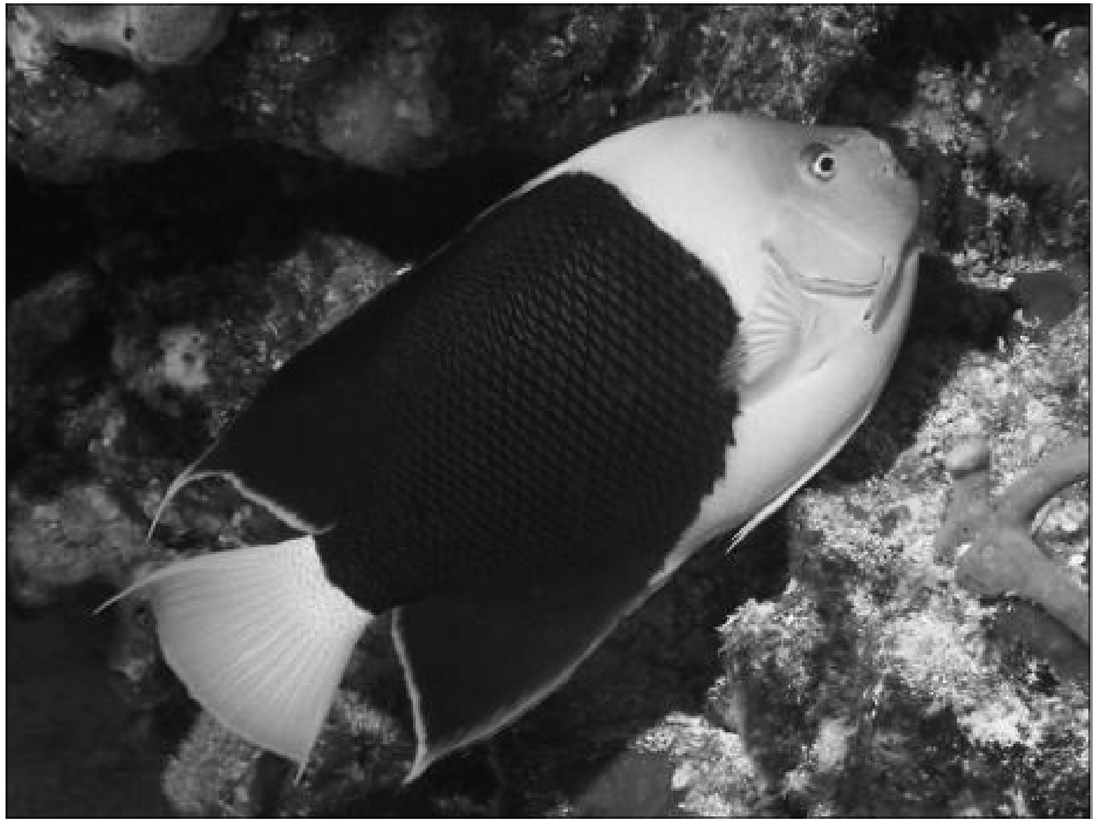
$\mathbf{X} \in \mathbb{R}^{H \times W \times 1}$

Color information: ab channels

$\hat{\mathbf{Y}} \in \mathbb{R}^{H \times W \times 2}$



[Zhang, Isola, Efros, ECCV 2016]



$\mathcal{F}$

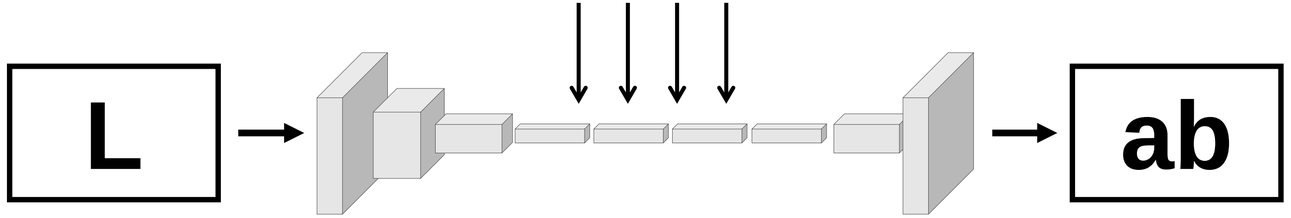
Grayscale image: L cha

$\mathbf{X} \in \mathbb{R}^{H \times W \times 1}$

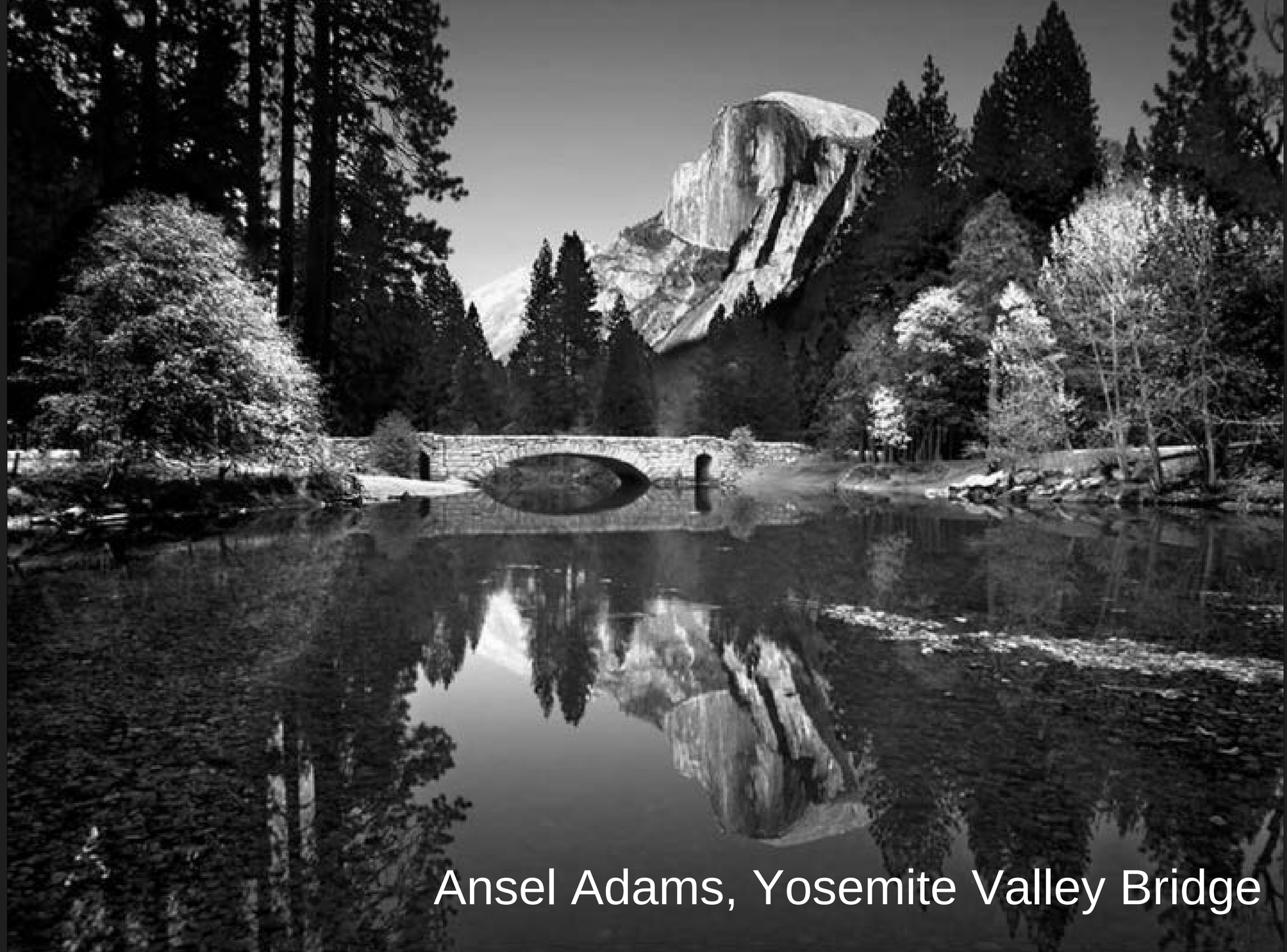
Semantics? Higher-level abstraction?

information: ab channels

$\hat{\mathbf{Y}} \in \mathbb{R}^{H \times W \times 2}$

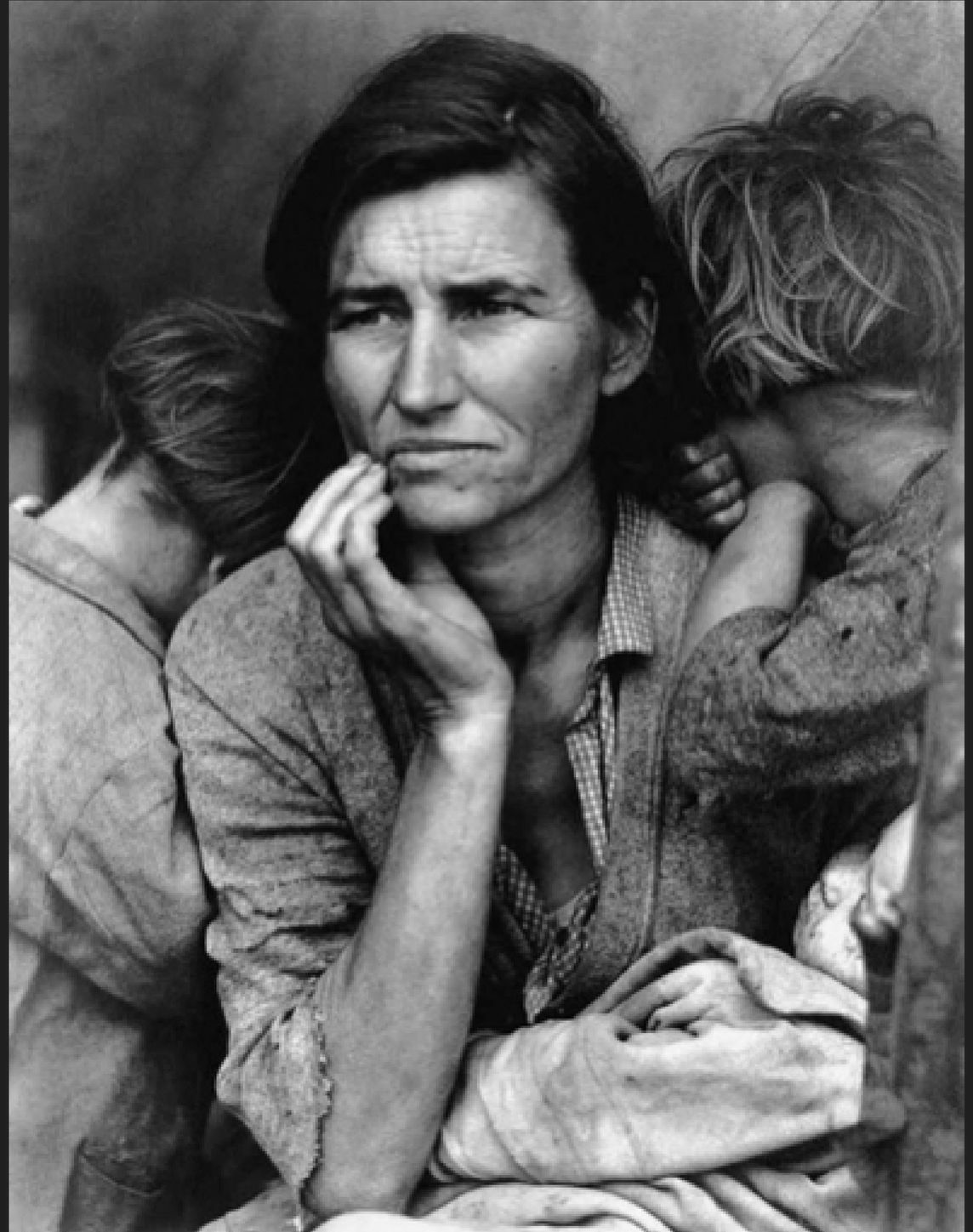


[Zhang, Isola, Efros, ECCV 2016]

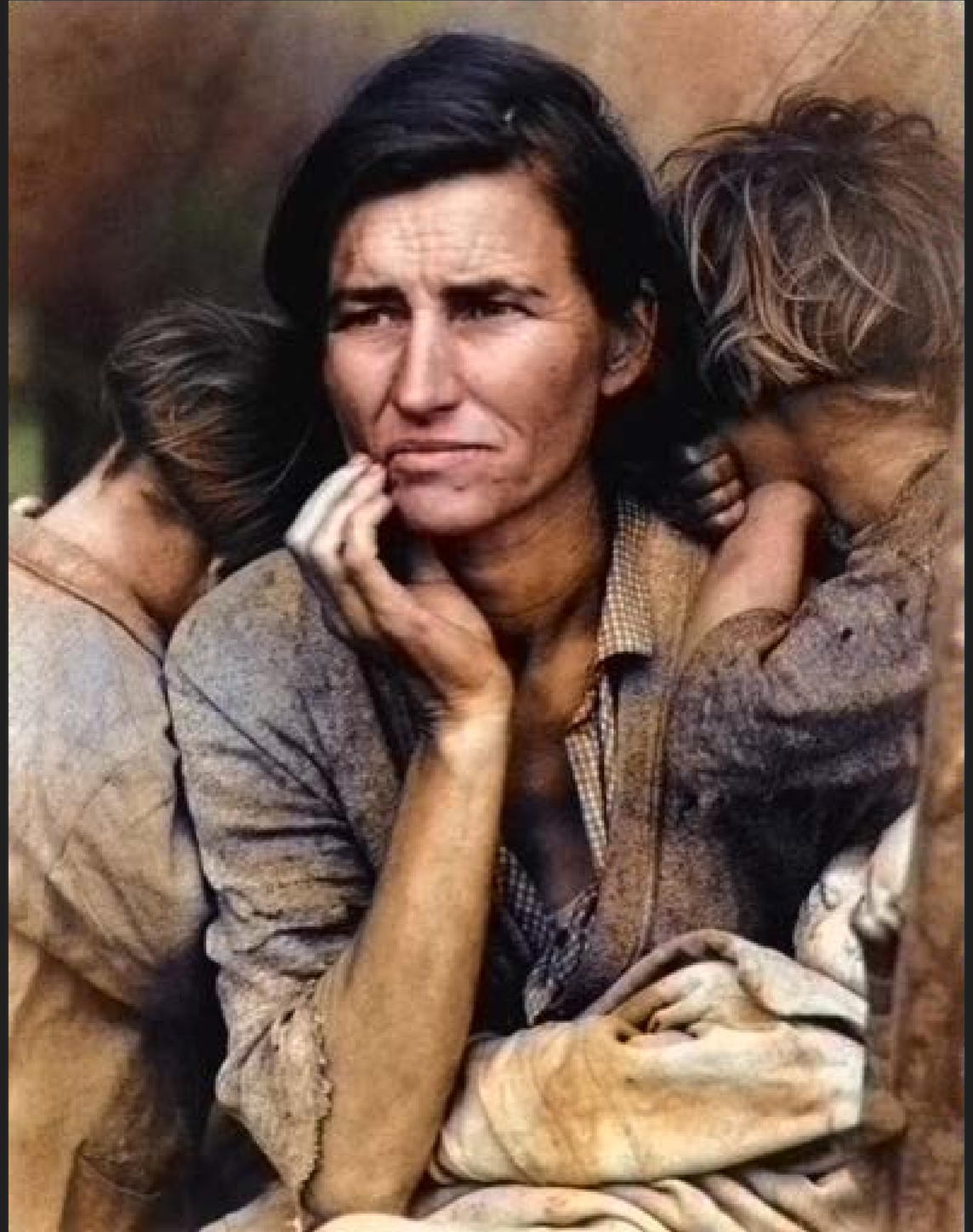


Ansel Adams, Yosemite Valley Bridge


Our result



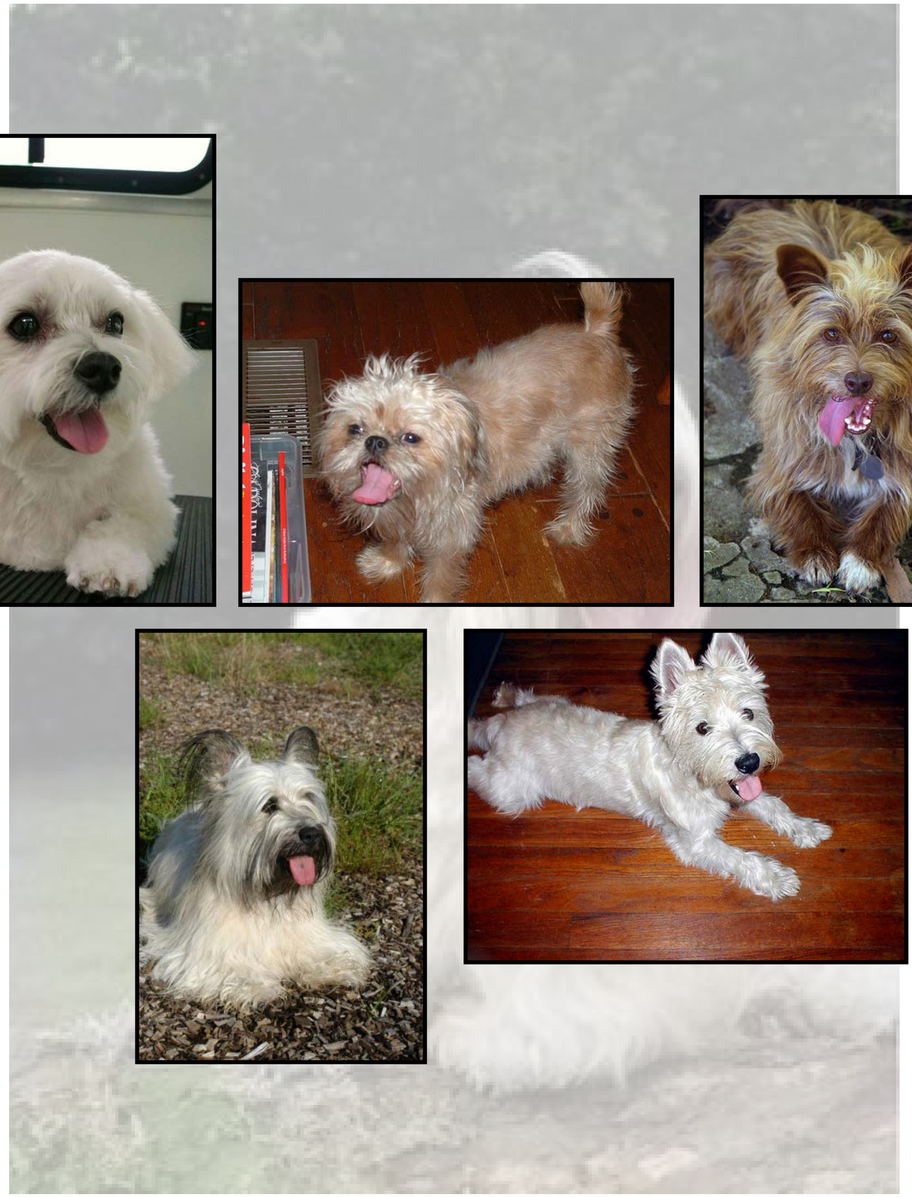
Migrant Mother
Dorothea Lange
1936



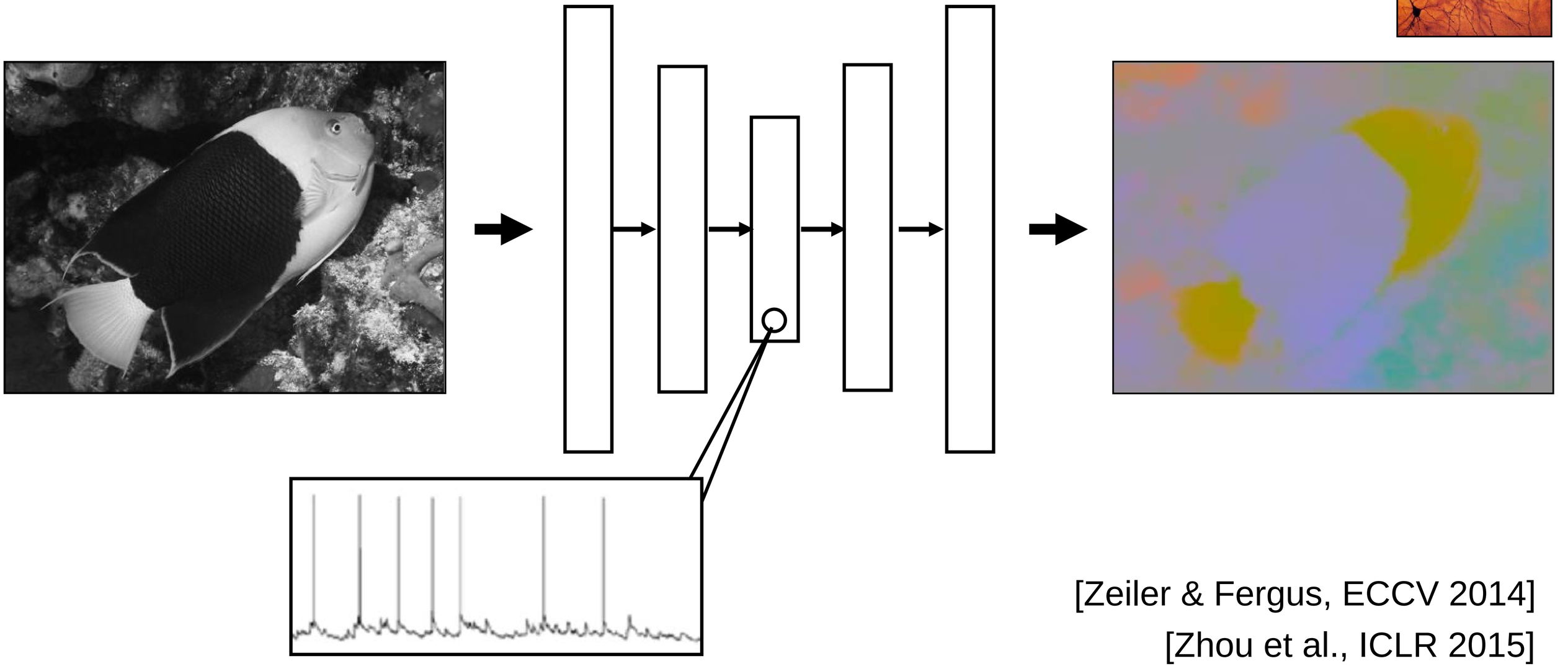
Our result

Instructive failure

Instructive failure



Deep Net “Electrophysiology”



[Zeiler & Fergus, ECCV 2014]

[Zhou et al., ICLR 2015]

Stimuli that drive selected neurons (conv5 layer)

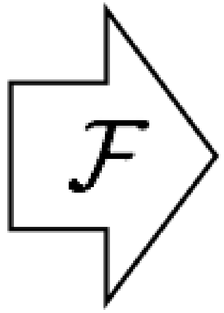
faces

dog faces

flowers

How to evaluate predictions?

Input



Output

$$\arg\min_{\mathcal{F}} \mathbb{E}_{\mathbf{x},\mathbf{y}}\left[L(\mathcal{F}(\mathbf{x}), \mathbf{y})\right]$$

Objective function (loss)

Neural Network

How to evaluate predictions?

Input

$\mathcal{F}$

Output

$$\arg\min_{\mathcal{F}} \mathbb{E}_{\mathbf{x},\mathbf{y}}\left[L(\mathcal{F}(\mathbf{x}), \mathbf{y})\right]$$

"**What** should I do"

"**How** should I do it?"

Designing objective functions

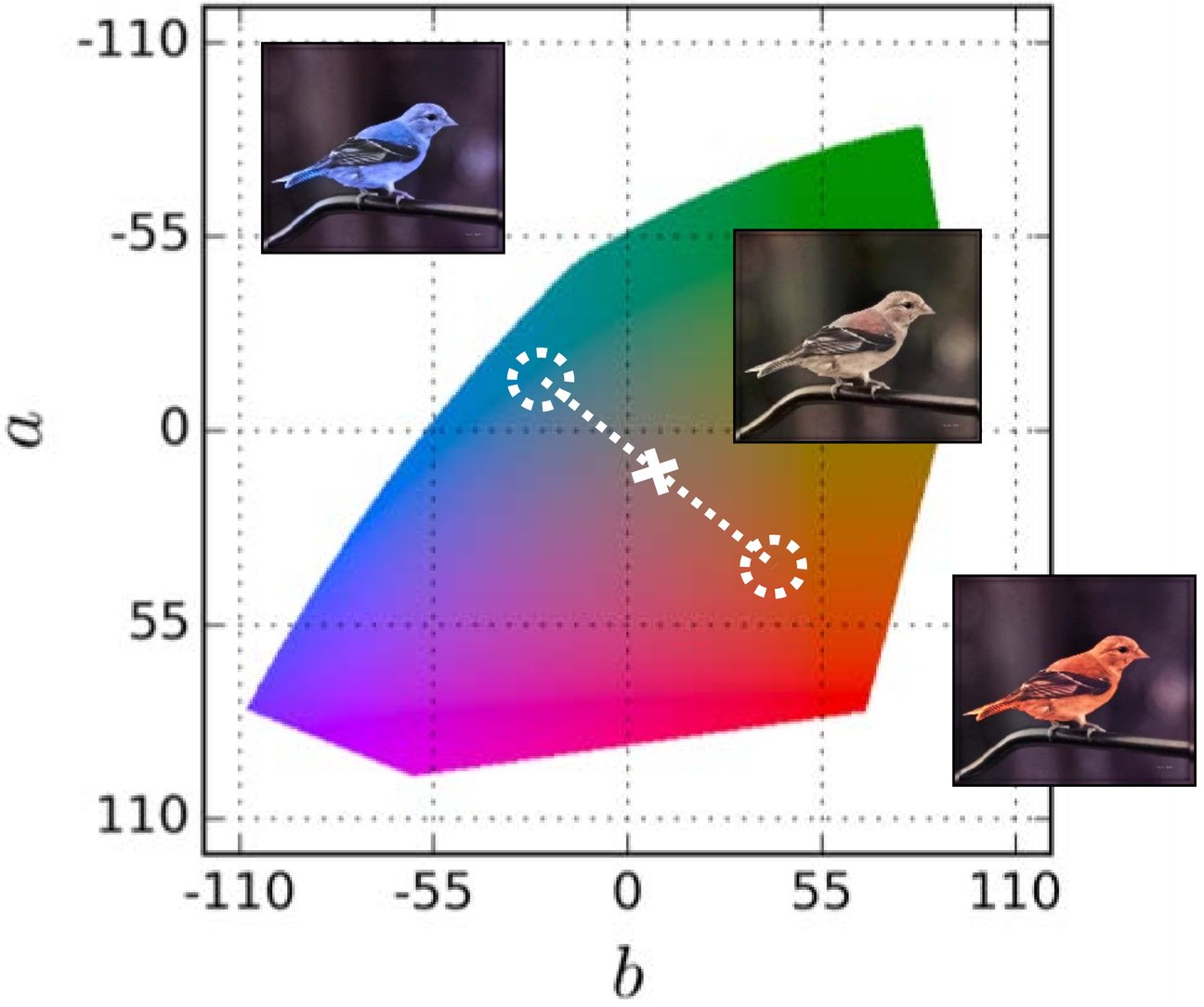
Input

Output

Ground truth

$$\mathbf{L_2}(\widehat{\mathbf{Y}}, \mathbf{Y}) = \frac{1}{2}\sum_{h,w} \|\mathbf{Y}_{h,w} - \widehat{\mathbf{Y}}_{h,w}\|_2^2$$

[Zhang, Isola, Efros, ECCV 2016]



$$\mathrm{L}_2(\widehat{\mathbf{Y}}, \mathbf{Y}) = \frac{1}{2} \sum_{h,w} \|\mathbf{Y}_{h,w} - \widehat{\mathbf{Y}}_{h,w}\|_2^2$$

Designing objective functions

Input

Zhang et al. 2016

Ground truth

Color distribution cross-entropy loss with colorfulness enhancing term.

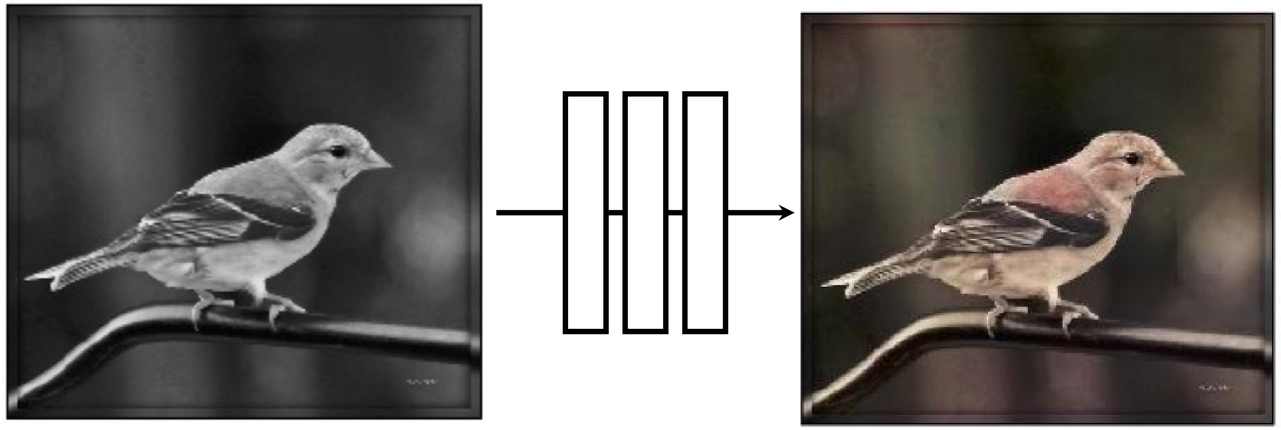
[Zhang, Isola, Efros, ECCV 2016]

Designing objective functions

Be careful what you wish for!

Designing objective functions

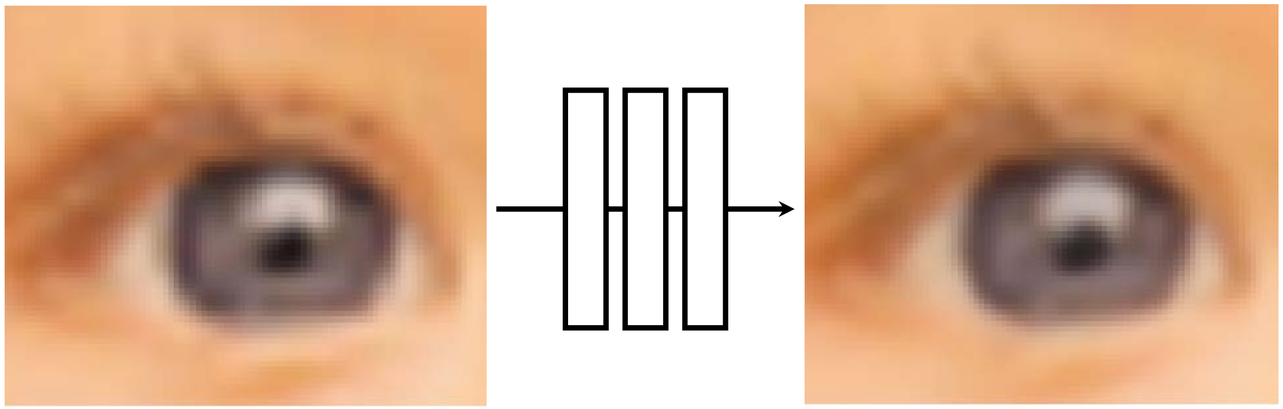
Image colorization



[Zhang, Isola, Efros, ECCV 2016]

L2 regression

Super-resolution

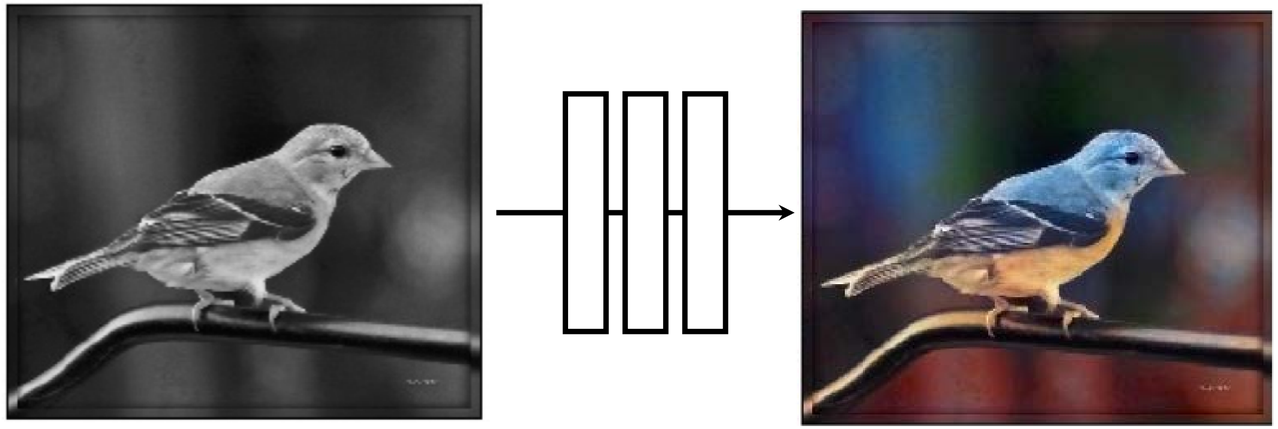


[Johnson, Alahi, Li, ECCV 2016]

L2 regression

Designing objective functions

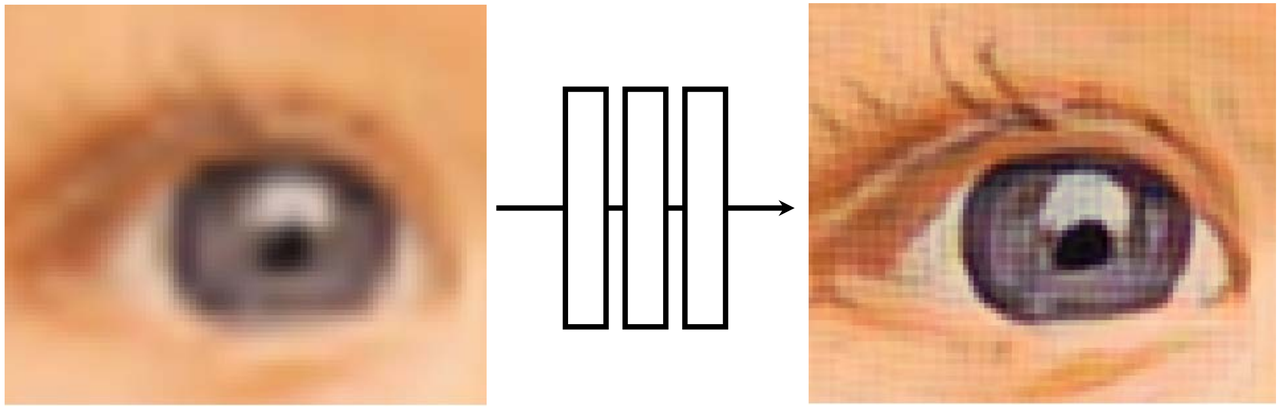
Image colorization



[Zhang, Isola, Efros, ECCV 2016]

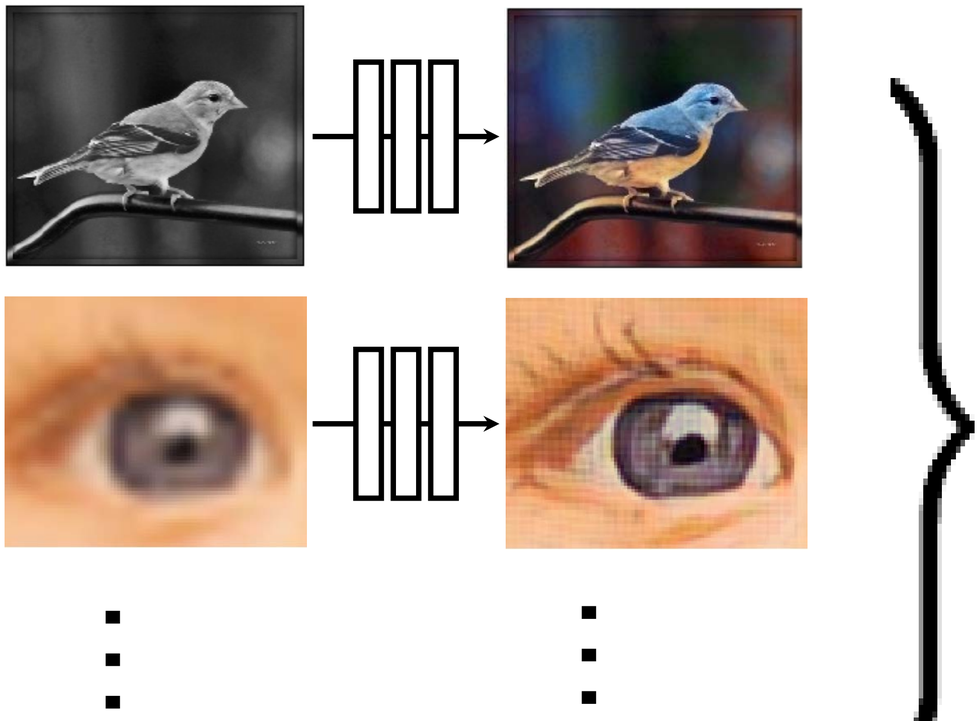
Cross entropy objective,
with colorfulness term

Super-resolution



[Johnson, Alahi, Li, ECCV 2016]

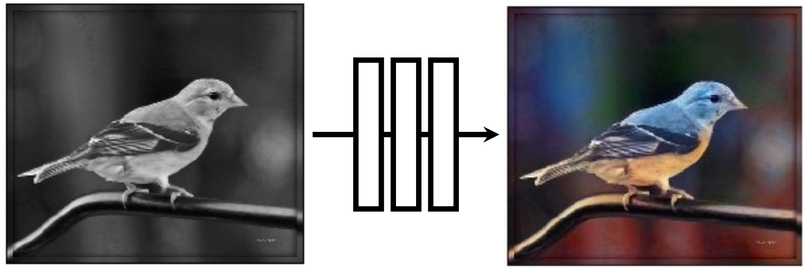
Deep feature covariance
matching objective

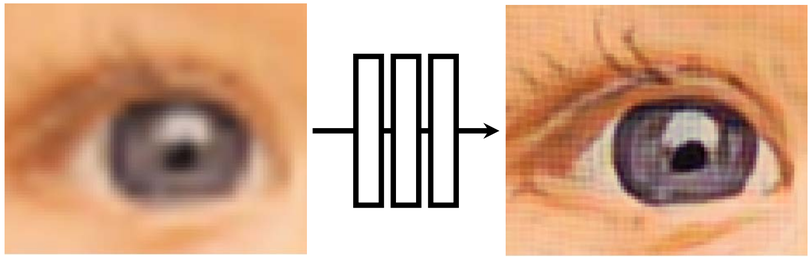


Universal loss?

“Generative Adversarial Network” (GANs)

Generated images





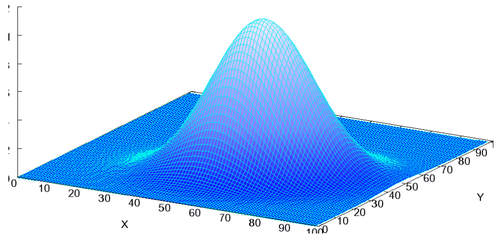
Generated vs Real (classifier)

Real photos

…

[Goodfellow, Pouget-Abadie, Mirza, Xu, Warde-Farley, Ozair, Courville, Bengio 2014]

GANs



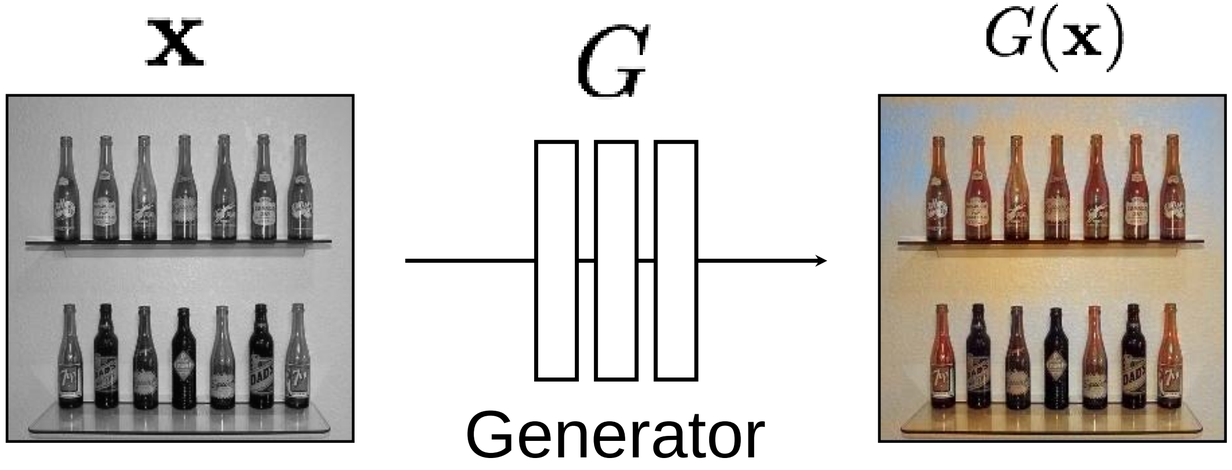
Z

[Goodfellow et al., 2014]

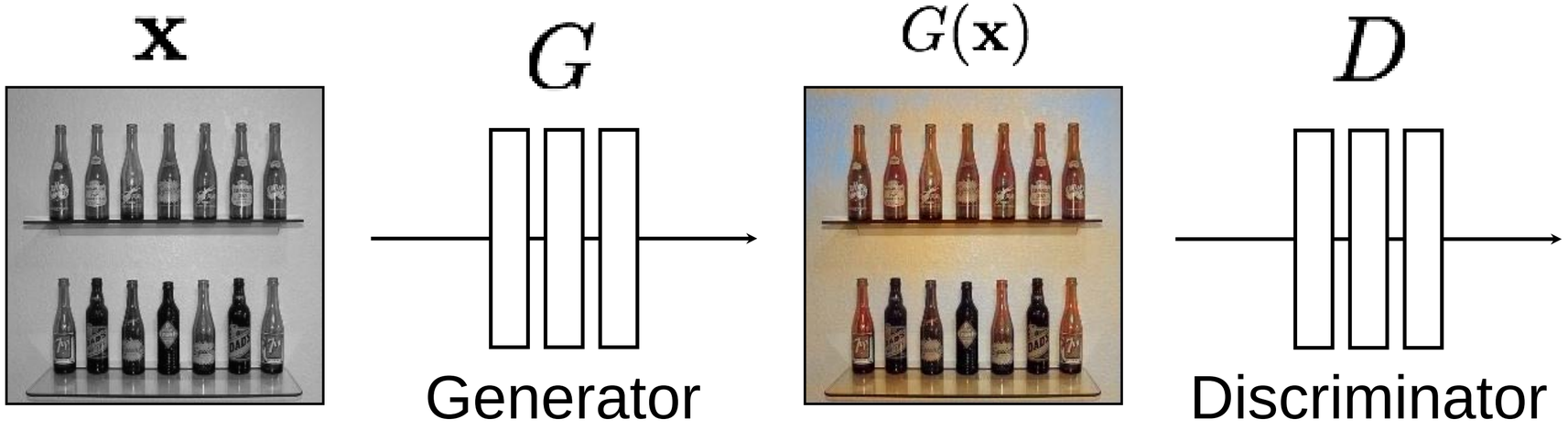
Conditional GANs

[Goodfellow et al., 2014]

[Isola et al., 2016]



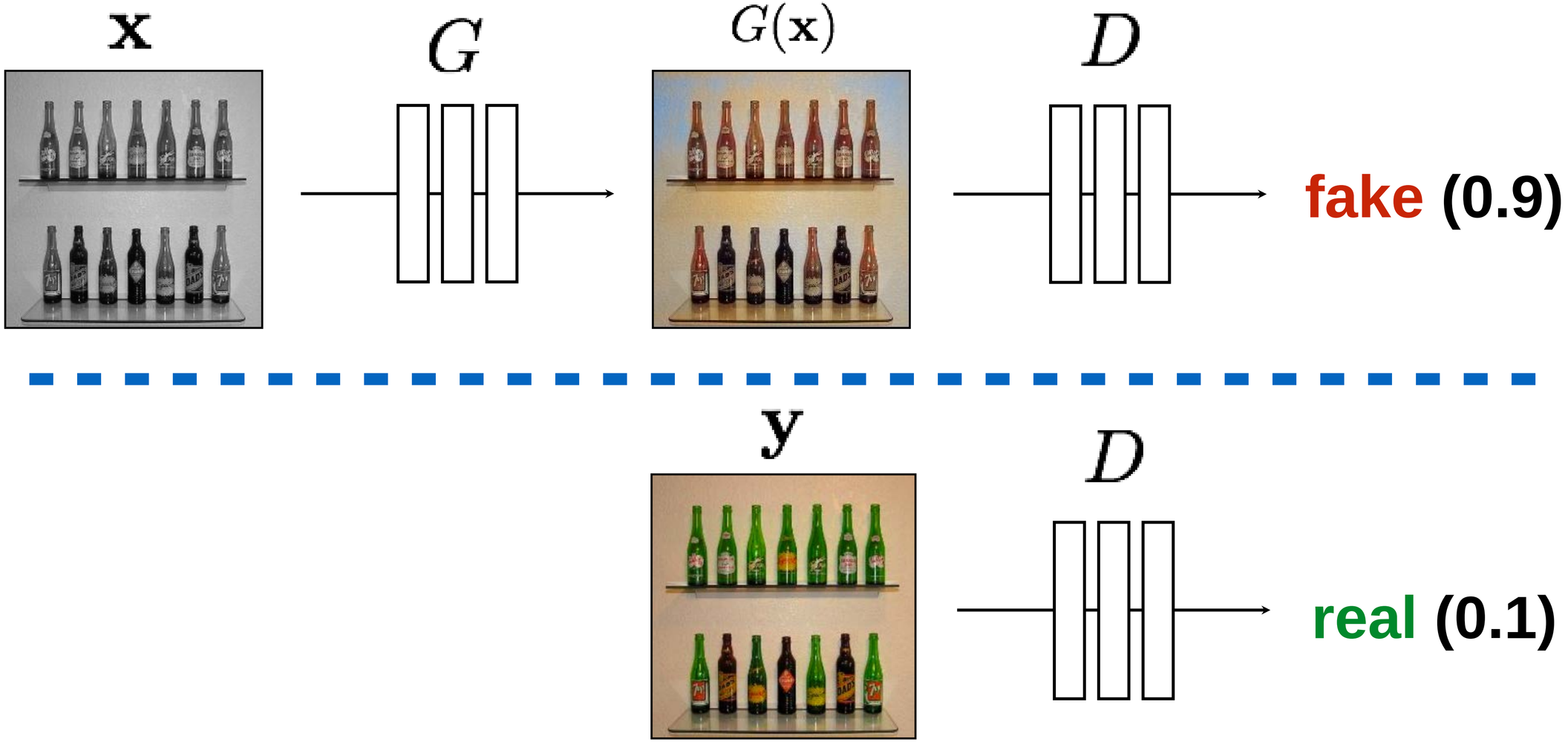
[Goodfellow et al., 2014]



G tries to synthesize fake images that fool D

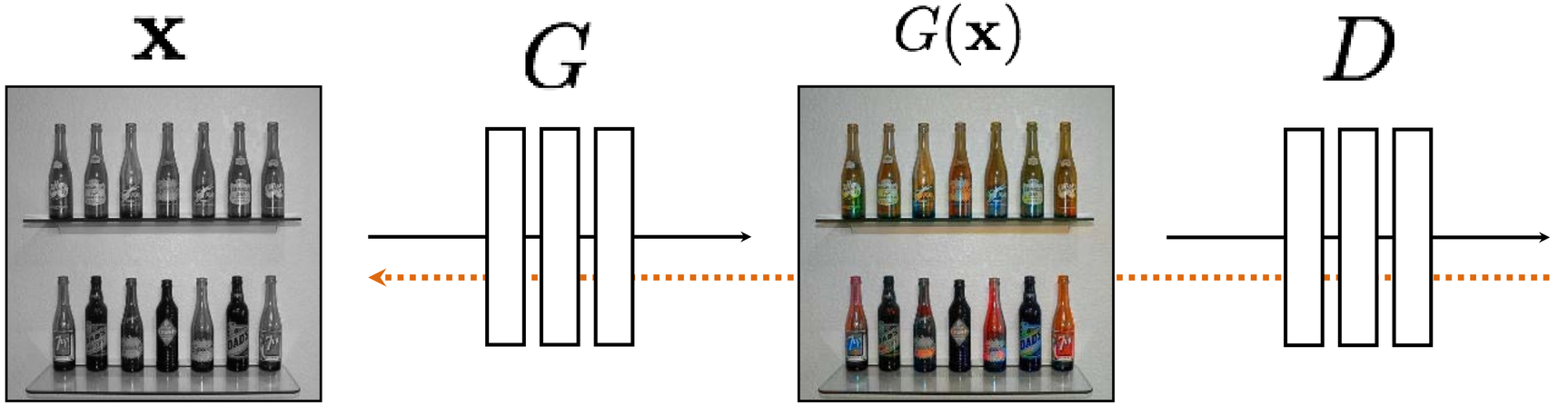
D tries to identify the fakes

[Goodfellow et al., 2014]



$$\arg\max_{D} \; \mathbb{E}_{\mathbf{x},\mathbf{y}}[\; \log D(G(\mathbf{x})) \; + \; \log(1 - D(\mathbf{y}))\;]$$

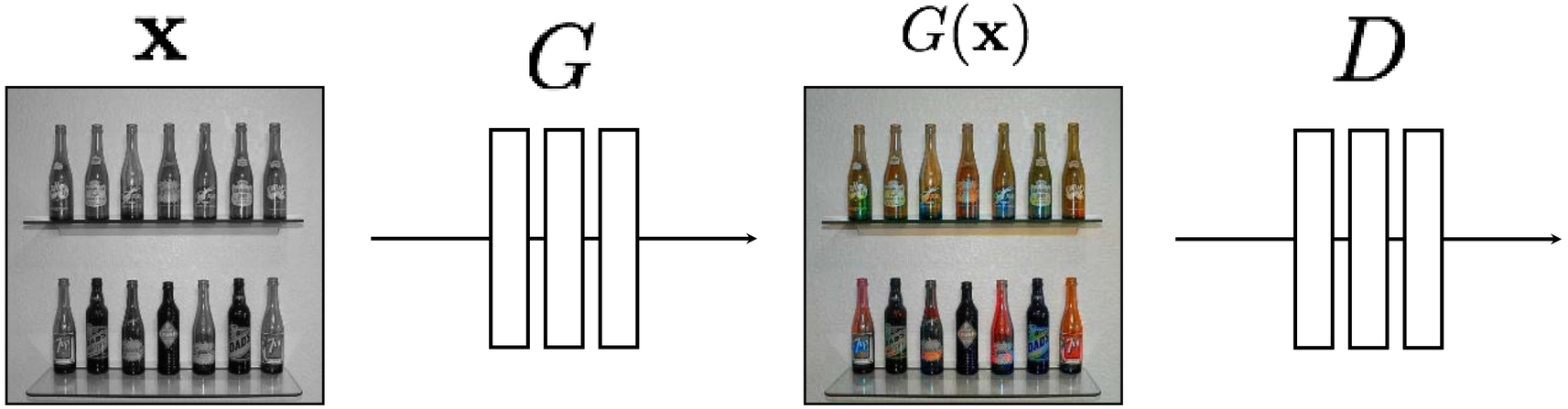
[Goodfellow et al., 2014]



G tries to synthesize fake images that *fool*
D:

$$\arg\min_{G} \mathbb{E}_{\mathbf{x},\mathbf{y}}[\ \log D(G(\mathbf{x})) \quad + \quad \log(1 - D(\mathbf{y}))\ ]$$

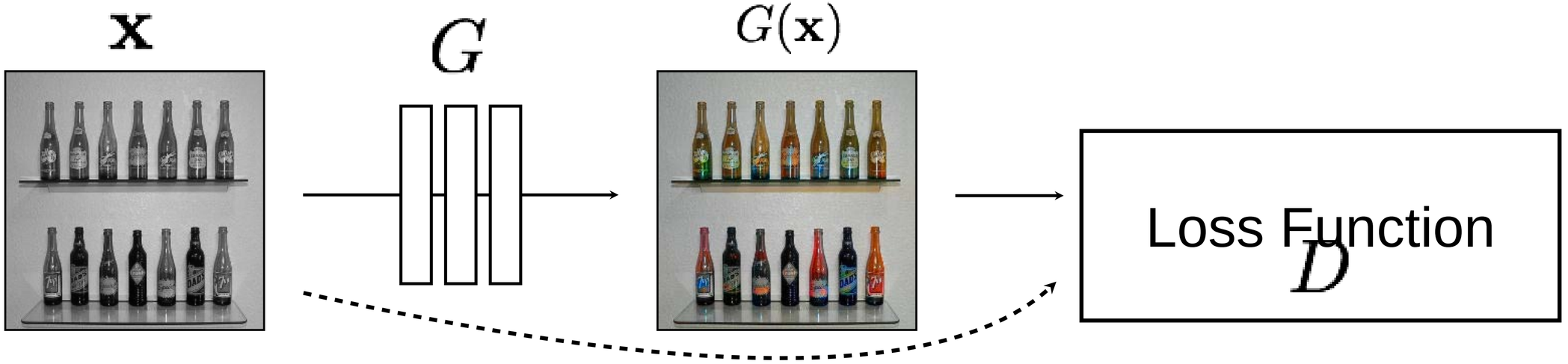
[Goodfellow et al., 2014]



G tries to synthesize fake images that *fool* the *best*
D:

$$\arg \min_{G} \max_{D} \mathbb{E}_{\mathbf{x},\mathbf{y}}[\ \log D(G(\mathbf{x})) \quad + \quad \log(1 - D(\mathbf{y}))\ ]$$

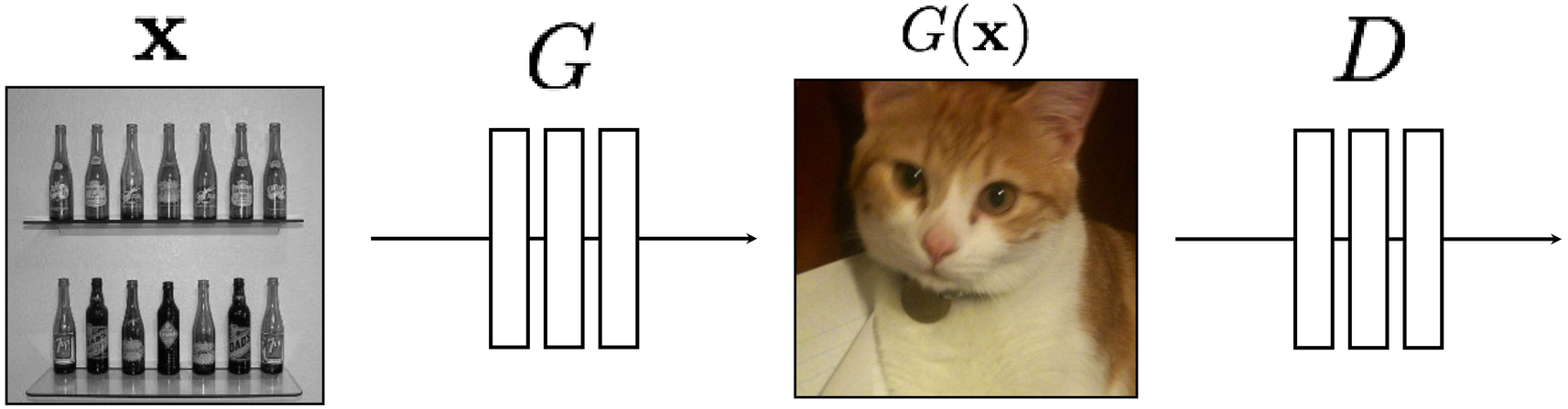
[Goodfellow et al., 2014]



G’s perspective: **D** is a loss function.

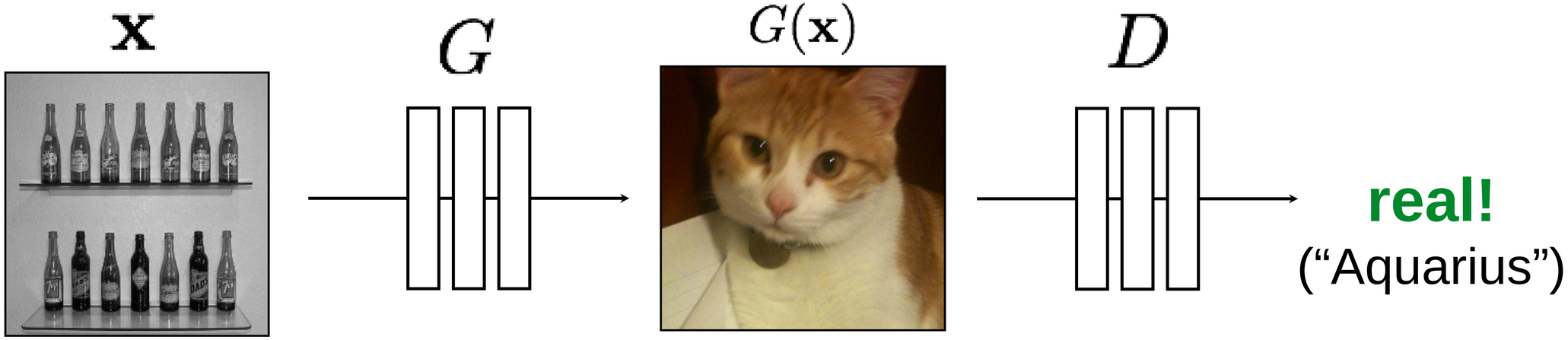
Rather than being hand-designed, it is *learned*.

[Goodfellow et al., 2014]
[Isola et al., 2016]



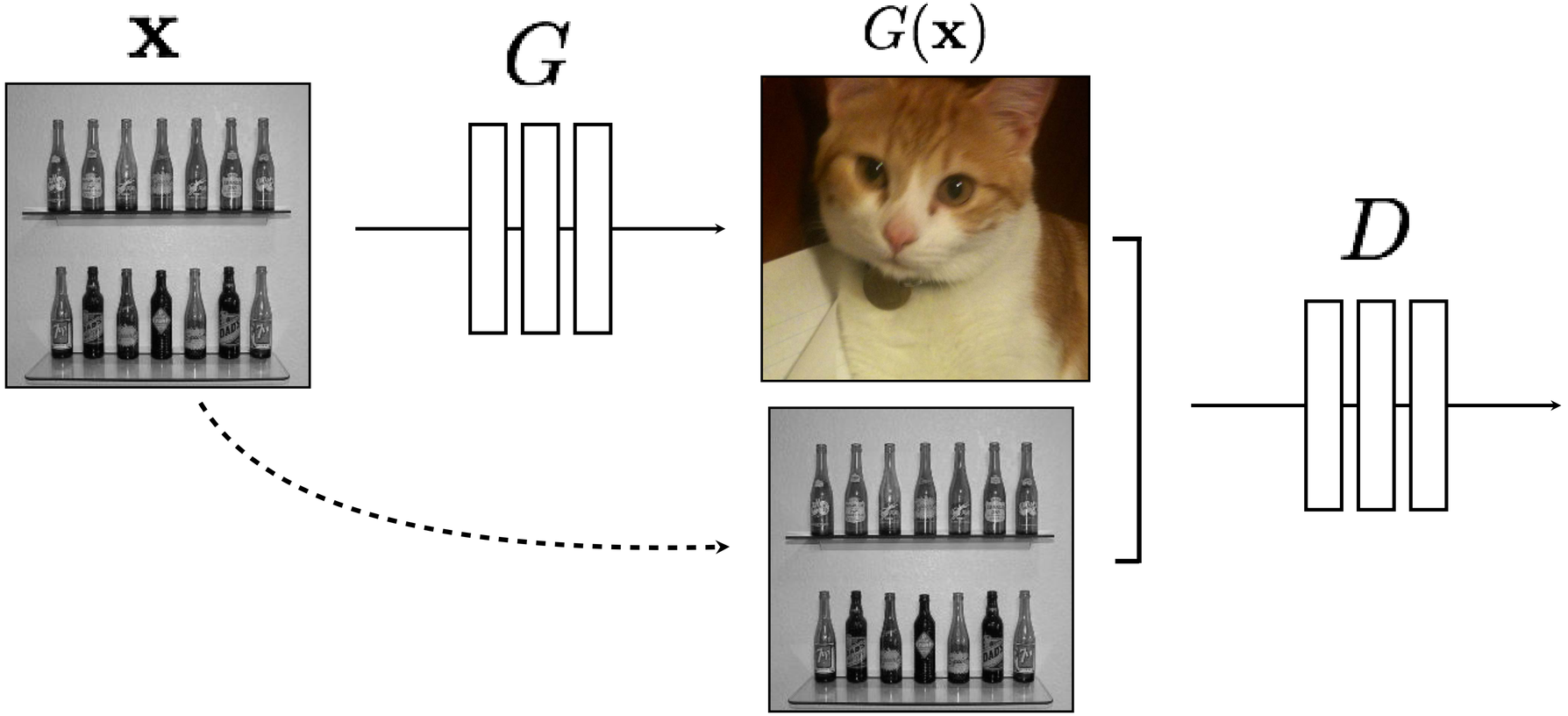
$$\arg\min_G \max_D \; \mathbb{E}_{\mathbf{x},\mathbf{y}}[\; \log D(G(\mathbf{x})) \quad + \quad \log(1 - D(\mathbf{y})) \;\;]$$

[Goodfellow et al., 2014]



$$\arg\min_G \max_D \; \mathbb{E}_{\mathbf{x},\mathbf{y}}[\; \log D(G(\mathbf{x})) \quad + \quad \log(1 - D(\mathbf{y})) \;\;]$$

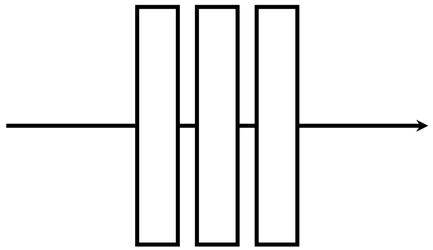
[Goodfellow et al., 2014]

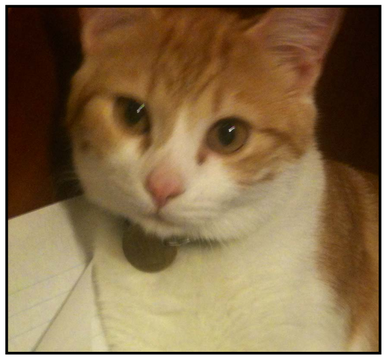


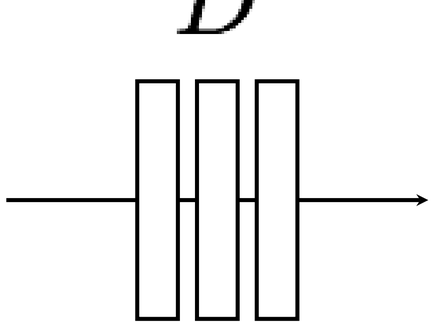
$$\arg\min_G \max_D \; \mathbb{E}_{\mathbf{x},\mathbf{y}}[\; \log D(G(\mathbf{x})) \quad + \quad \log(1 - D(\mathbf{y}))\;]$$

[Goodfellow et al., 2014]

[Isola et al., 2016]







real or fake *pair* ?

$$\arg\min_G \max_D \; \mathbb{E}_{\mathbf{x},\mathbf{y}}[\;\; \log D(\mathbf{x}, G(\mathbf{x})) + \log(1 - D(\mathbf{x}, \mathbf{y}))\;\;]$$

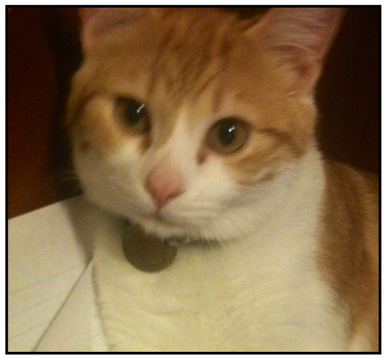
[Goodfellow et al., 2014]

[Isola et al., 2016]

$\mathbf{x}$

G

$G(\mathbf{x})$



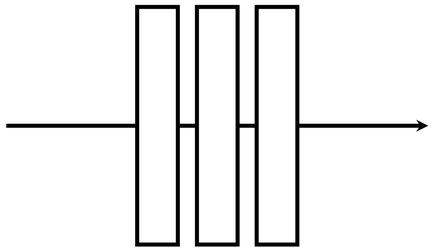
D

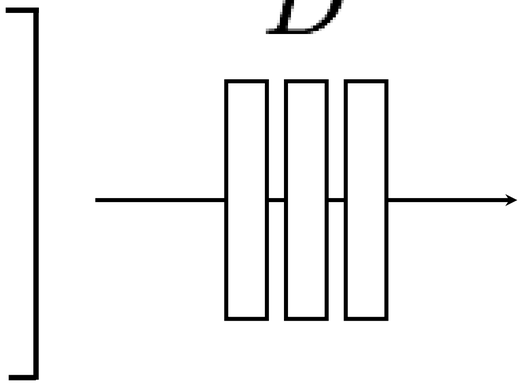
fake *pair*

$$\arg\min_G \max_D \; \mathbb{E}_{\mathbf{x},\mathbf{y}}[\;\; \log D(\mathbf{x}, G(\mathbf{x})) + \log(1 - D(\mathbf{x}, \mathbf{y}))\;\;]$$

[Goodfellow et al., 2014]

[Isola et al., 2016]



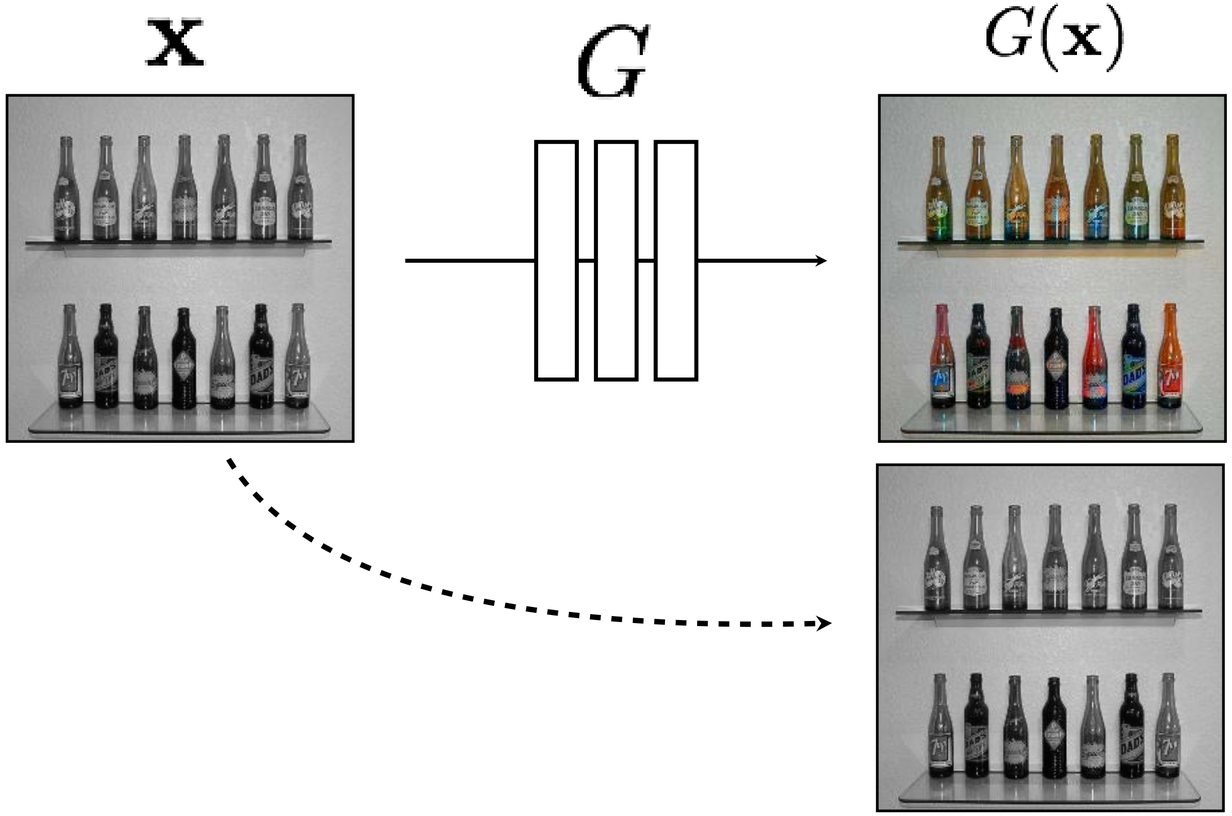


real *pair*

$$\arg\min_G \max_D \; \mathbb{E}_{\mathbf{x},\mathbf{y}}[\;\; \log D(\mathbf{x}, G(\mathbf{x})) + \log(1 - D(\mathbf{x}, \mathbf{y}))\;\;]$$

[Goodfellow et al., 2014]

[Isola et al., 2016]

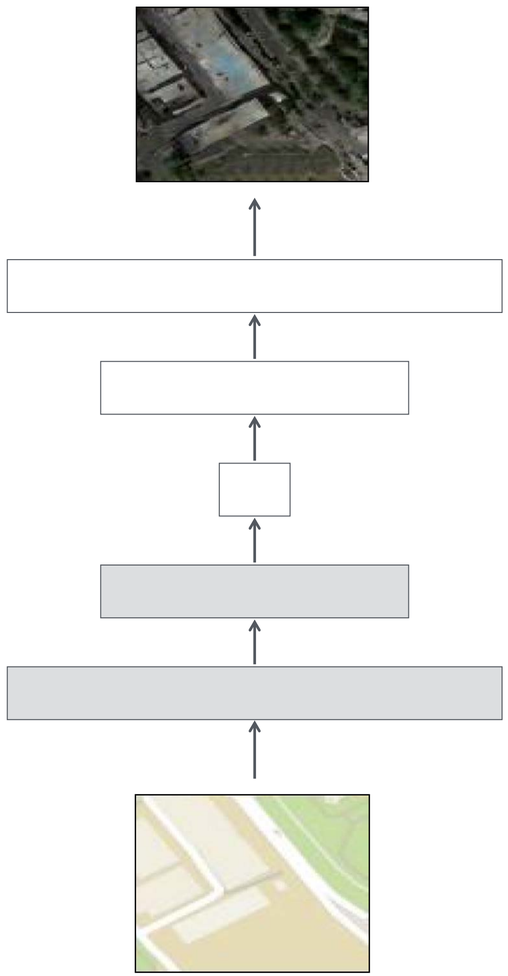


$$\arg\min_G \max_D \; \mathbb{E}_{\mathbf{x},\mathbf{y}}[\;\; \log D(\mathbf{x}, G(\mathbf{x})) + \log(1 - D(\mathbf{x}, \mathbf{y}))\;\;]$$

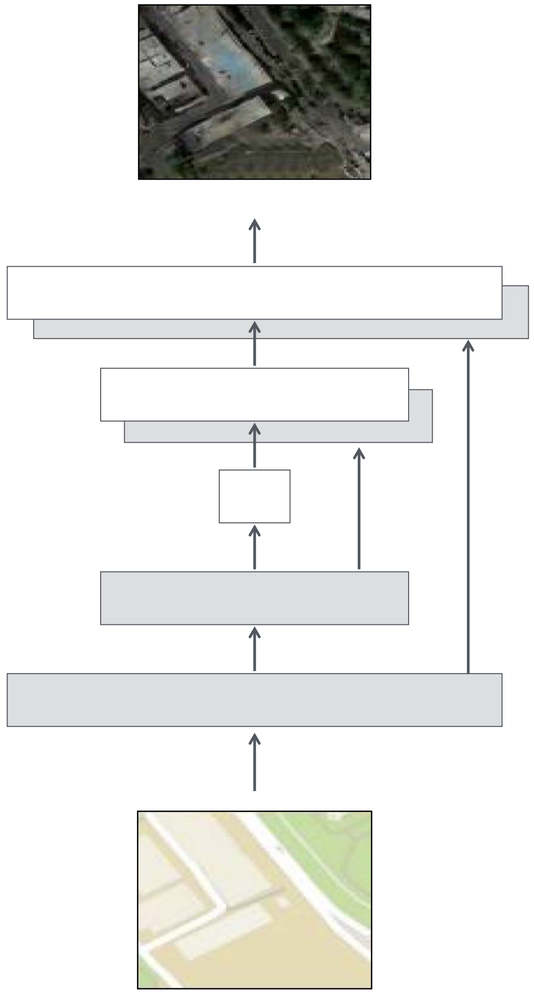
[Goodfellow et al., 2014]

[Isola et al., 2016]

Generator architecture



Encoder-decoder



U-Net

BW → Color

Input Output Input Output Input Output

Data from [Russakovsky et al. 2015]

Input

Output

Groundtruth

Data from
[maps.google.com]

[Isola et al, 2016]

Input

Output

Groundtruth

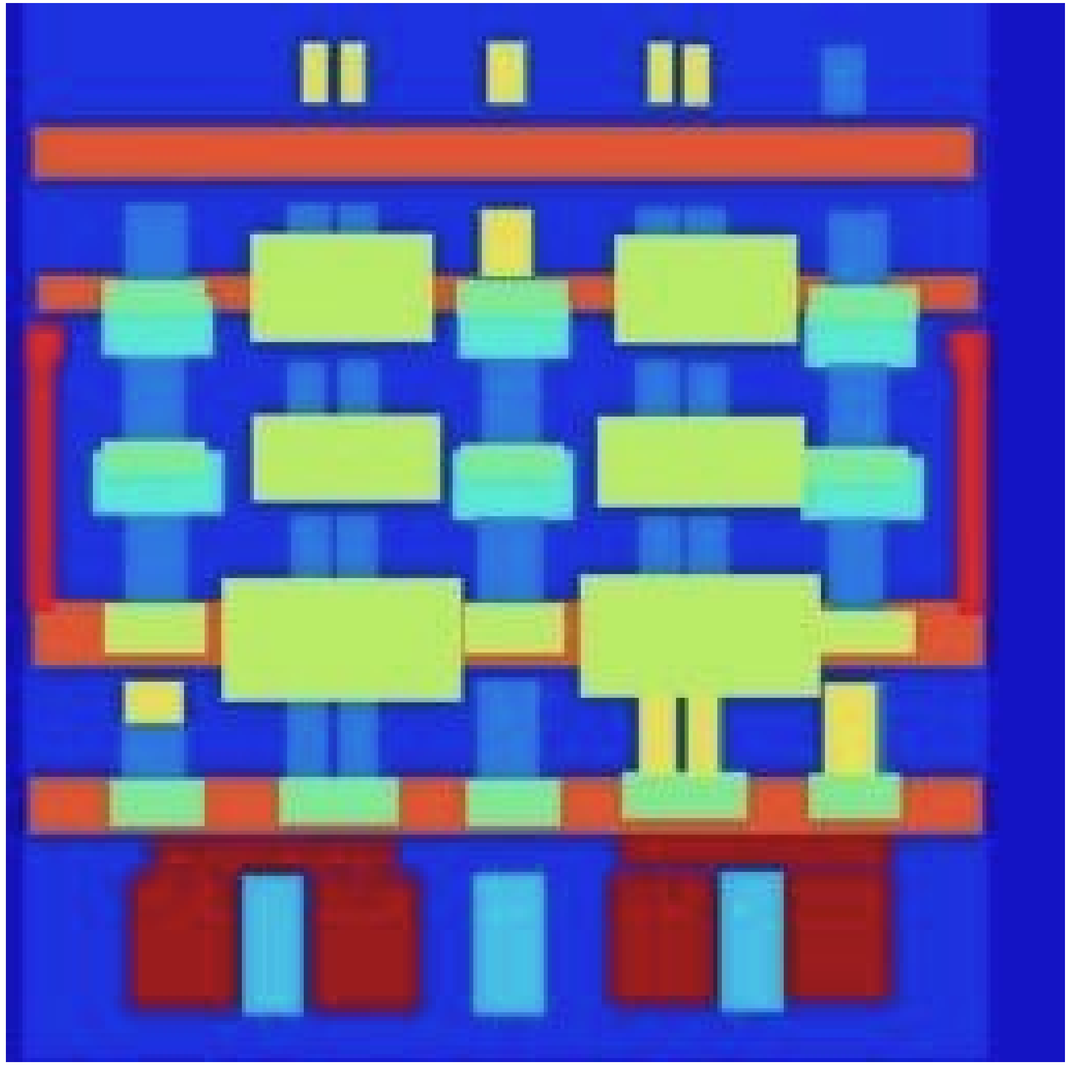
Data from [maps.google.com]

Map → Aerial Photo

Aerial Photo → Map

Labels → Facades

Input

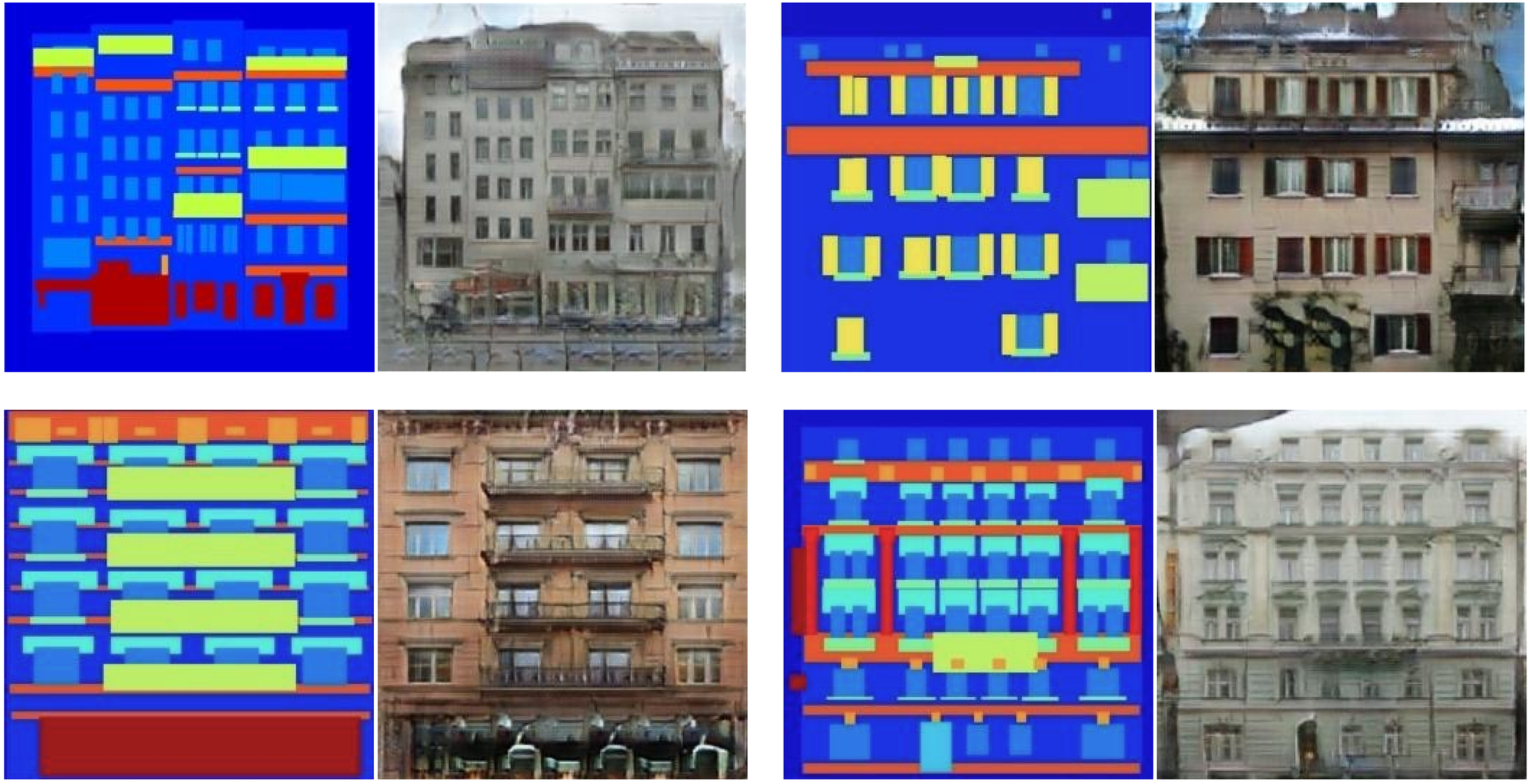


Output

Data from [Tylecek, 2013]

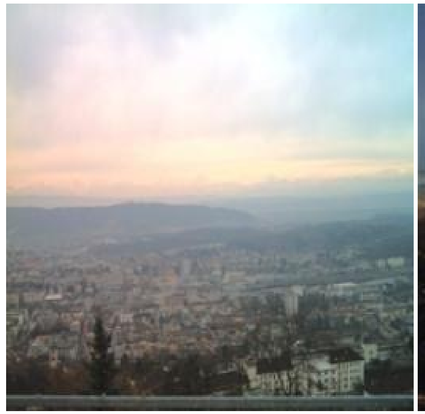
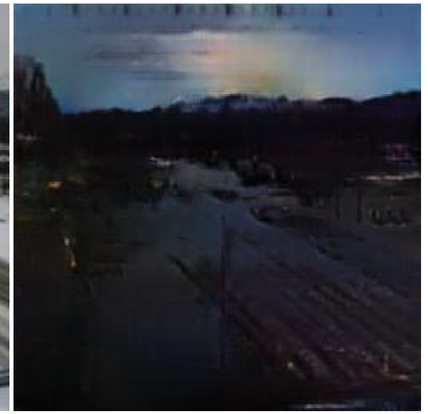
Labels → Facades

Input Output Input Output



Data from [Tylecek, 2013]

Day → Night

Input	Output	Input	Output	Input	Output
					
					

Data from [Laffont et al., 2014]

Edges → Images

Input Output Input Output Input Output

Edges from [Xie & Tu, 2015]

Sketches → Images

Input | Output | Input | Output | Input | Output

Trained on Edges → Images

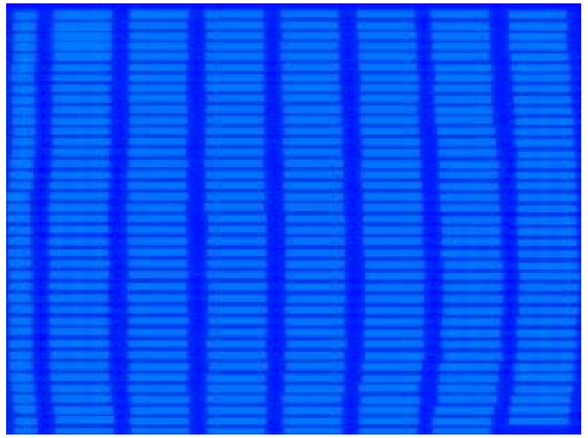
Data from [Eitz, Hays, Alexa, 2012]

Failure cases

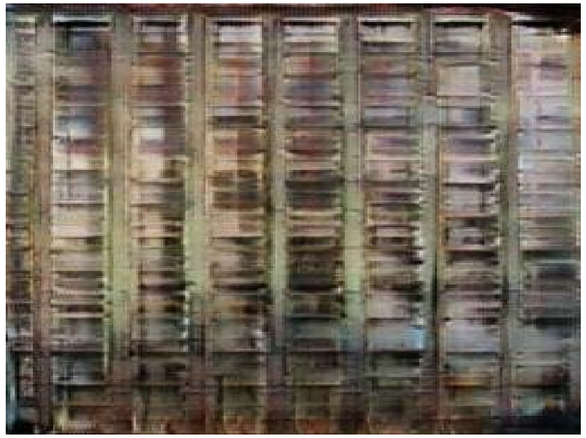
Input

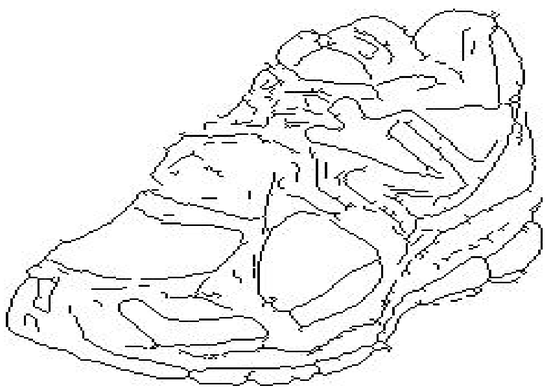
Output

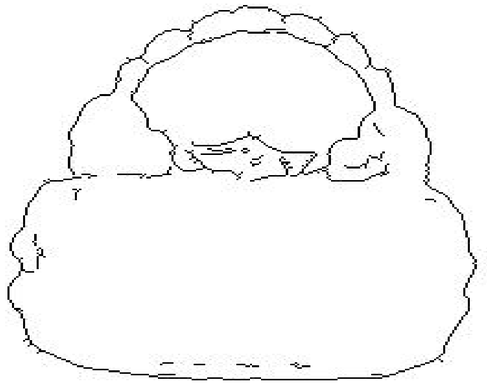
Input



Output







Code online: https://github.com/phillipi/pix2pix

phillipi / pix2pix

Unwatch 80 | Star 1,638 | Fork 193

Code | Issues 8 | Pull requests 1 | Projects 0 | Wiki | Pulse | Graphs | Settings

Image-to-image translation using conditional adversarial nets https://phillipi.github.io/pix2pix/

Edit

New Add topics

73 commits | 1 branch | 0 releases | 5 contributors

Branch: master | New pull request | Create new file | Upload files | Find file | Clone or download

junyanz committed on GitHub add day2night model — Latest commit c2bcba4 4 days ago

data	Add preprocess mode for inpainting	a month ago
datasets	add more datasets	2 months ago
imgs	smaller example image	3 months ago
models	add pre-trained models	2 months ago
scripts	update edge scripts	2 months ago
util	Merge pull request #24 from brannondorsey/plot-loss	12 days ago
.gitignore	add pre-trained models	2 months ago
LICENSE	adding license	3 months ago
README.md	add day2night model	4 days ago
models.lua	Added a generator net for 128x128 images	a month ago
test.lua	fix cudnn test error	2 months ago
train.lua	Merge pull request #24 from brannondorsey/plot-loss	12 days ago

Twitter-driven research: #pix2pix

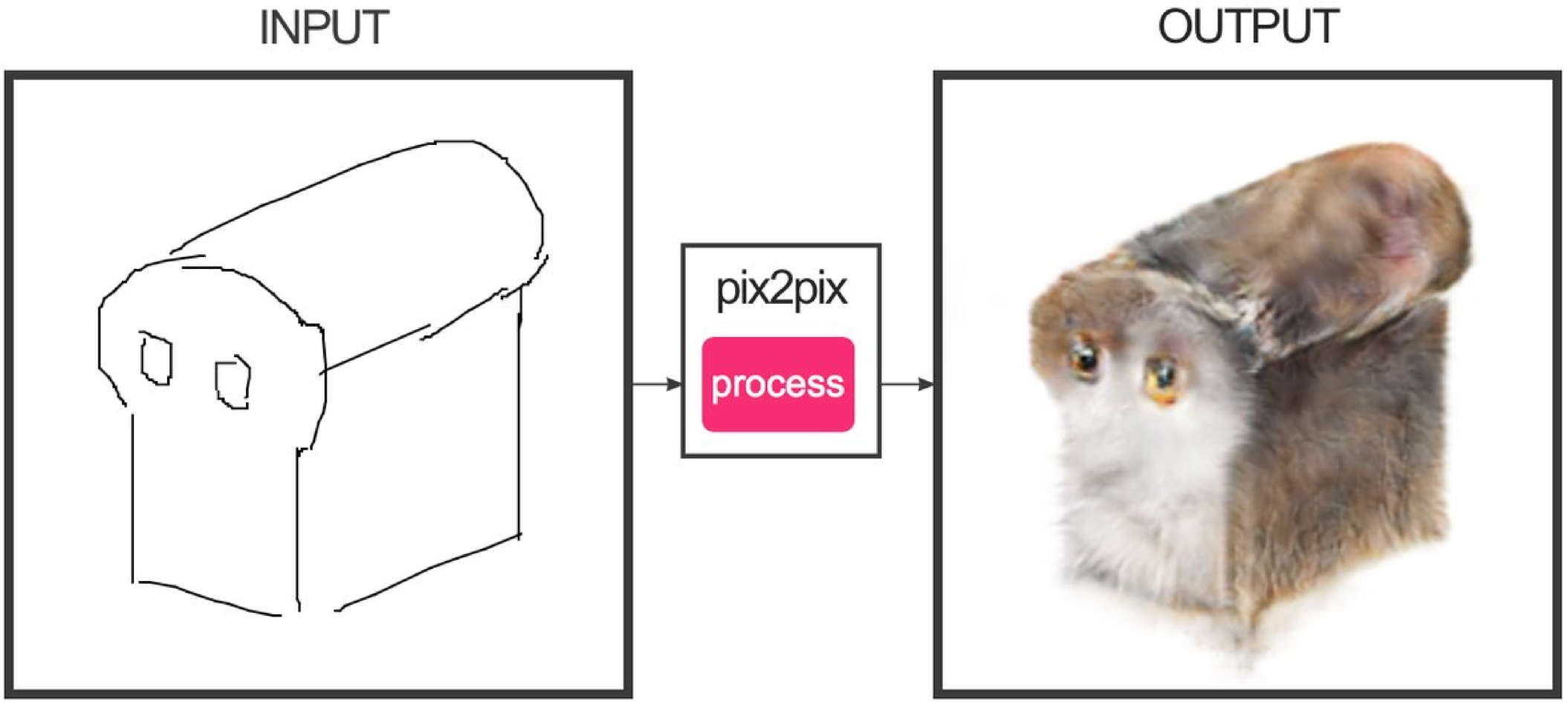
Kaihu Chen @kaihuchen

Mario Klingemann @quasimondo

Brannon Dorsey @brannondorsey

Bertrand Gondouin @bgondouin

Online Demo!



Pix2pix [Isola et al, 2016]