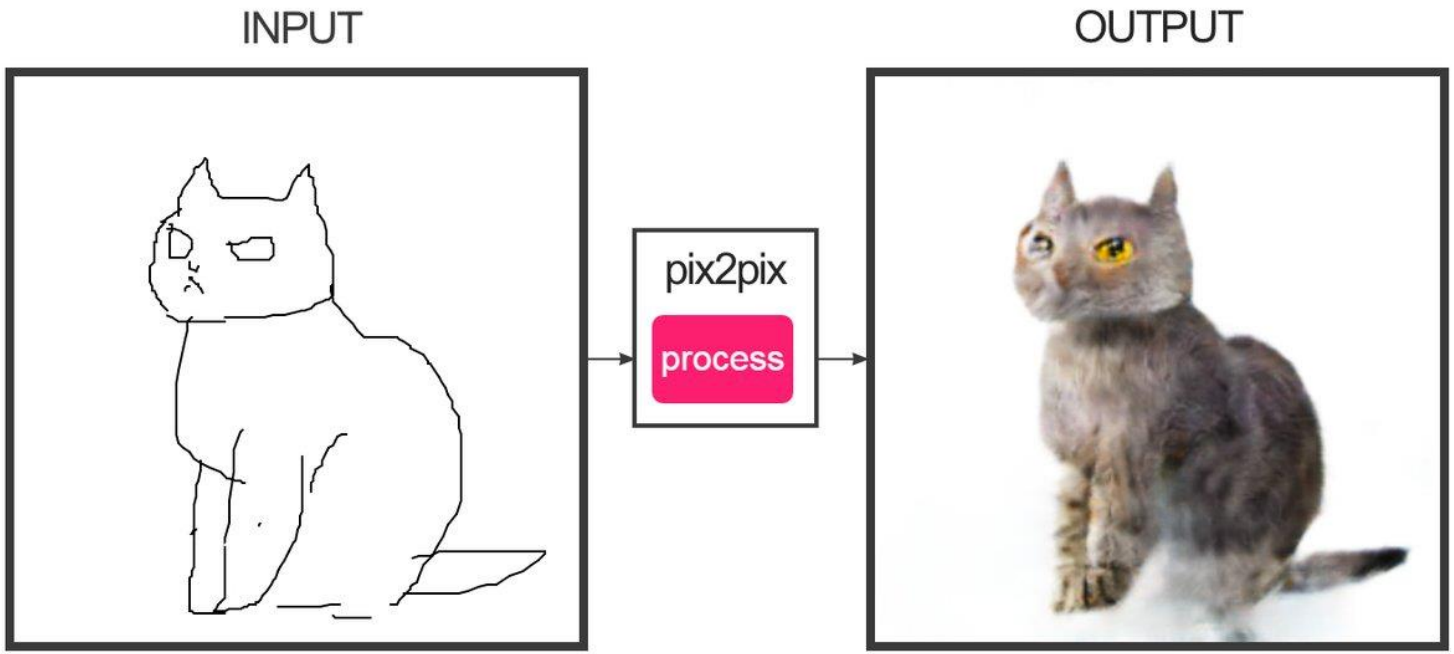
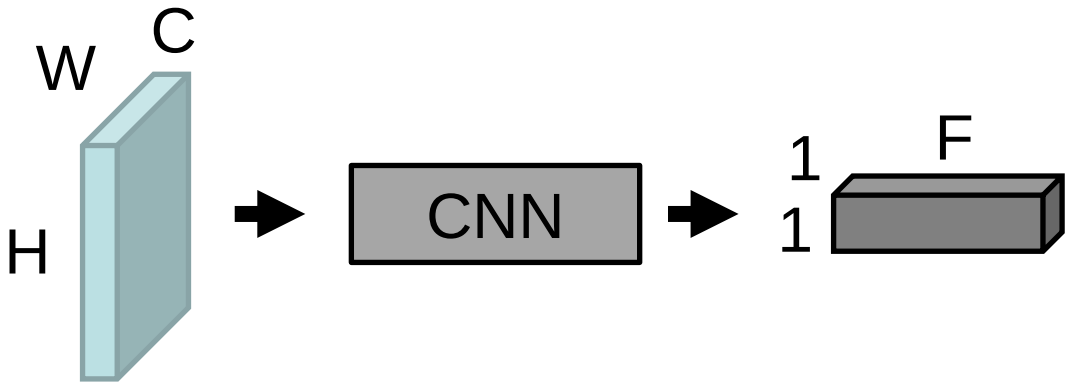
ConvNets as a Versatile Tool

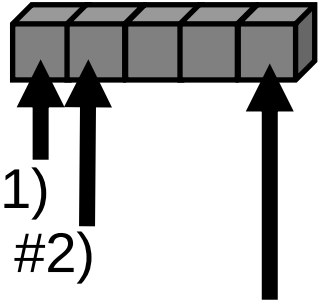


CS194: Intro to Comp. Vision and Comp. Photo
Alexei Efros, UC Berkeley, Fall 2021

Project 2 winner!

Classification Loss (e.g. ImageNet)





P(image is class #1)

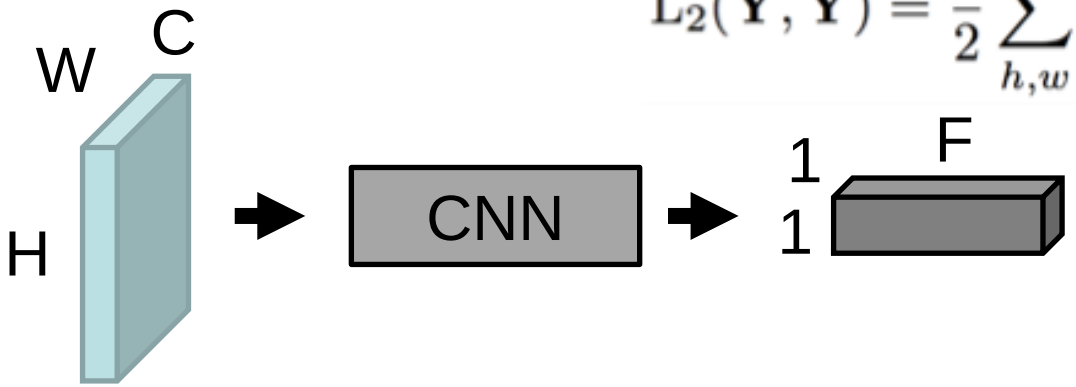
P(image is class #2)

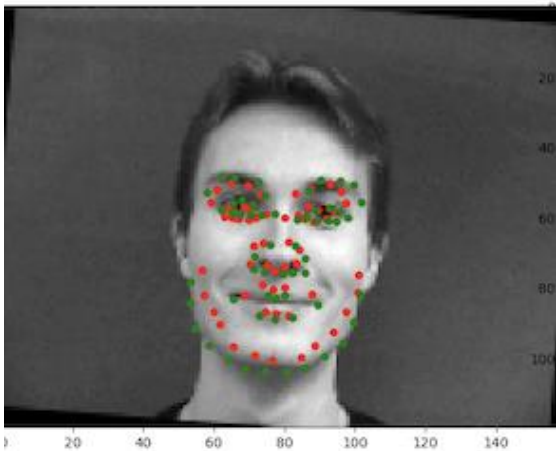
P(image is class #F)

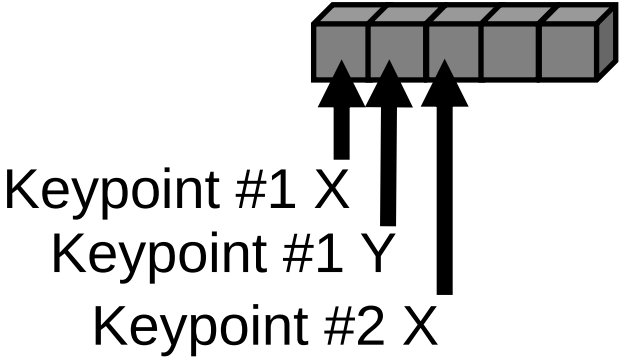
Slide by David Fouhey

Regression Loss (e.g. Project 5)

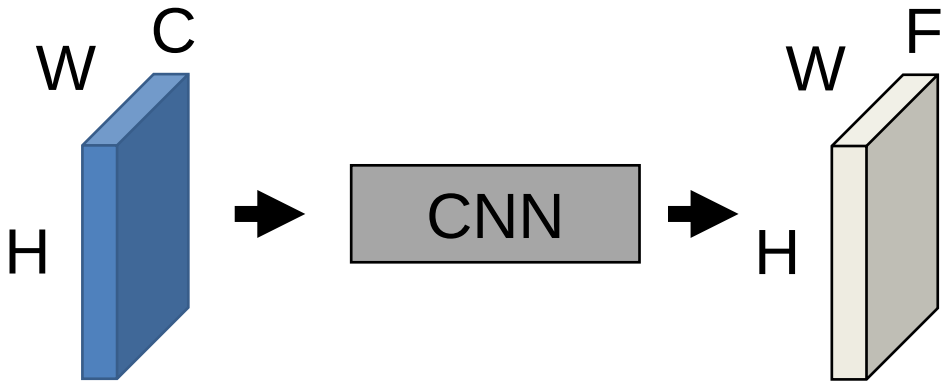
$$\mathrm{L}_2(\widehat{\mathbf{Y}}, \mathbf{Y}) = \frac{1}{2} \sum_{h,w} \| \mathbf{Y}_{h,w} - \widehat{\mathbf{Y}}_{h,w} \|_2^2$$







Pixel Labeling



<u>Convert HxW image into a F-dimensional vector</u>

Which pixels in this image are a cat?
How far is each pixel away from the camera?
Which pixels of this image are fake?

“Semantic Segmentation”

Each pixel has label, inc. **background**, and unknown
Usually visualized by colors.
Note: don’t distinguish between object *instances*

Input Label Input Label

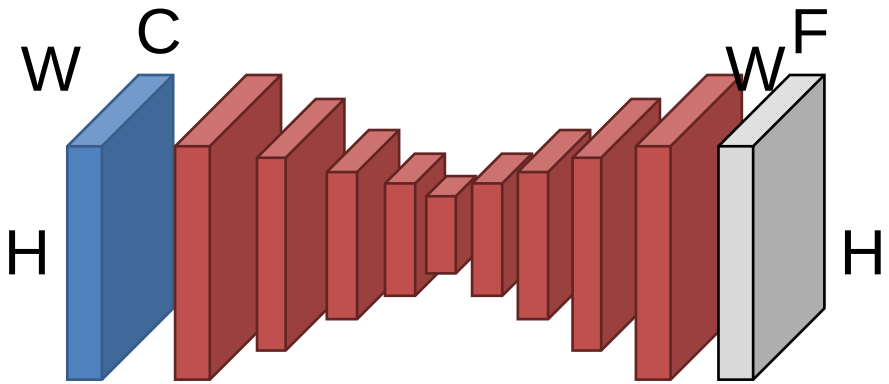
Image Credit: Everingham et al. Pascal VOC 2012.

Encoder-Decoder

Key idea: First **downsample** towards middle of network. Then **upsample** from middle.

How do we downsample?

Convolutions, pooling

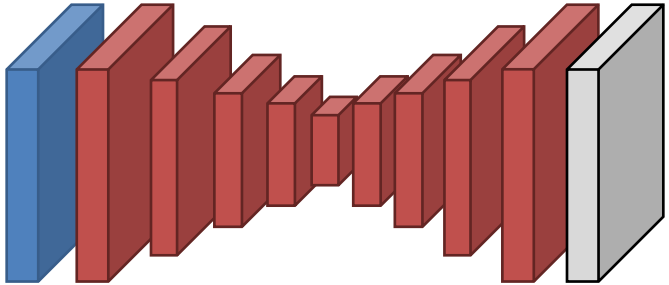


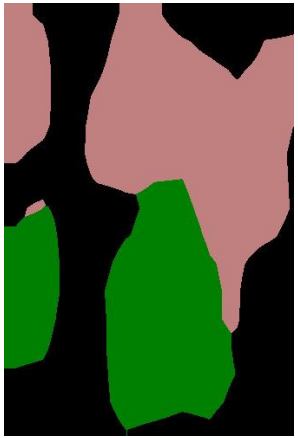
Missing Details

While the output *is* HxW, just upsampling often produces results without details/not aligned with the image.

Why?

Information about details lost when downsampling!

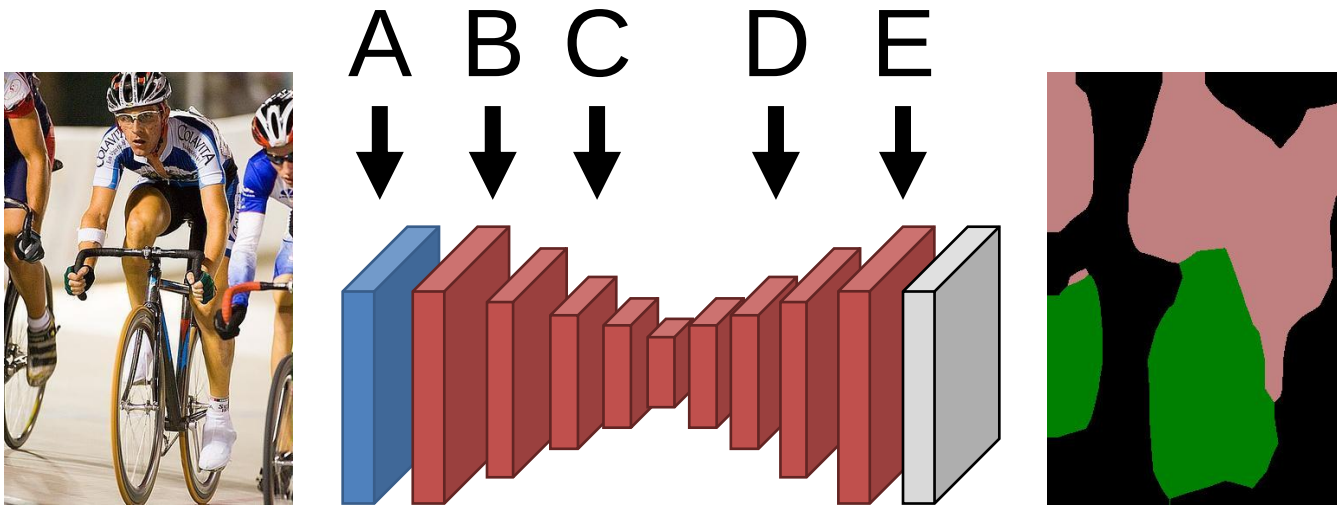




Result from Long et al. *Fully Convolutional Networks For Semantic Segmentation*. CVPR 2014

Missing Details

Where is the useful information about the high-frequency details of the image?



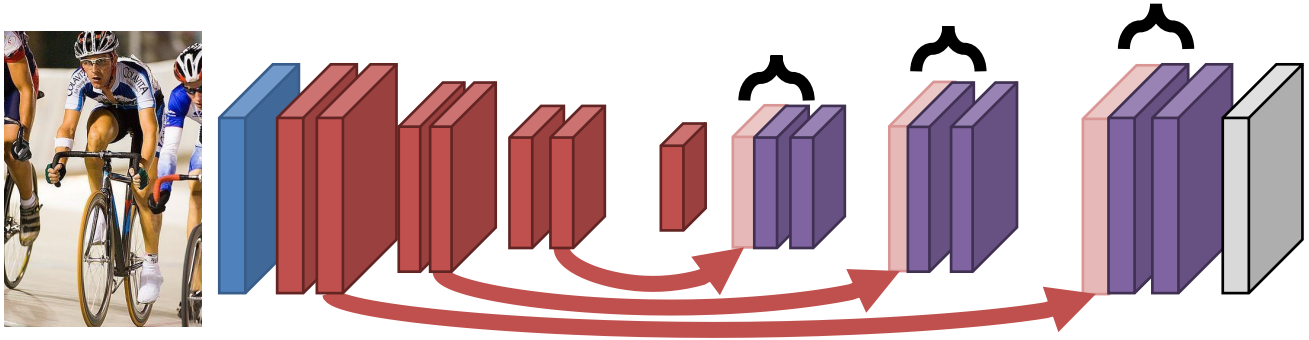
Result from Long et al. *Fully Convolutional Networks For Semantic Segmentation*. CVPR 2014

Missing Details

How do you send details forward in the network?

You copy the activations forward.

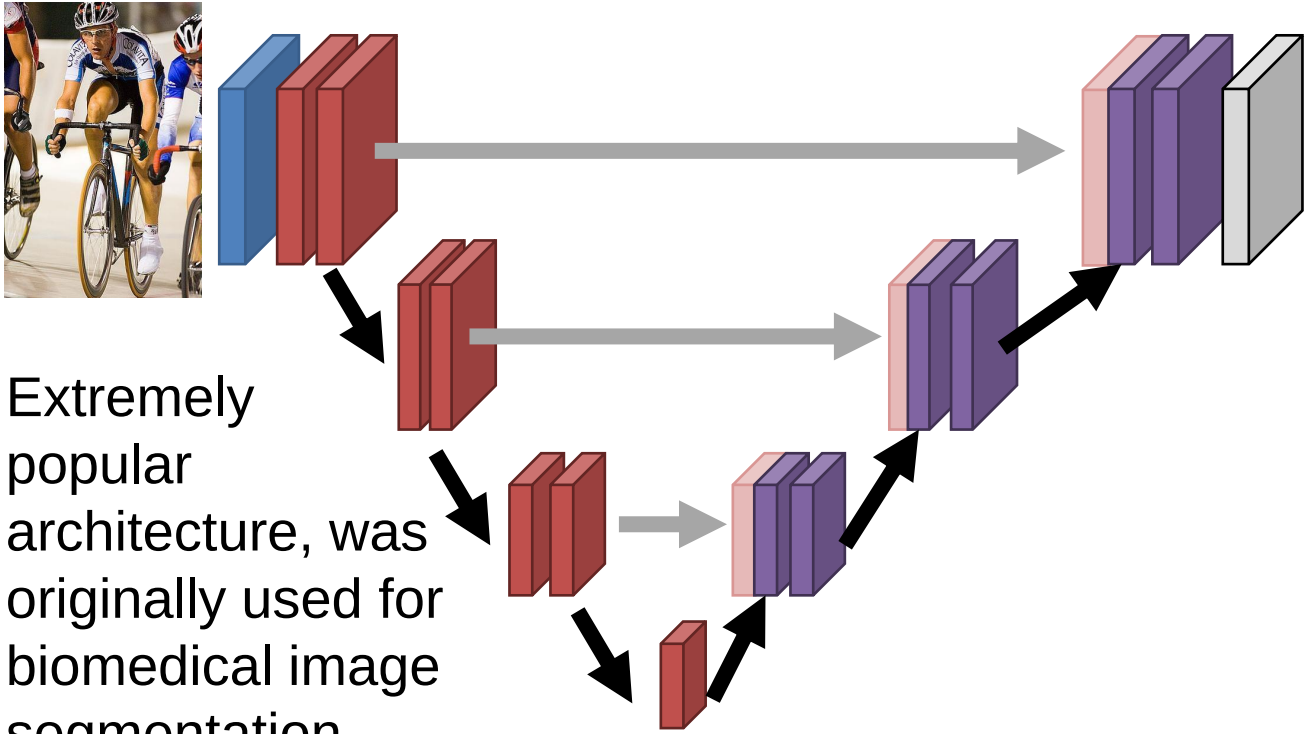
Subsequent layers at the same resolution figure out how to fuse things.



Copy

Result from Long et al. *Fully Convolutional Networks For Semantic Segmentation*. CVPR 2014

U-Net



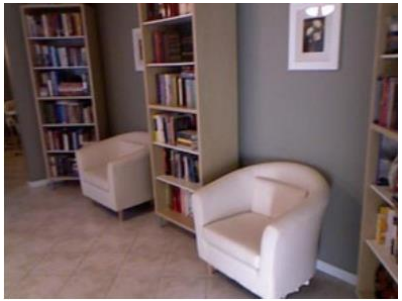
Extremely popular architecture, was originally used for biomedical image segmentation.

Ronneberger et al. “U-Net: Convolutional Networks for Biomedical Image Segmentation. MICCAI 2015

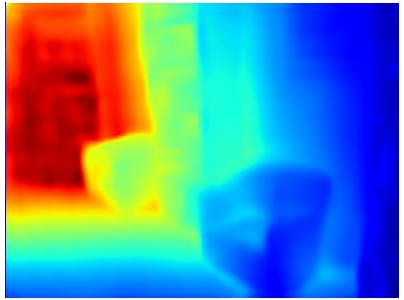
e.g. Depth Prediction

Instead: give label of depthmap, train network to do regression (e.g., $\|z_i - \hat{z}_i\|$ where z_i is the ground-truth and $\hat{z}_i$ the prediction of the network at pixel i).

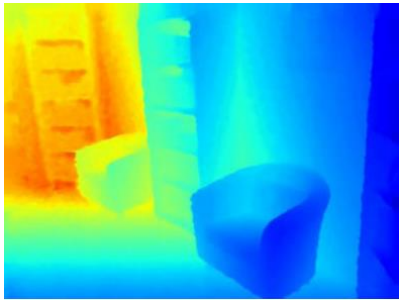
Input HxWx3
RGB Image



Output HxWx1
Depth Image

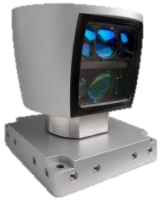


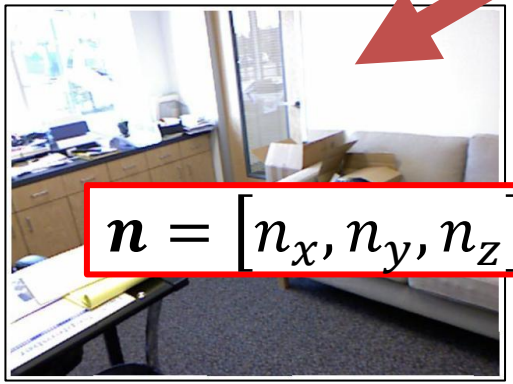
True HxWx1
Depth Image



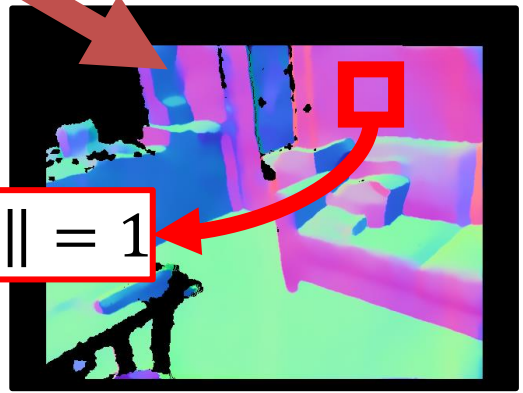
Result credit: Eigen and Fergus, ICCV 2015

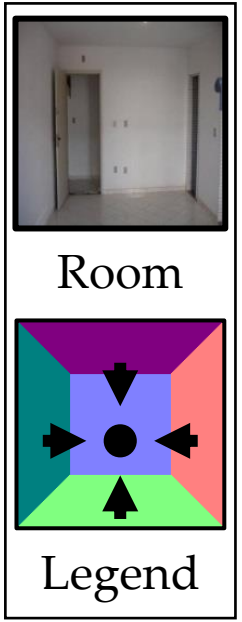
Surface Normals





$$\boldsymbol{n} = [n_x, n_y, n_z], \|\boldsymbol{n}\| = 1$$





Color Image

Normals

Image credit: NYU Dataset, Silberman et al. ECCV 2012

Surface Normals

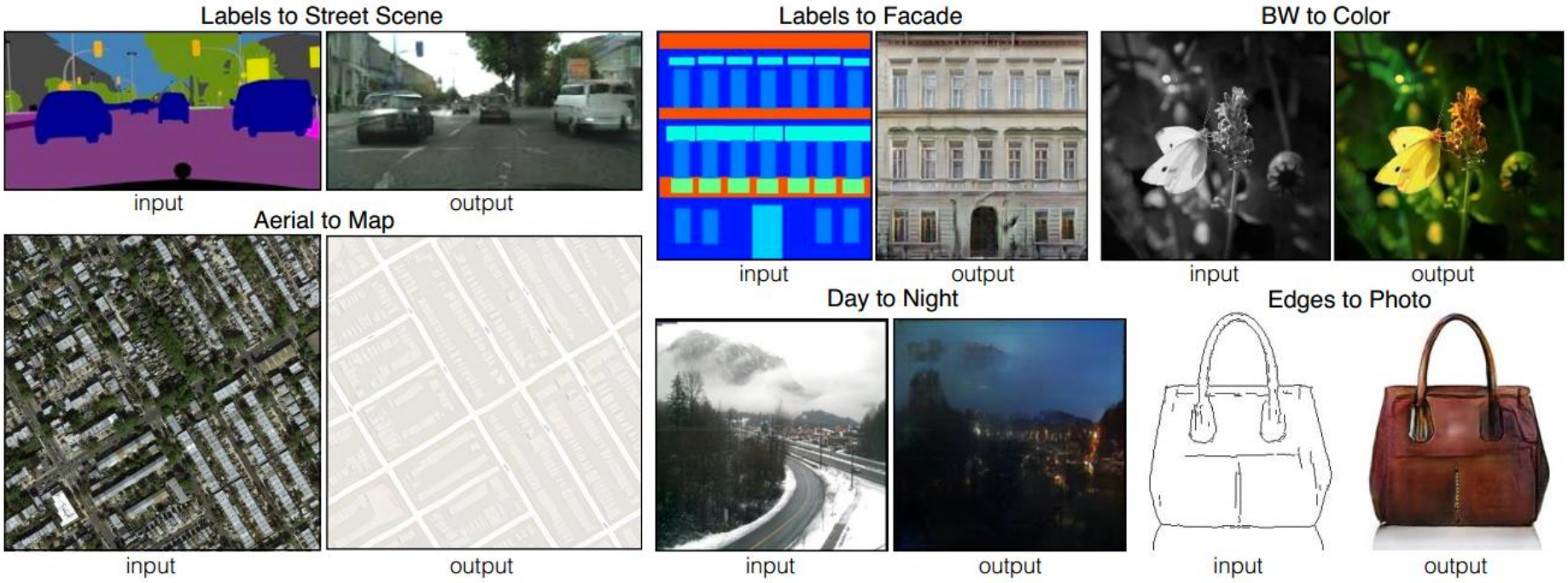
Instead: train normal network to minimize $\|\boldsymbol{n}_i - \widehat{\boldsymbol{n}_i}\|$ where $\boldsymbol{n}_i$ is ground-truth and $\widehat{\boldsymbol{n}_i}$ prediction at pixel i.

Input: HxWx3
RGB Image

Output: HxWx3
Normals

Result credit: X. Wang, D. Fouhey, A. Gupta, *Designing Deep Networks for Surface Normal Estimation.* CVPR 2014

Generic Image-to-Image Translation



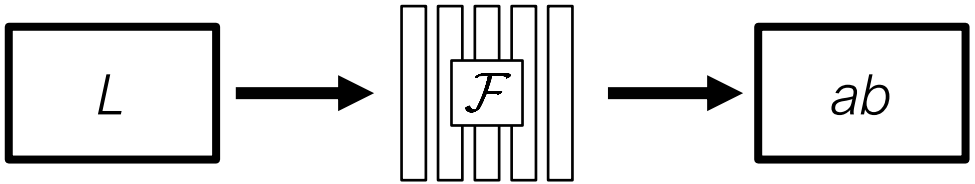
$\mathcal{F}$

Grayscale image: *L* channel

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

Color information: *ab* channels

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



Zhang, Isola, Efros. ***Colorful Image Colorization.*** In *ECCV*, 2016.

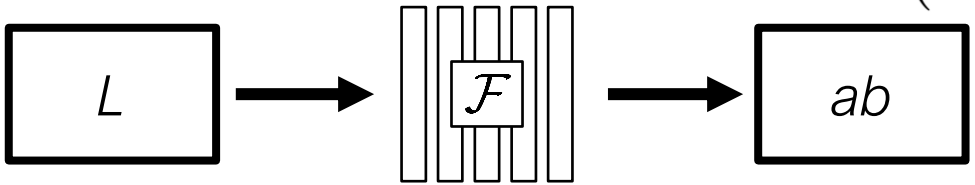
$\mathcal{F}$

Grayscale image: *L* channel

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

Concatenate (*L*,*ab*) channels

$$(\mathbf{X}, \widehat{\mathbf{Y}})$$



Zhang, Isola, Efros. ***Colorful Image Colorization.*** In *ECCV*, 2016.

Simple L2 regression doesn't work ☹

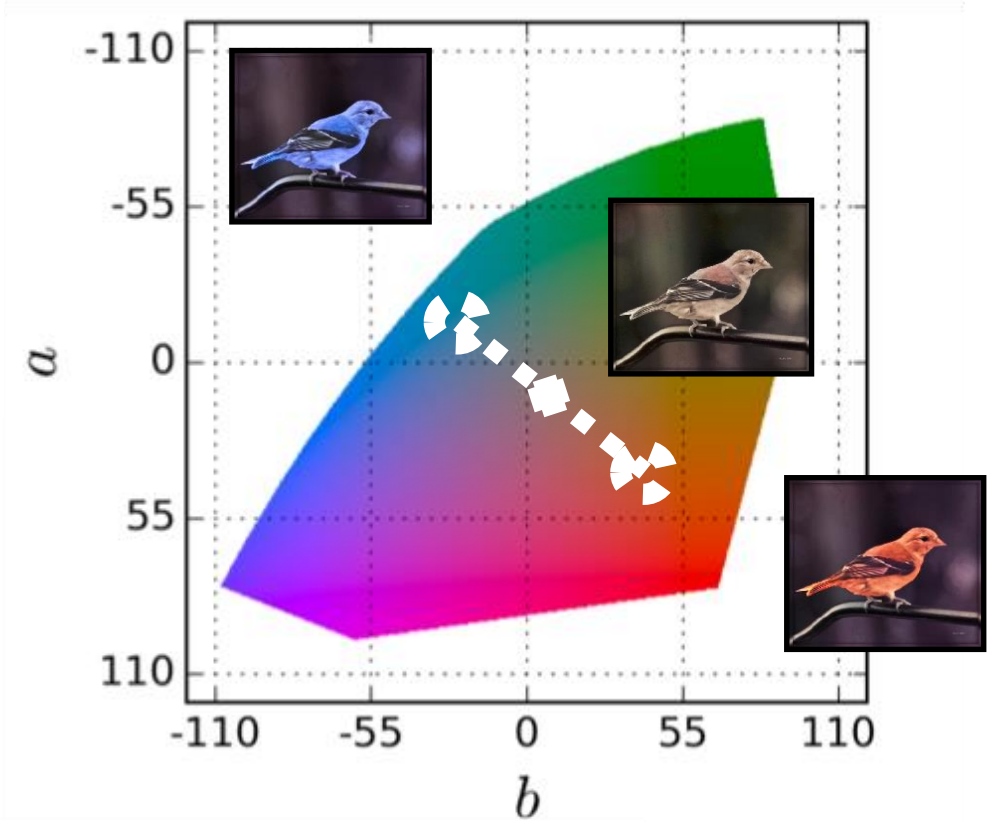
Input

Output

Ground truth

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

Slide by Richard Zhang



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

Slide by Richard Zhang

Better Loss Function

$$\theta^* = \arg\min_{\theta} \ell(\mathcal{F}_\theta(\mathbf{X}), \mathbf{Y})$$

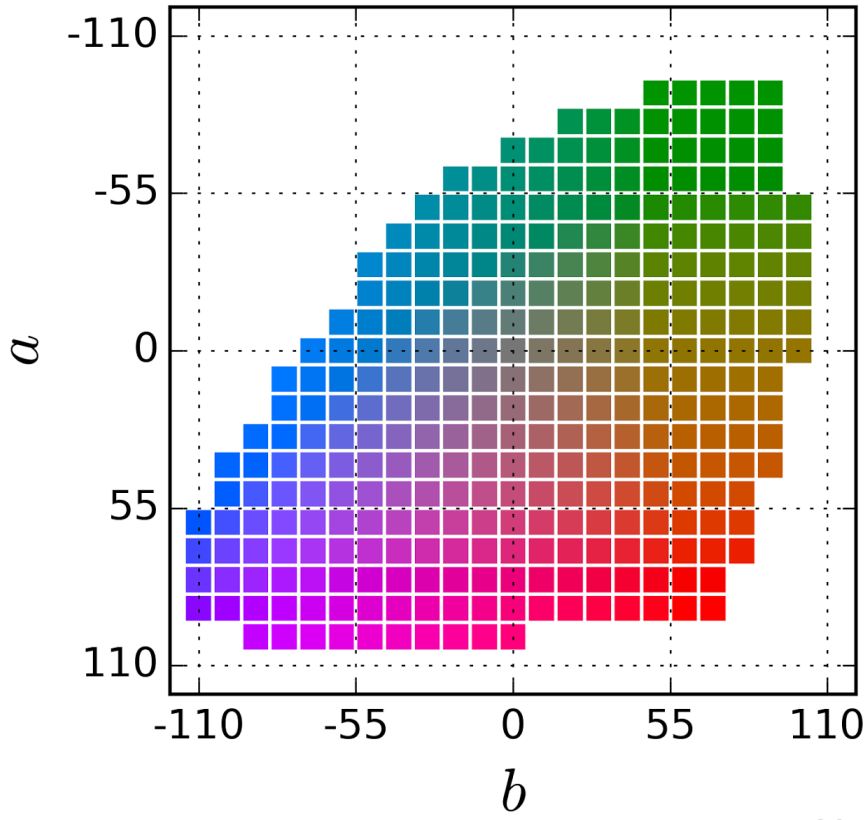
- Regression with L2 loss inadequate

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

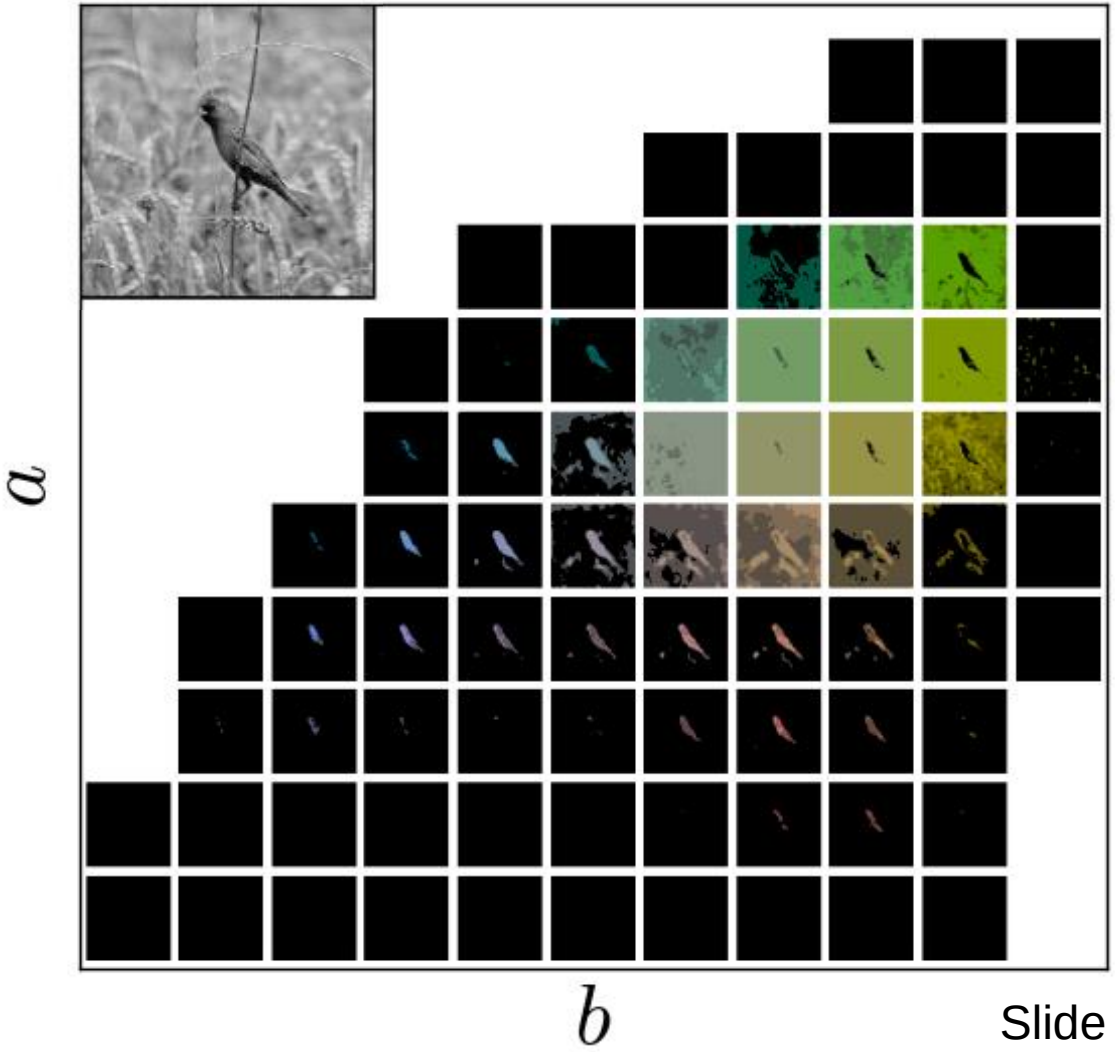
- Use per-pixel multinomial classification

$$\mathrm{L}(\widehat{\mathbf{Z}}, \mathbf{Z}) = -\frac{1}{HW} \sum_{h,w} \sum_{q} \mathbf{Z}_{h,w,q} \log(\widehat{\mathbf{Z}}_{h,w,q})$$

Colors in *ab* space
(discrete)



Slide by Richard Zhang



Slide by Richard Zhang

Designing loss functions

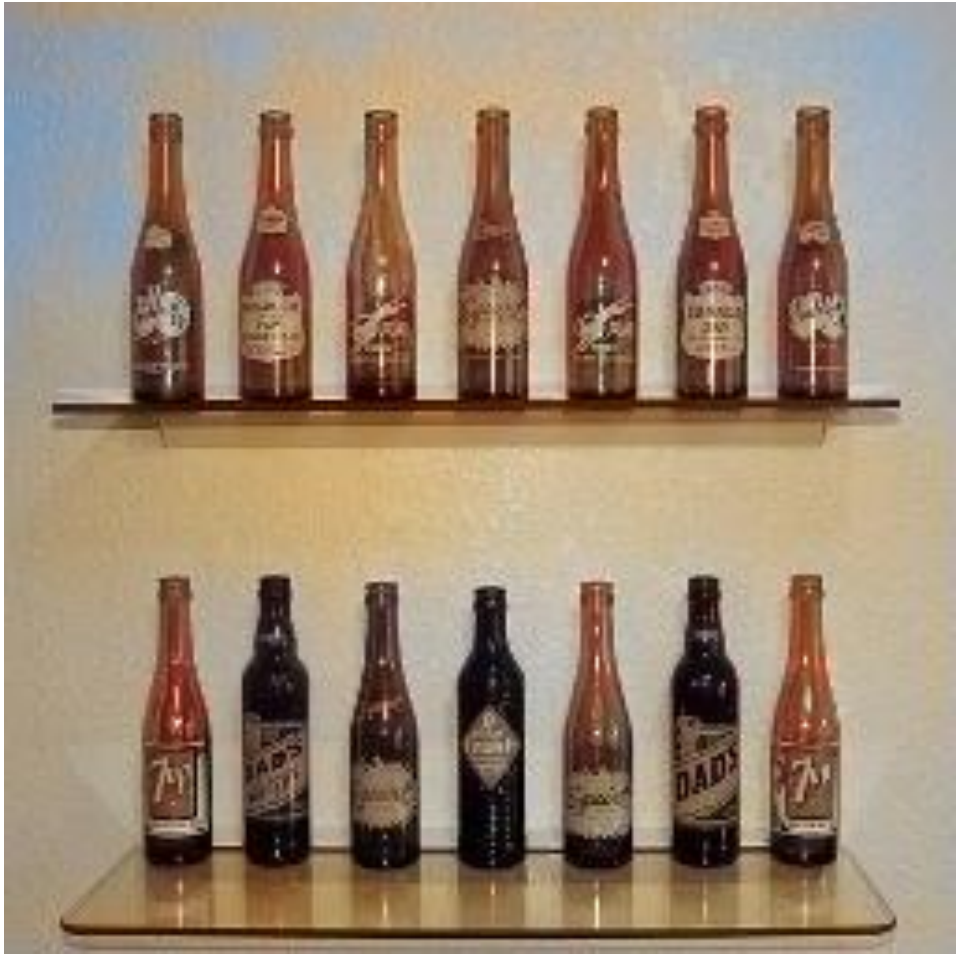
Input

Zhang et al. 2016

Ground truth

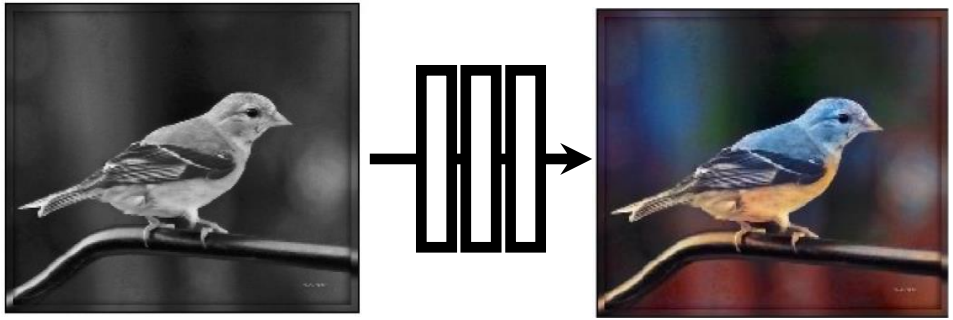
Color distribution cross-entropy loss with colorfulness enhancing term.

[Zhang, Isola, Efros, ECCV 2016]


DADS

Designing loss functions

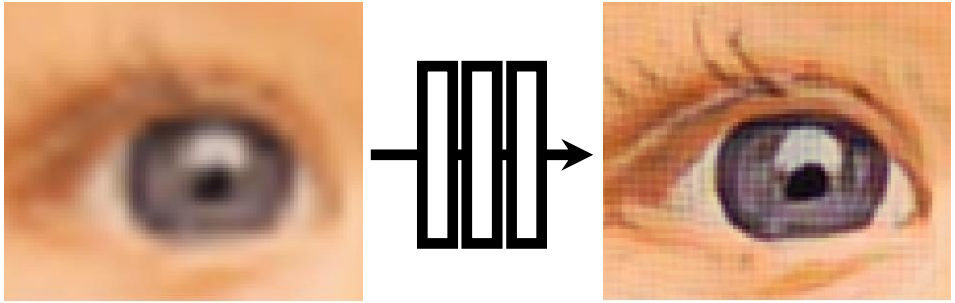
Image colorization



[Zhang et al. 2016]

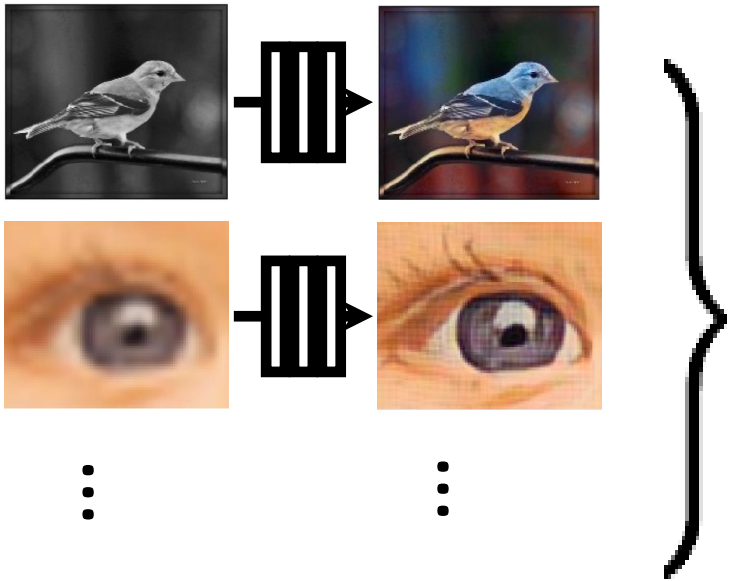
Cross entropy loss, with colorfulness term

Super-resolution



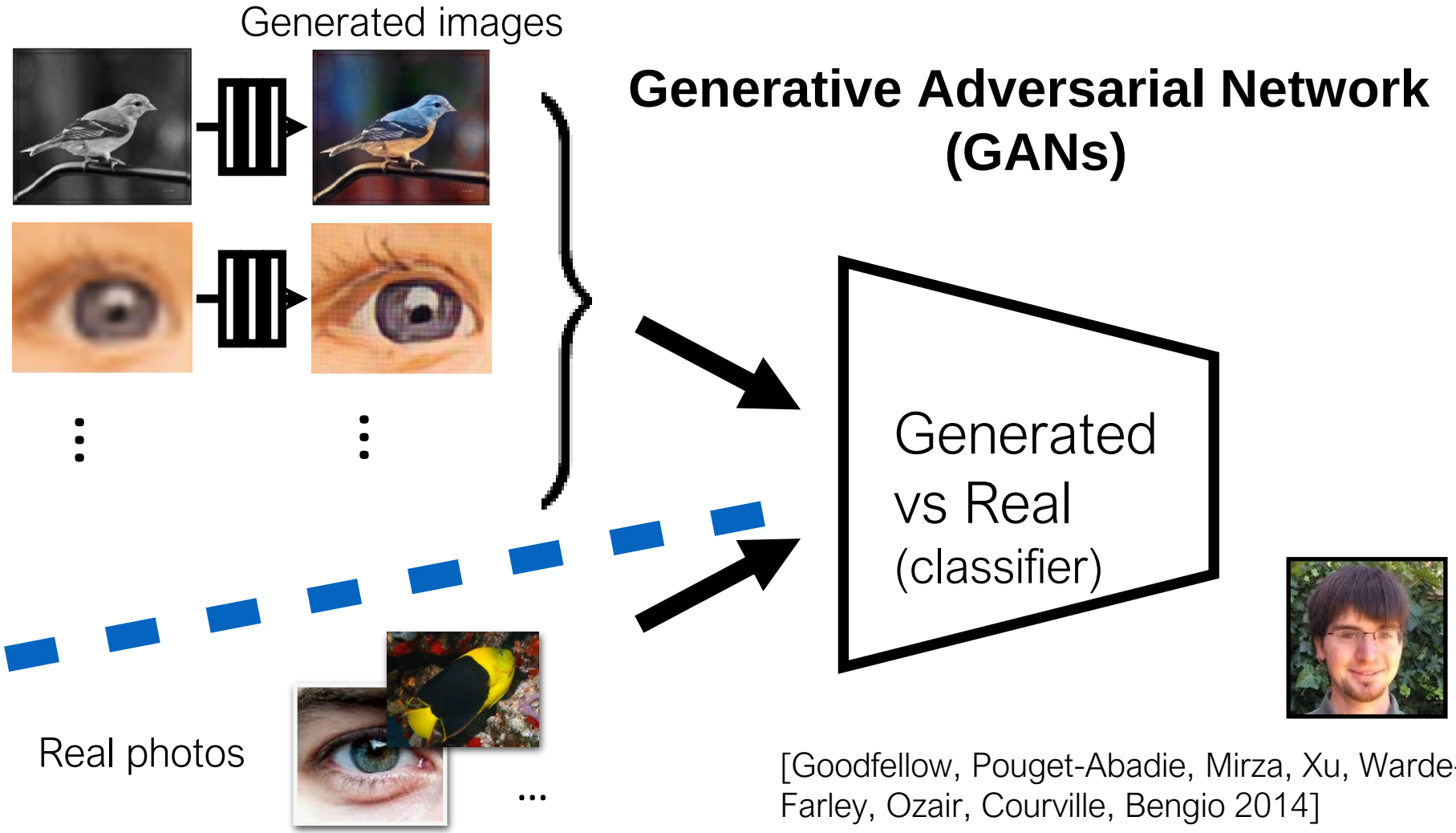
[Johnson et al. 2016]

“semantic feature loss” (VGG feature covariance matching objective)

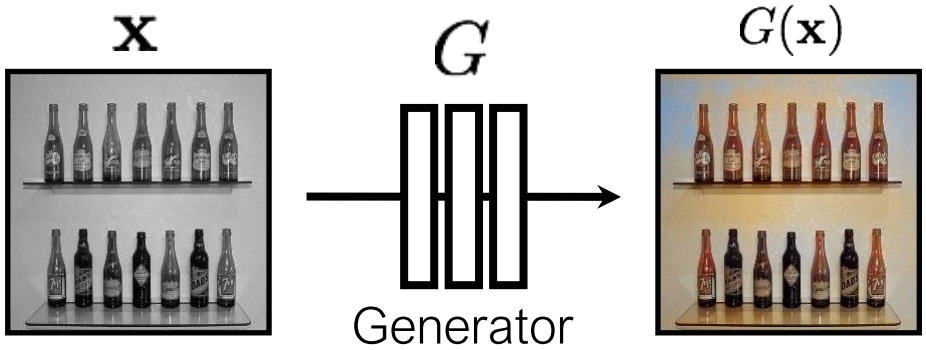


Universal loss?

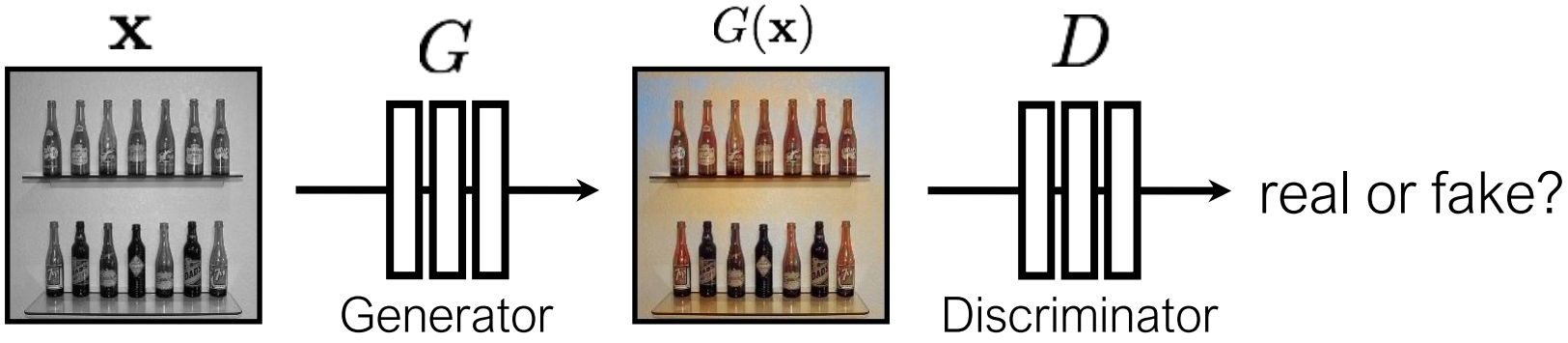
Generative Adversarial Network (GANs)



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



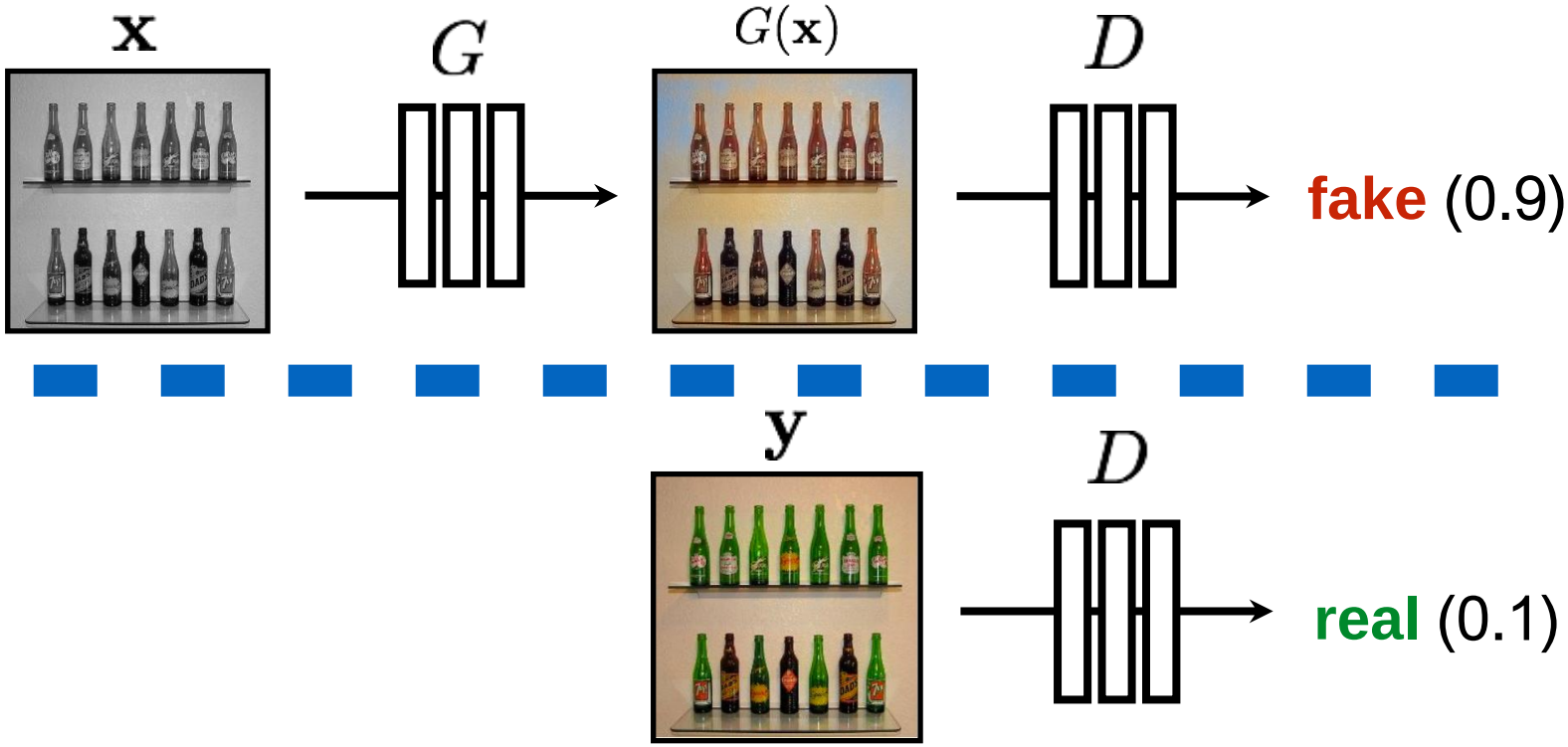
[Goodfellow et al., 2014]



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

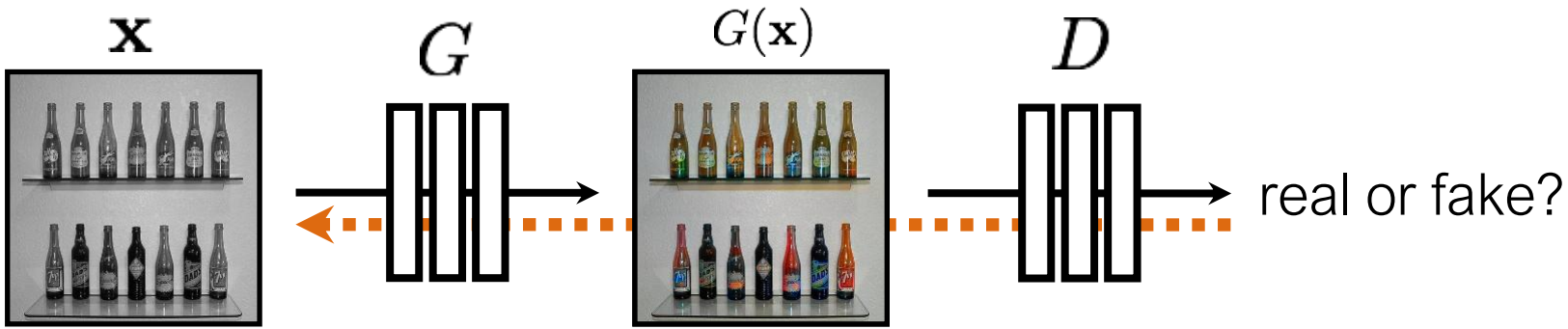
D tries to identify the fakes

[Goodfellow et al., 2014]



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

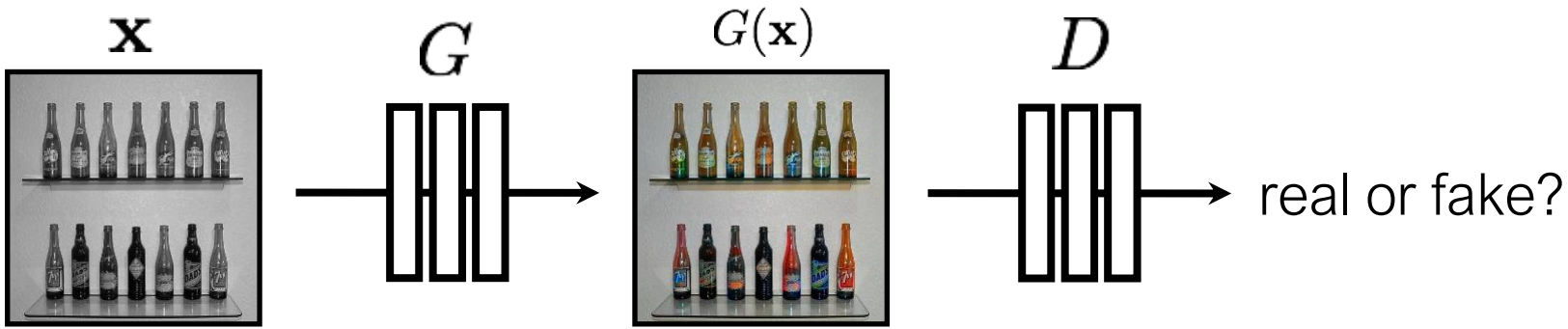
[Goodfellow et al., 2014]



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

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

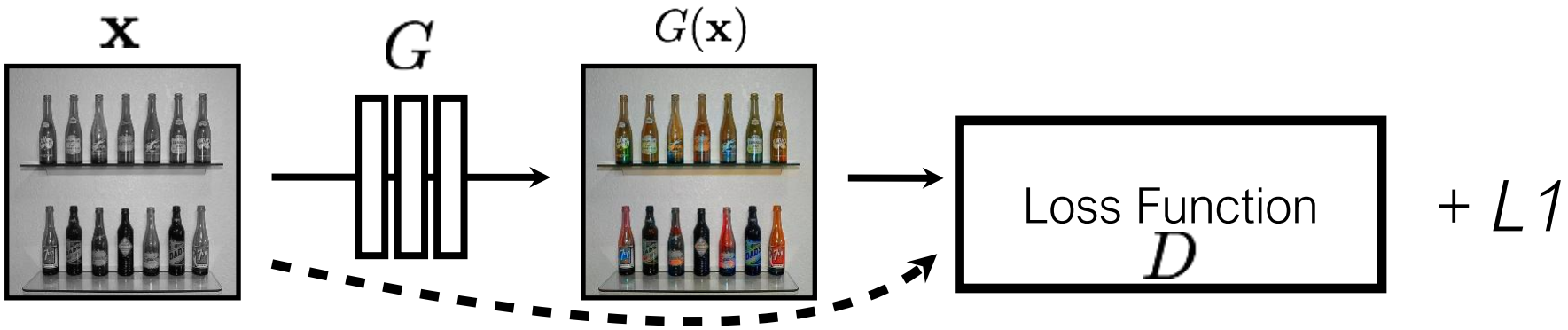
[Goodfellow et al., 2014]



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

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

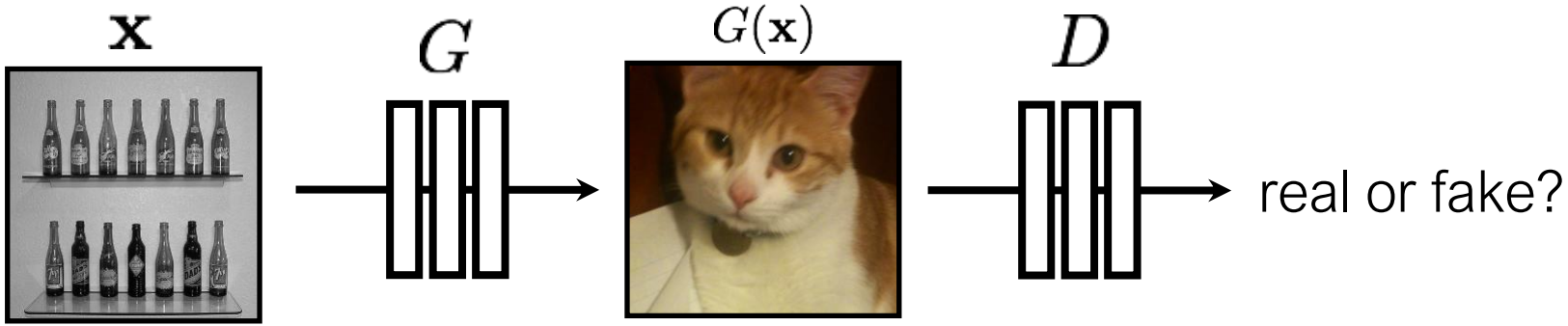
[Goodfellow et al., 2014]



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

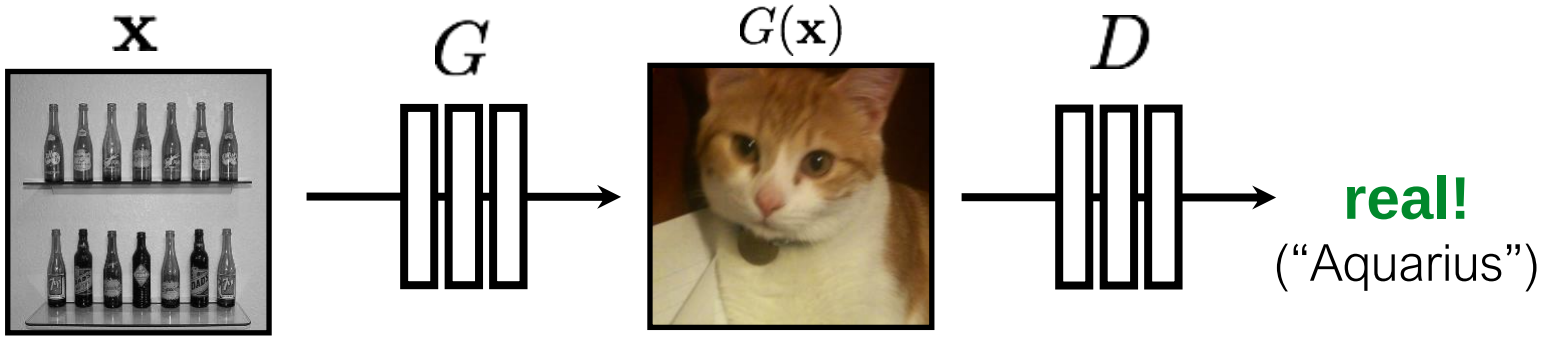
Rather than being hand-designed, it is *learned*.

[Goodfellow et al., 2014]
[Isola et al., 2017]



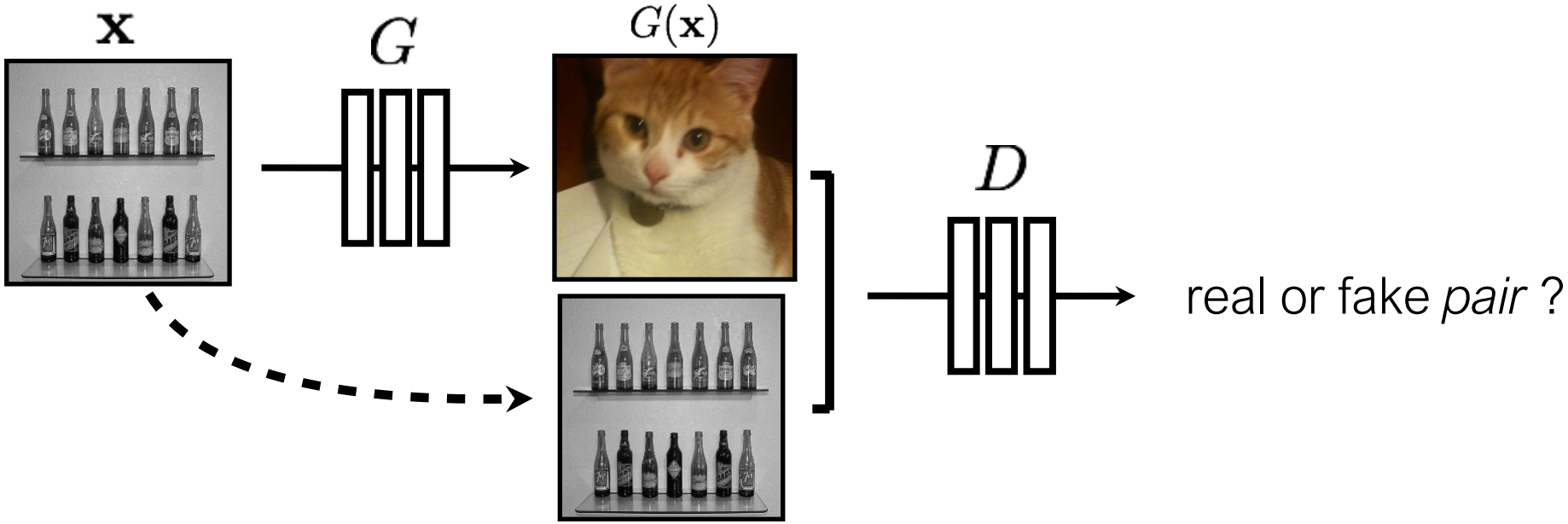
$$\arg\min_{G}\max_{D}\ \mathbb{E}_{\mathbf{x},\mathbf{y}}[\ \log D(G(\mathbf{x})) \quad + \quad \log(1 - D(\mathbf{y}))\ ]$$

[Goodfellow et al., 2014]



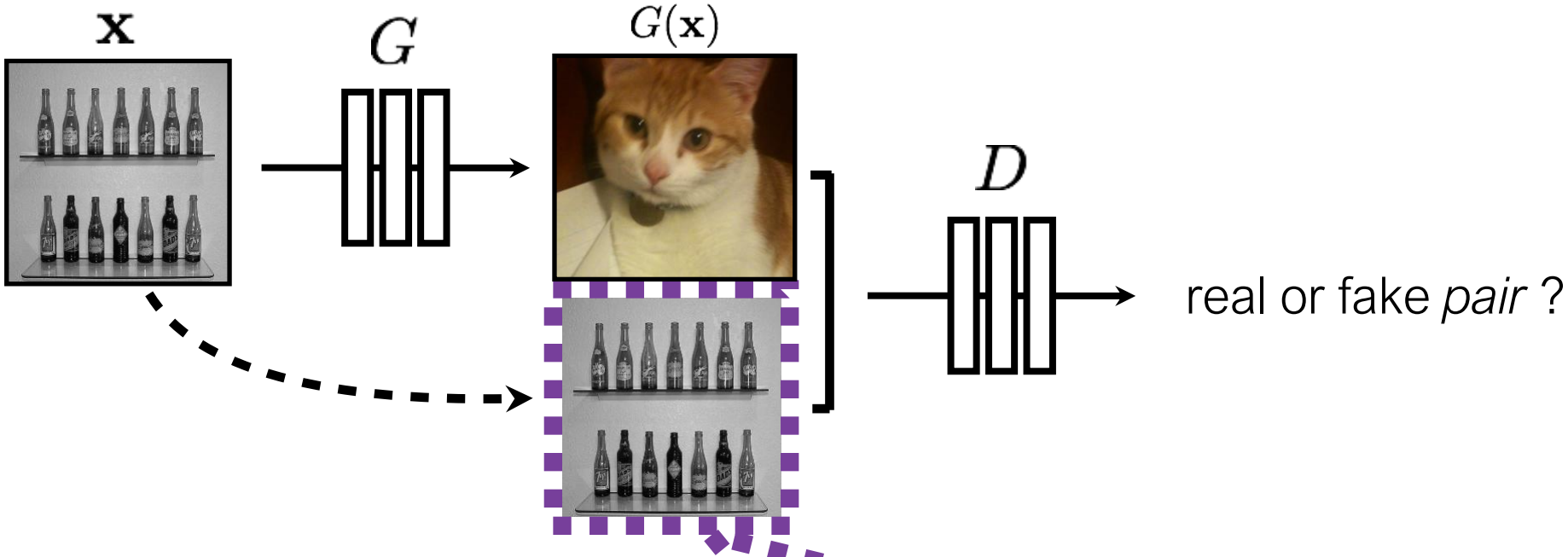
$$\arg\min_{G}\max_{D}\ \mathbb{E}_{\mathbf{x},\mathbf{y}}[\ \log D(G(\mathbf{x})) \quad + \quad \log(1 - D(\mathbf{y}))\ ]$$

[Goodfellow et al., 2014]



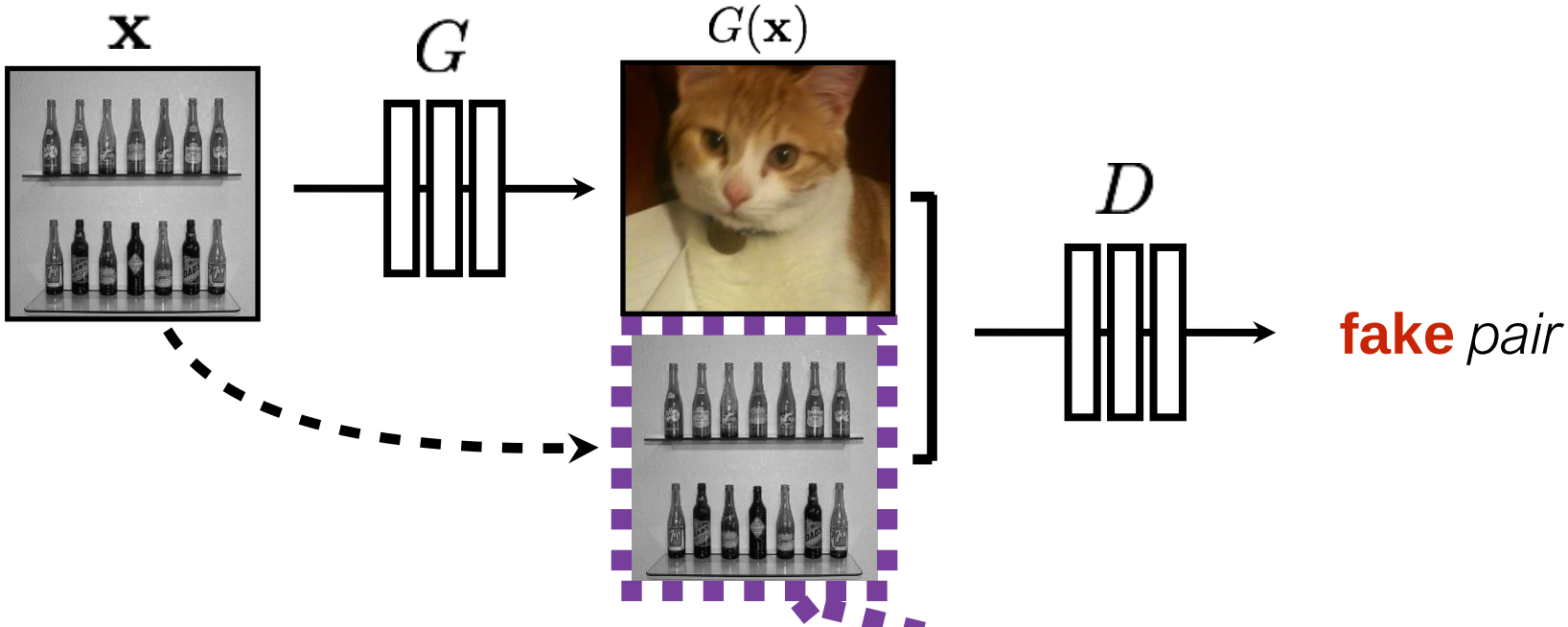
$$\arg\min_{G}\max_{D}\ \mathbb{E}_{\mathbf{x},\mathbf{y}}[\ \log D(G(\mathbf{x})) \quad + \quad \log(1 - D(\mathbf{y}))\ ]$$

[Goodfellow et al., 2014]
[Isola et al., 2017]



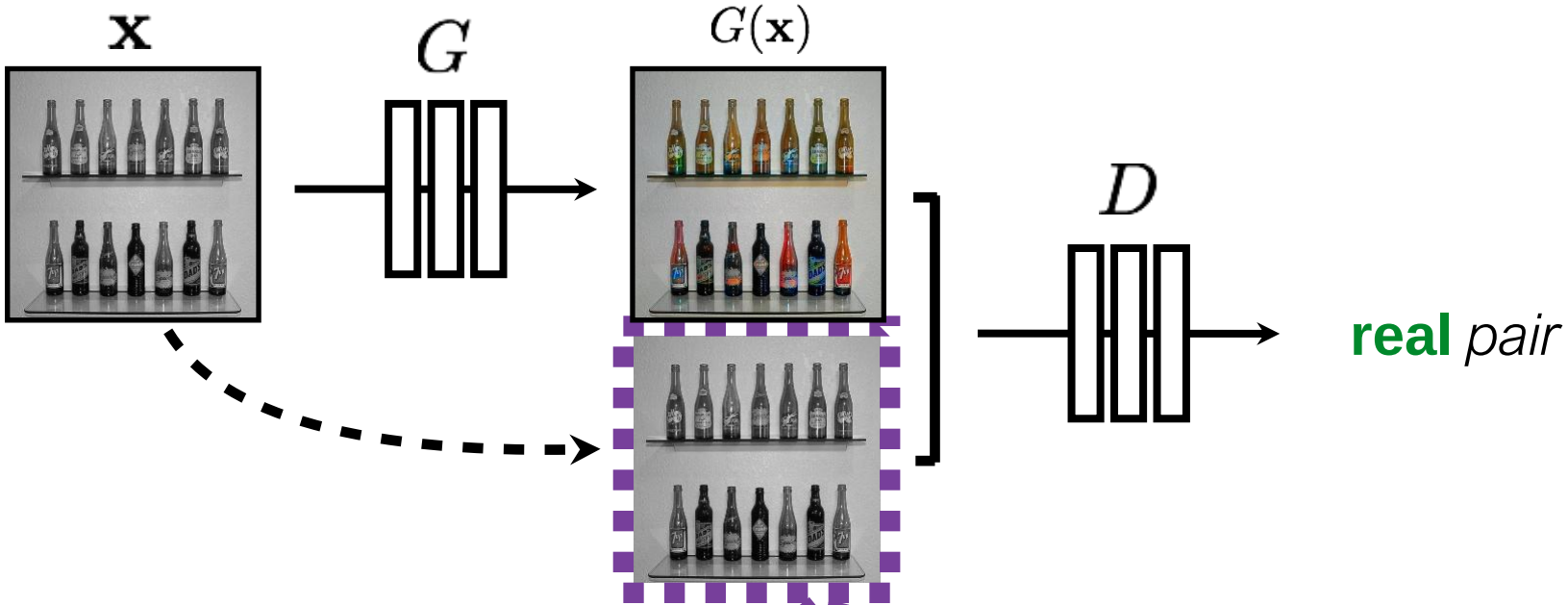
$$\arg\min_G \max_D \; \mathbb{E}_{\mathbf{x},\mathbf{y}}[\; \log D(\mathbf{x}, G(\mathbf{x})) + \log(1 - D(\mathbf{x}, \mathbf{y}))\;]$$

[Goodfellow et al., 2014]
[Isola et al., 2017]



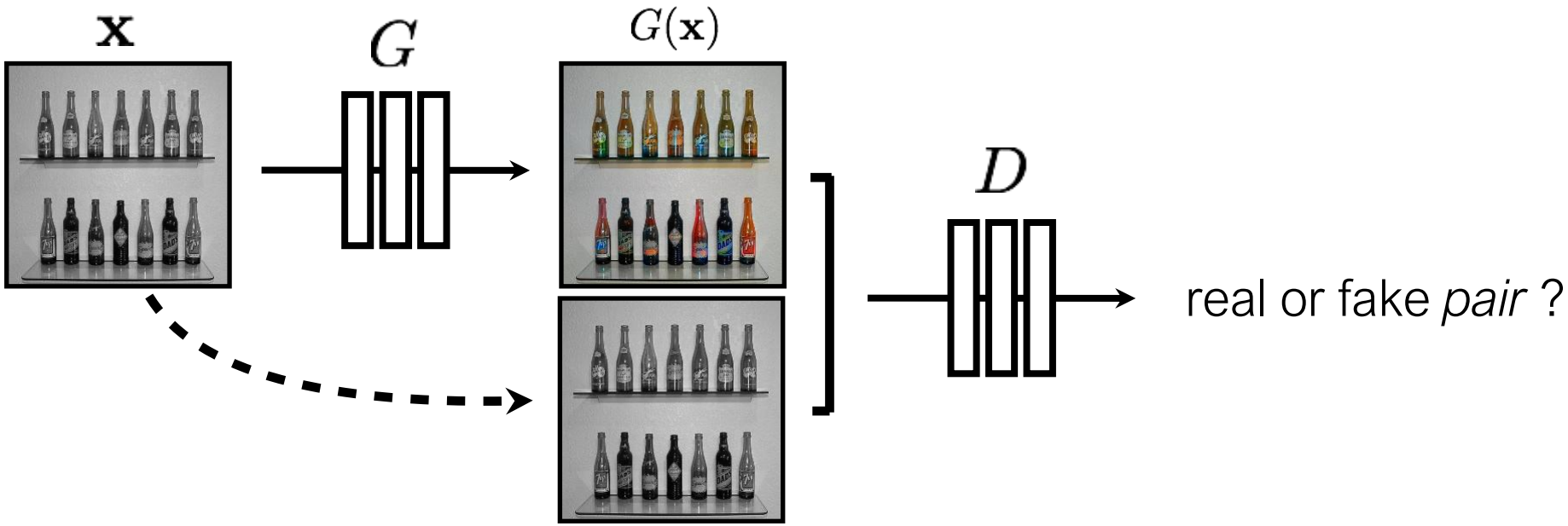
$$\arg\min_G \max_D \; \mathbb{E}_{\mathbf{x},\mathbf{y}}[\; \log D(\mathbf{x}, G(\mathbf{x})) + \log(1 - D(\mathbf{x}, \mathbf{y}))\;]$$

[Goodfellow et al., 2014]
[Isola et al., 2017]



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

[Goodfellow et al., 2014]
[Isola et al., 2017]



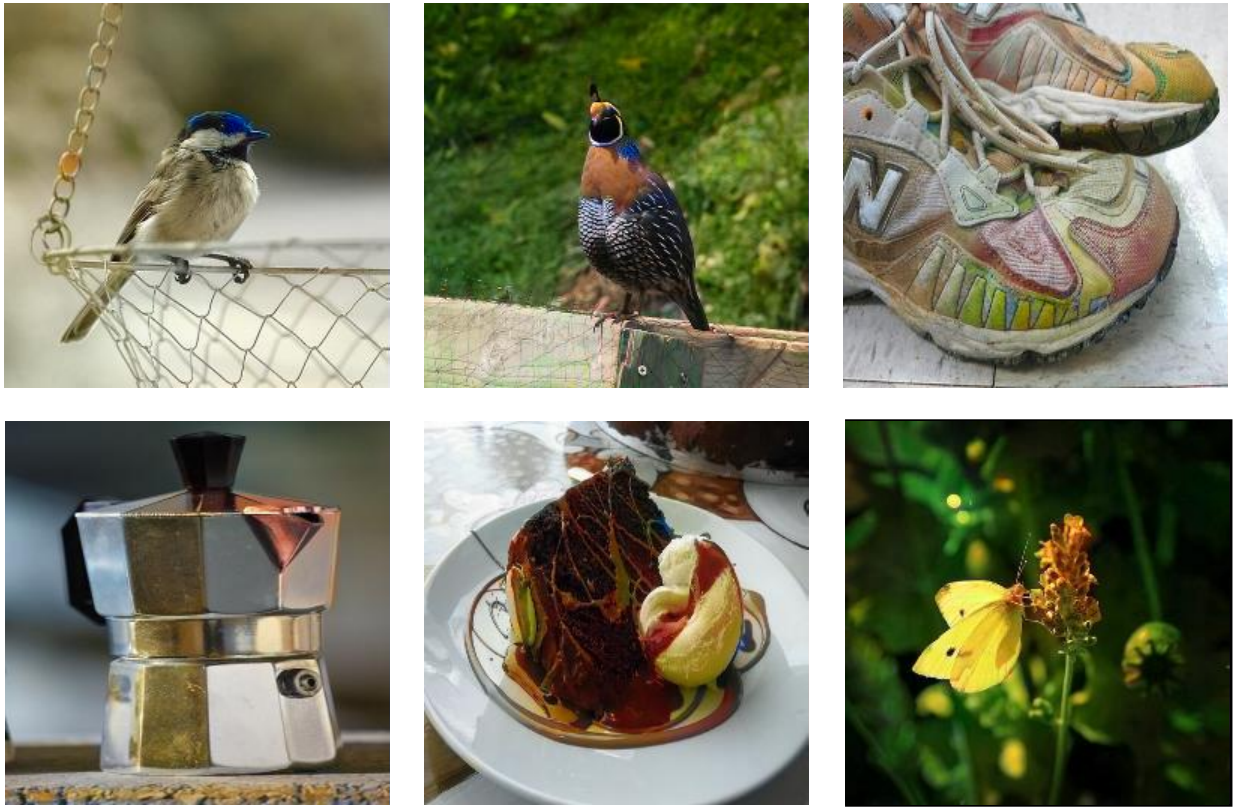
$$\arg\min_{G}\max_{D}\ \mathbb{E}_{\mathbf{x},\mathbf{y}}[\ \log D(\mathbf{x}, G(\mathbf{x})) + \log(1 - D(\mathbf{x}, \mathbf{y}))\ ]$$

[Goodfellow et al., 2014]
[Isola et al., 2017]

BW → Color

Data from [Russakovsky et al. 2015]

BW → Color



Data from [Russakovsky et al. 2015]

Input | Output | Groundtruth

Data from
[maps.google.com]

Input

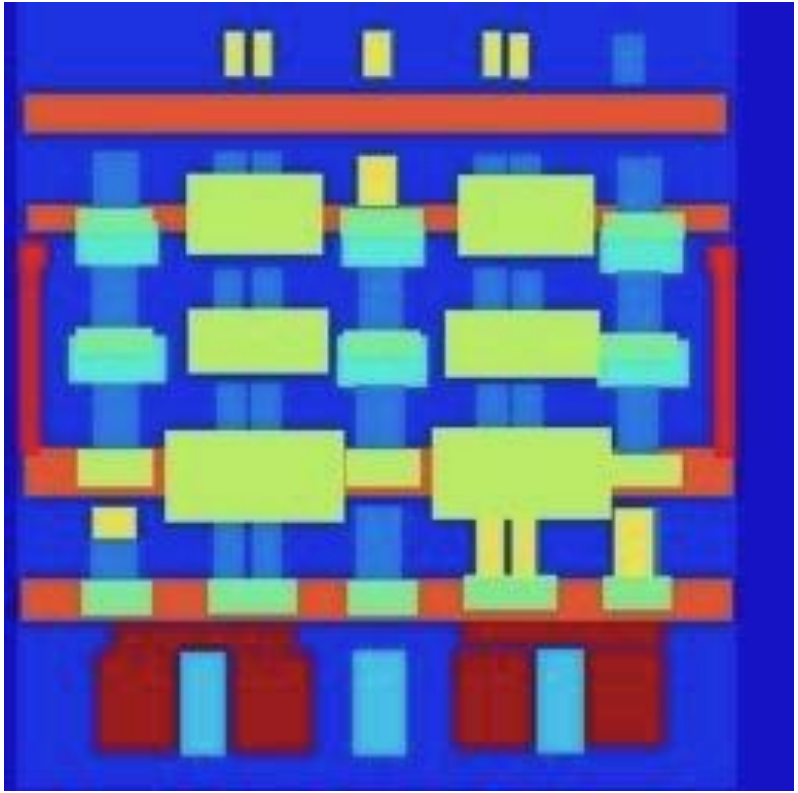
Output

Groundtruth

Data from [maps.google.

Labels → Facades

Input

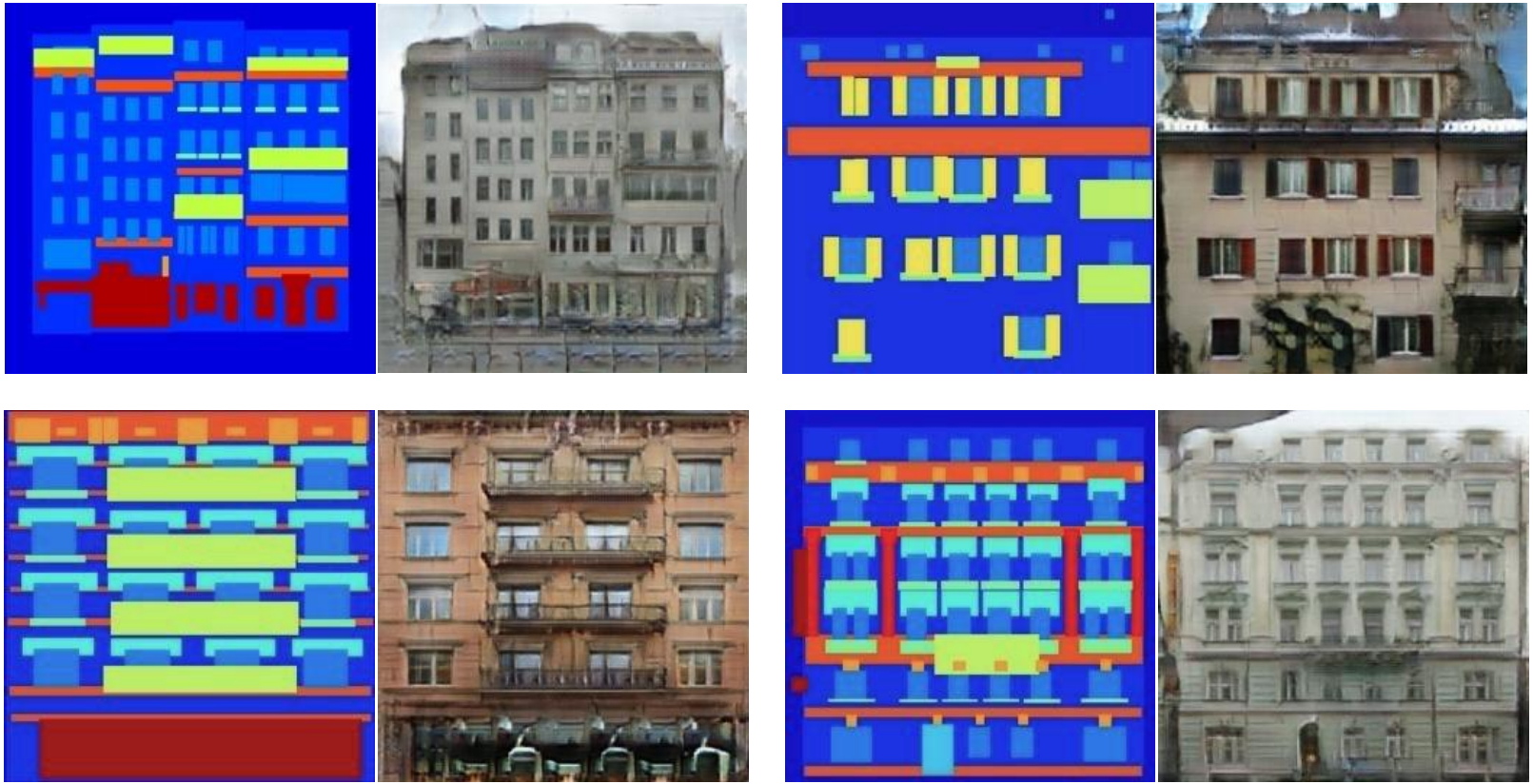


Output

Data from [Tylecek, 2013]

Labels → Facades

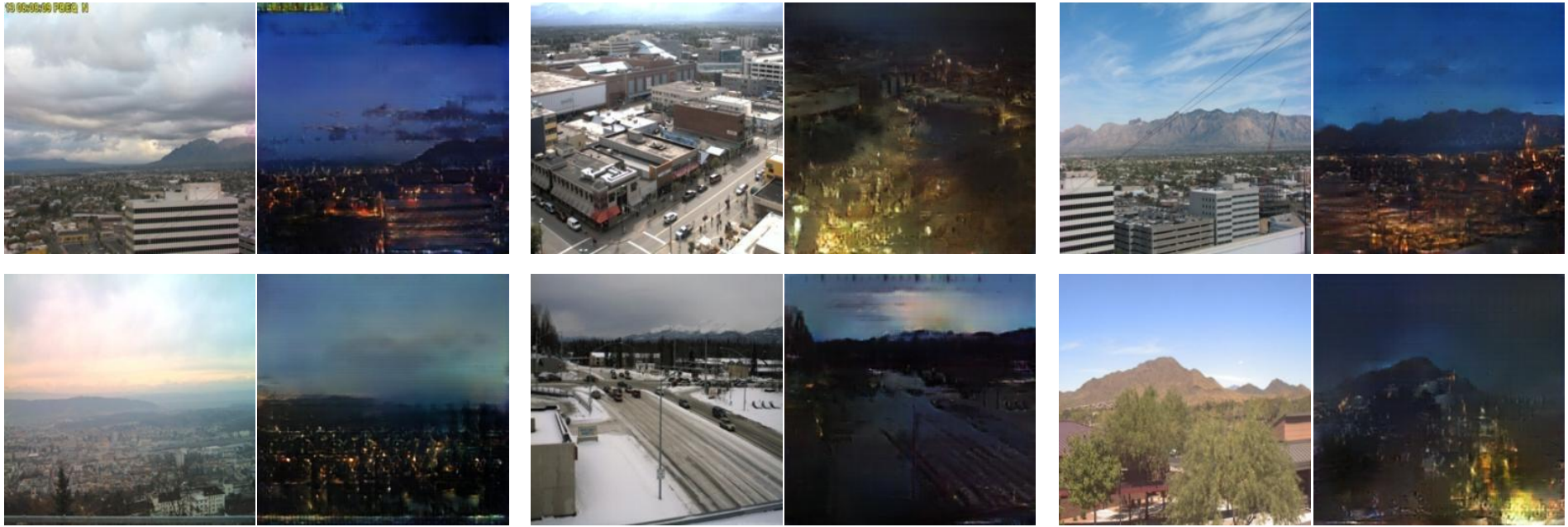
Input Output Input Output



Data from [Tylecek, 2013]

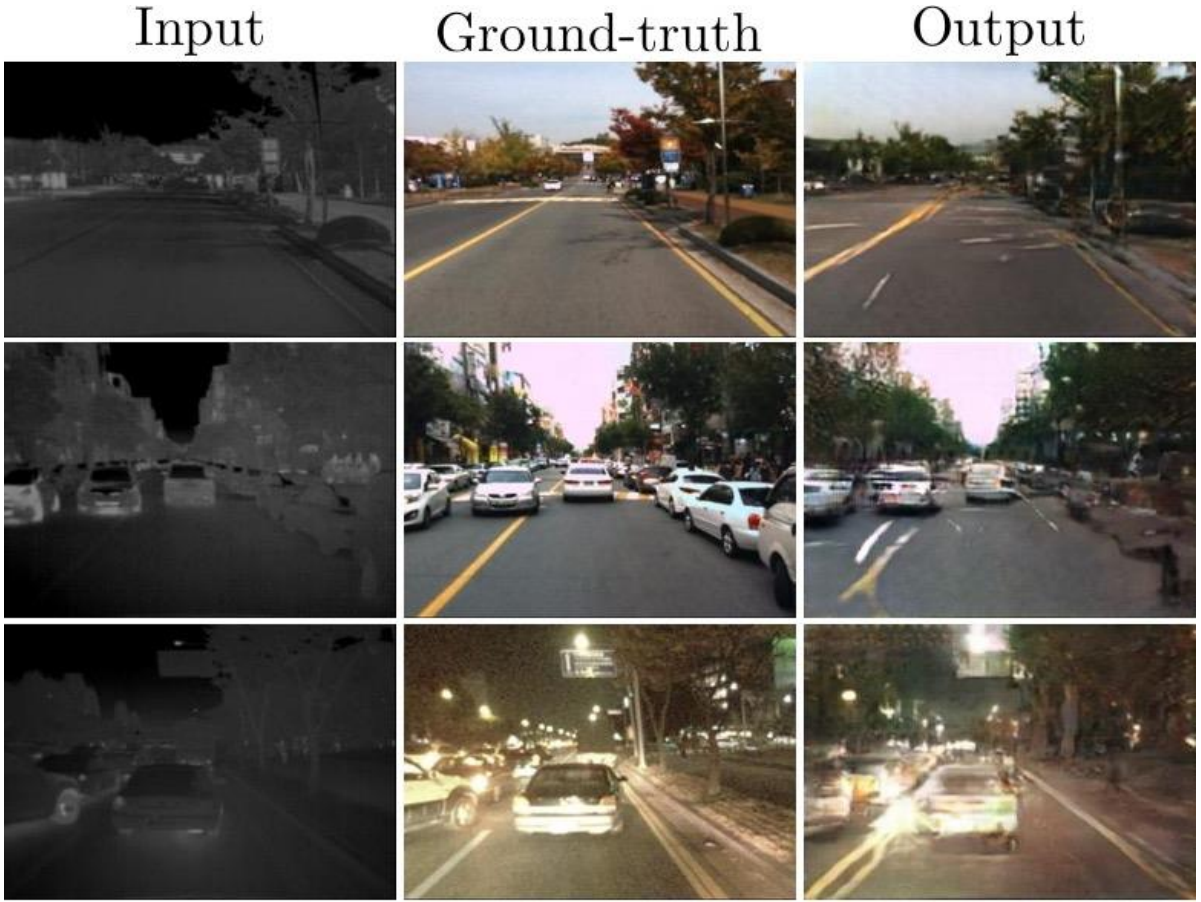
Day → Night

Input Output Input Output Input Output



Data from [Laffont et al., 2014]

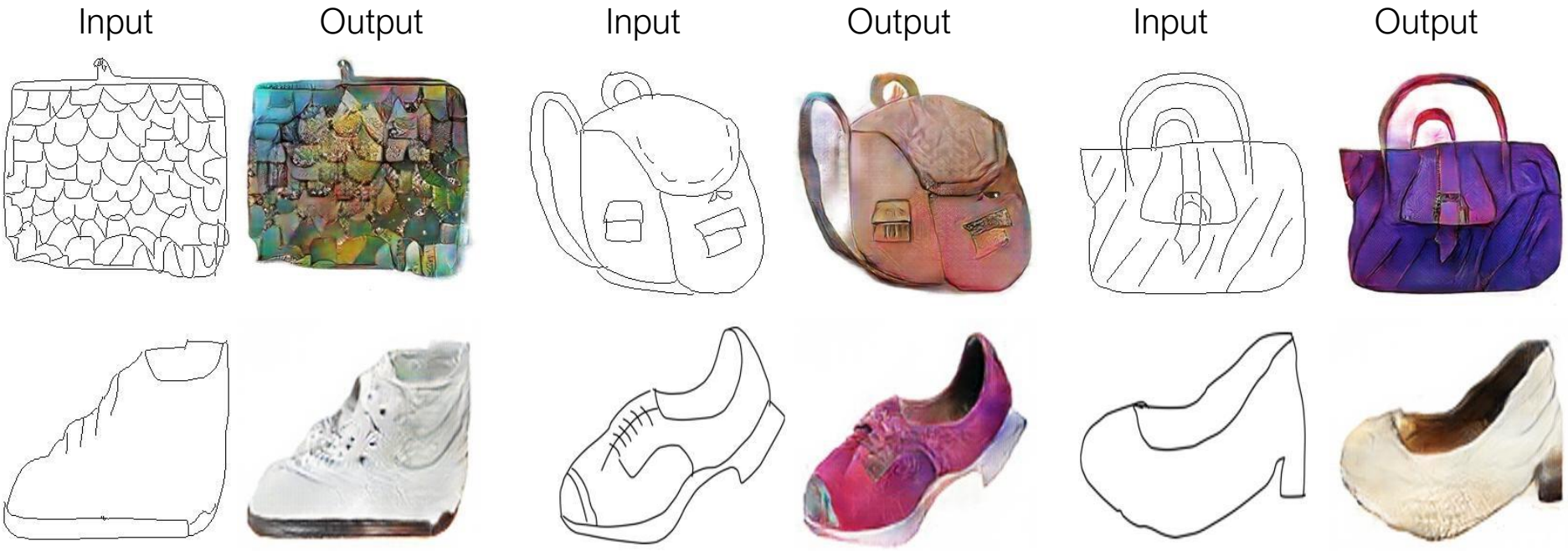
Thermal → RGB



Edges → Images

Edges from [Xie & Tu, 2015]

Sketches → Images



Trained on Edges → Images

Data from [Eitz, Hays, Alexa, 2012]

GitHub - junyanz/pytorch

GitHub, Inc. [US] | https://github.com/junyanz/pytorch-CycleGAN-and...

Apps | Research | Paris Stuff | Кино онлайн беспл | teaching | Bay Area Stuff | Other bookm

Features Business Explore Marketplace Pricing Search / Sign in or Sign up

junyanz / pytorch-CycleGAN-and-pix2pix

Watch 176 | Star 4,351 | Fork 982

Code | Issues 28 | Pull requests 3 | Projects 0 | Insights

Image-to-image translation in PyTorch (e.g., horse2zebra, edges2cats, and more)

pytorch gan cyclegan pix2pix deep-learning computer-vision computer-graphics image-manipulation image-generation generative-adversarial-network gans

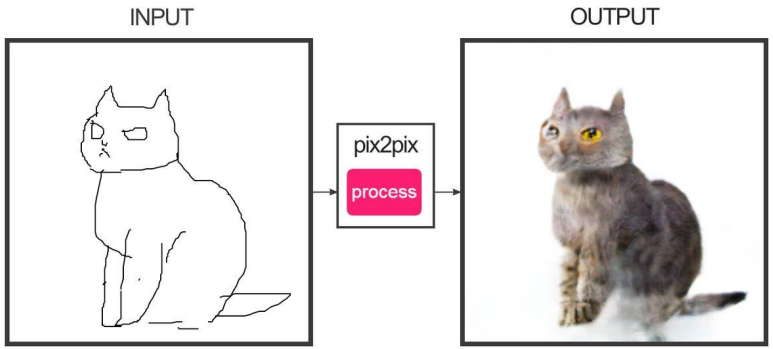
223 commits | 3 branches | 0 releases | 26 contributors

Branch: master | New pull request | Find file | Clone or download

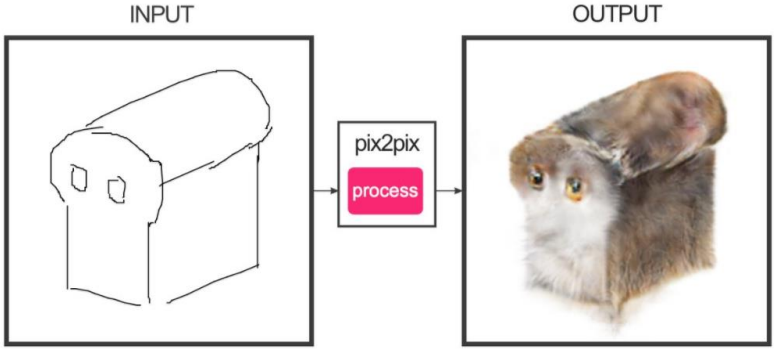
taesung89 Update README.md — Latest commit 6d9e173 10 Jun 13, 2018, 12:04 AM PDT

data	1. datasets are now configured automatically based on dataset_mode op...	10 days ago
datasets	Multiple changes regarding option management. See below.	15 days ago
imgs	add edges2cats demo	a year ago
models	TestModel now supports model_suffix option that can change the name o...	10 days ago

#edges2cats [Christopher Hesse]



@gods_tail



Ivy Tasi @ivymyt

Vitaly Vidmirov @vvid

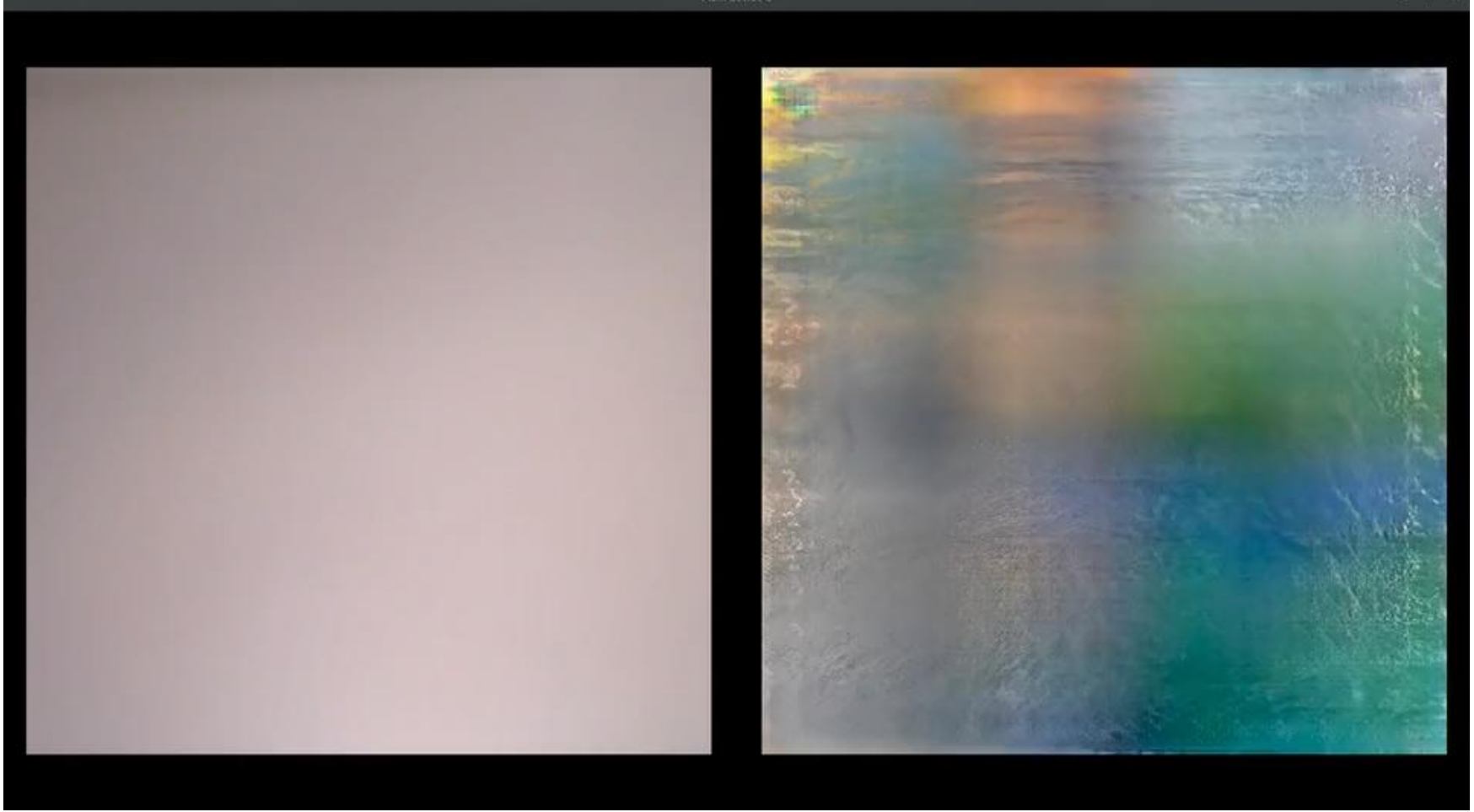
@ka92

Twitter-driven research: #pix2pix

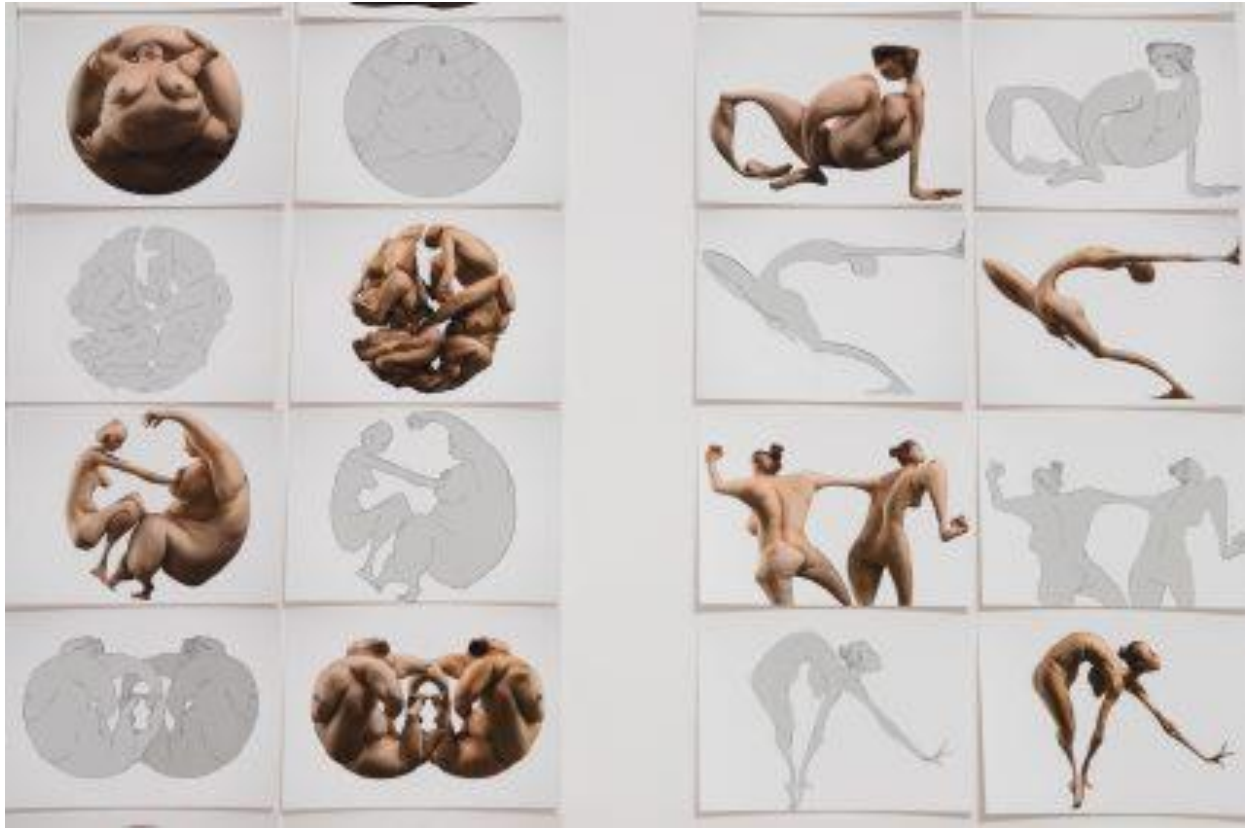
Brannon Dorsey @brannondorsey

Mario Klingemann @quasimondo

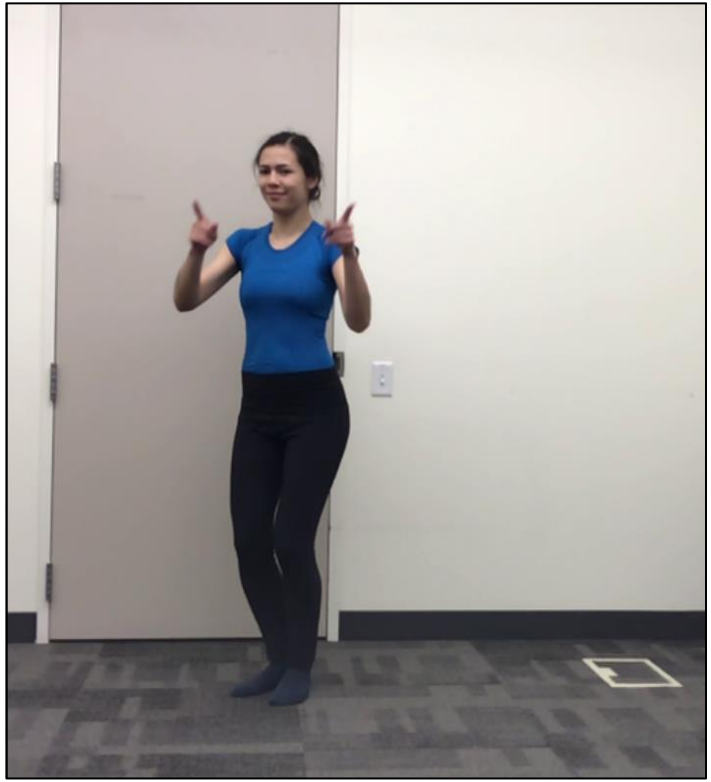
Bertrand Gondouin @bgondouin



Scott Eaton (http://www.scott-eaton.com/)

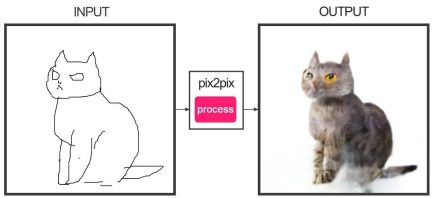


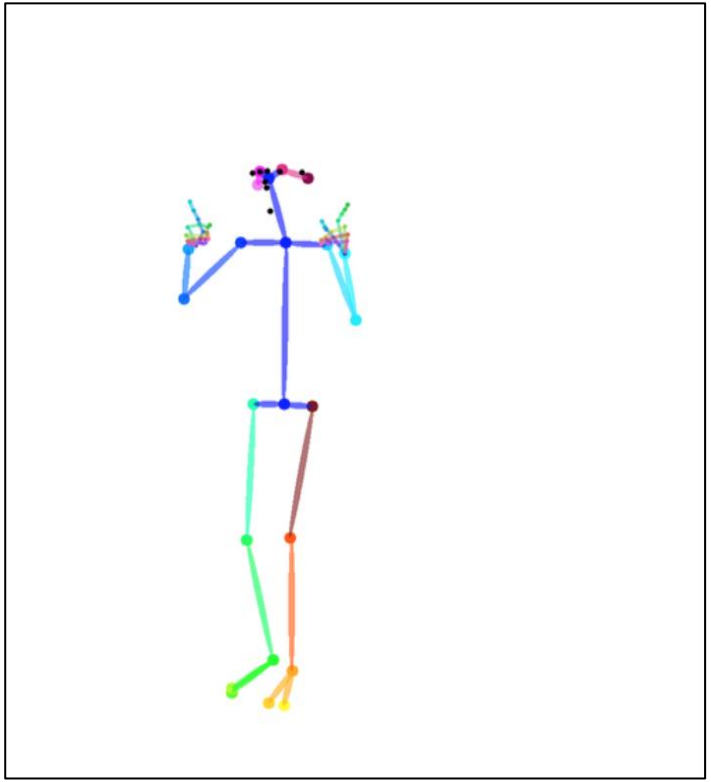
“Do as I Do”



OpenPose

pix2pix





Everybody Dance Now

Caroline Chan, Shiry Ginosar, Tinghui Zhou, Alexei A. Efros
UC Berkeley

Source Subject

Target Subject

Results

https://www.youtube.com/watch?v=PCBTZh41Ris&feature=youtu.be

Mac iPad iPhone Watch TV Music Support

App Store Preview

This app is available only on the App Store for iPhone and iPad.

Sway: Magic Dance 12+

Motion Effects & AI Filters

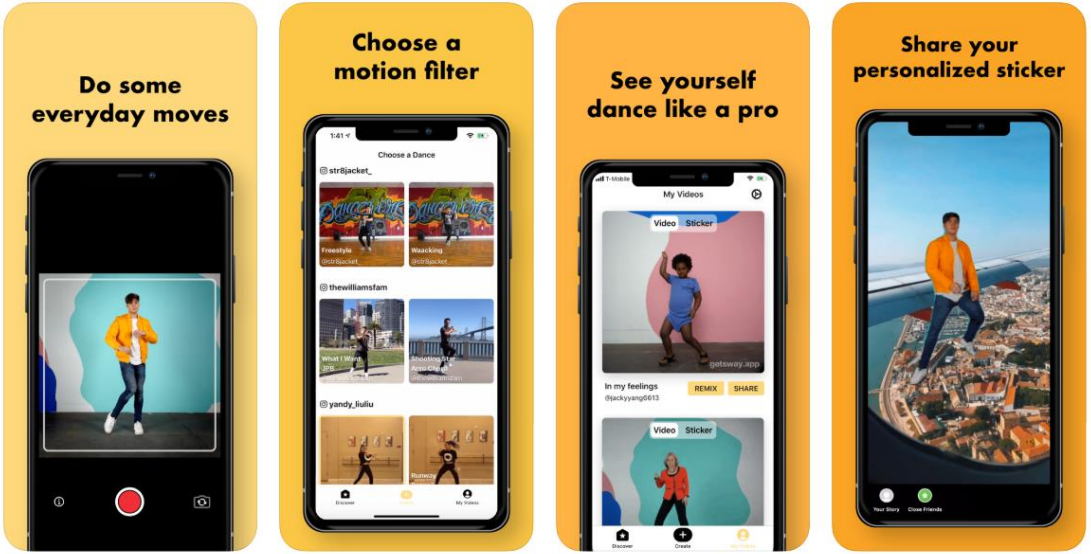
Humen, Inc.

#11 in Entertainment

★★★★☆ 3.9, 207 Ratings

Free

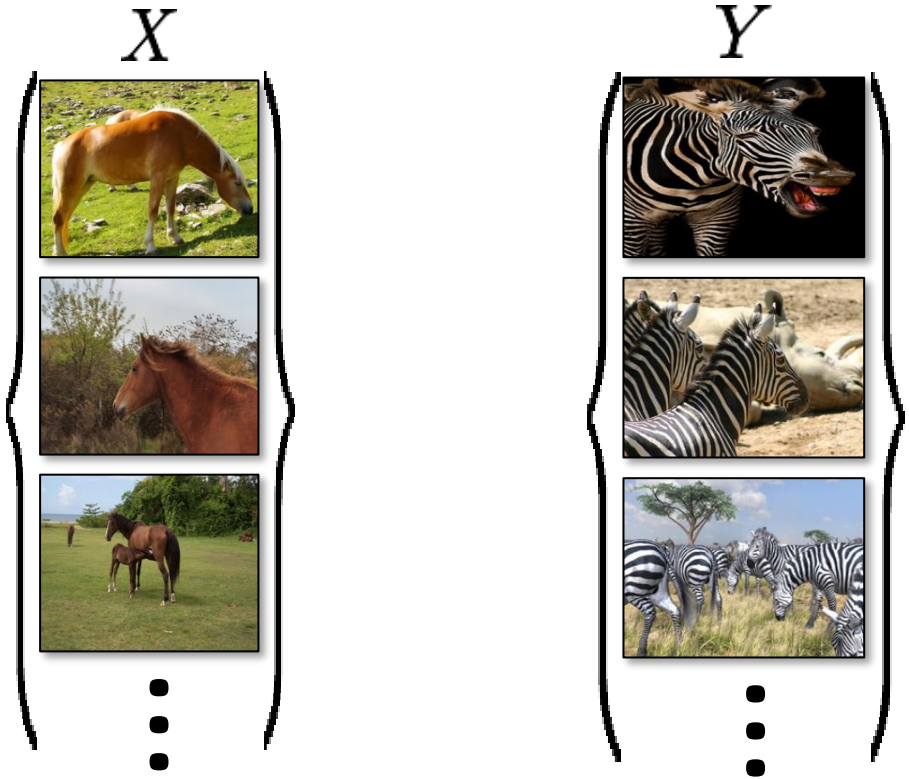
iPhone Screenshots



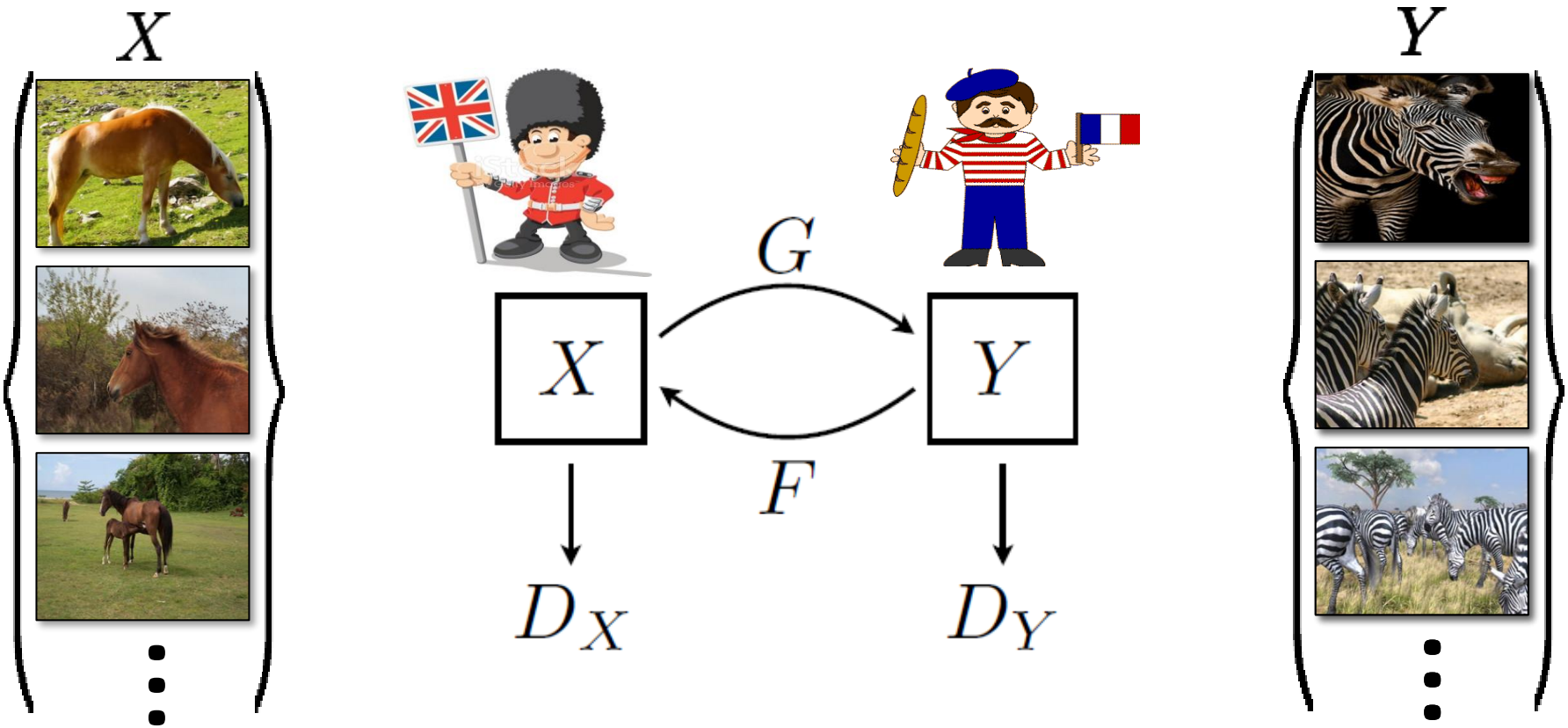
CEO: our own Dr. **Tinghui Zhou**

Paired training examples

Unpaired training examples

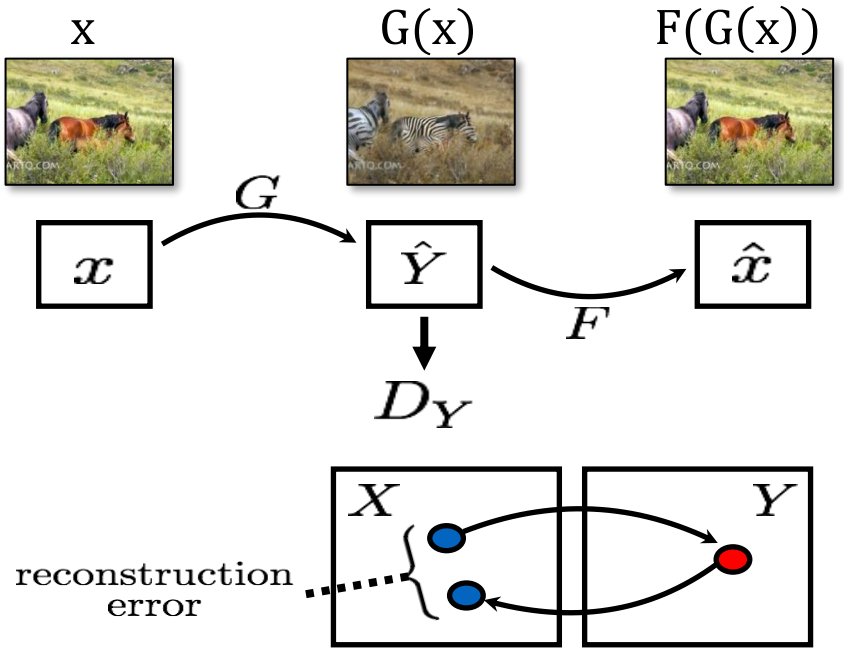


CycleGAN, or “there and back aGAN”



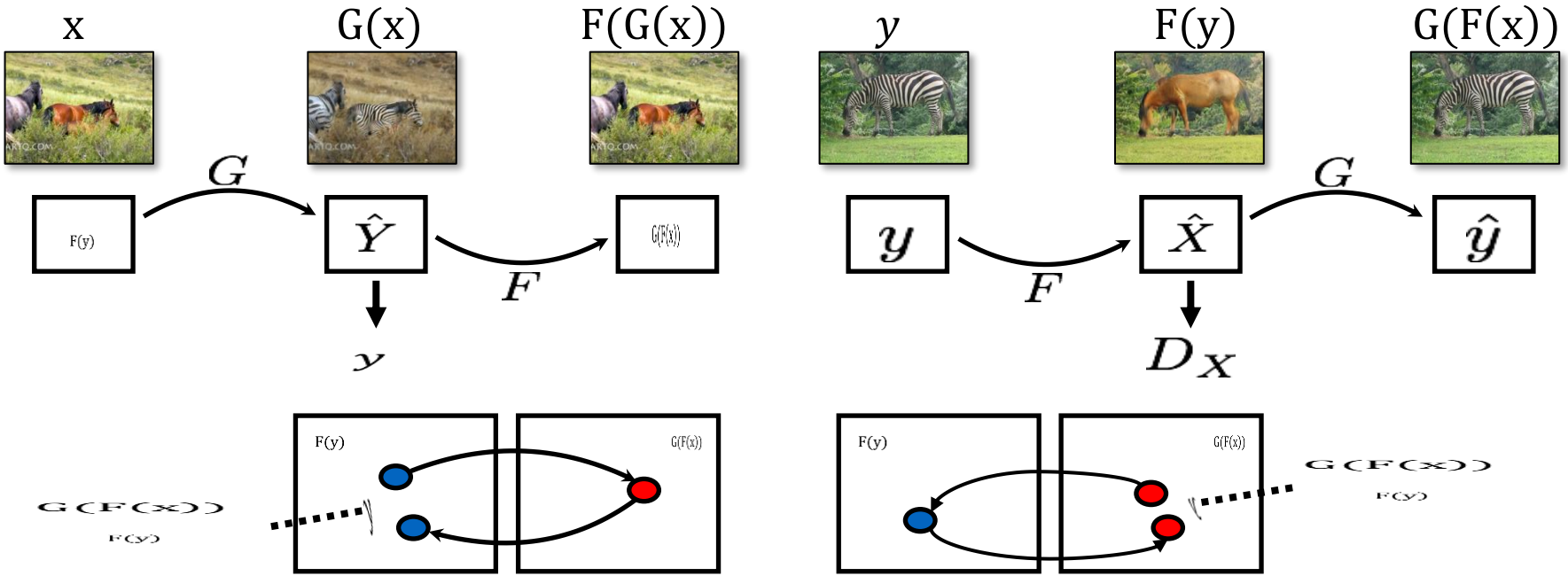
[Zhu*, Park*, Isola, Efros. ICCV 2017]

Cycle-Consistency Loss



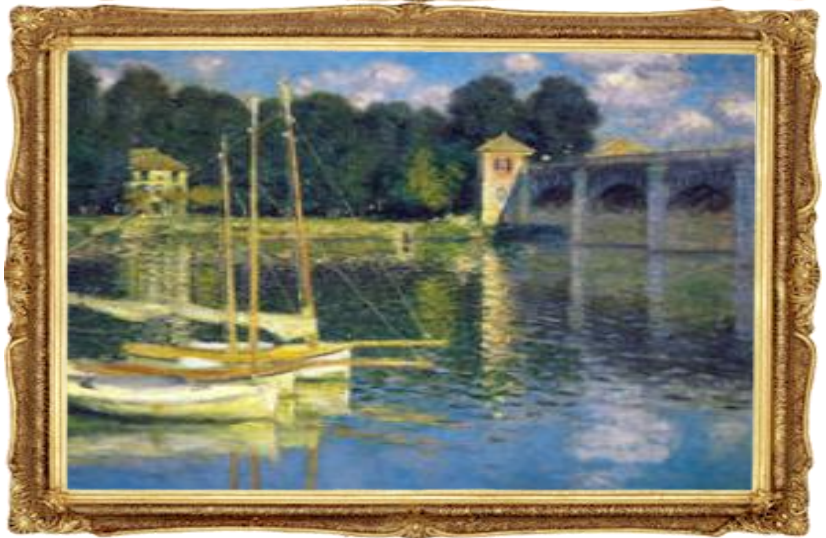
$$\left\| F\big(G(x)\big) - x \right\|_1$$

Cycle-Consistency Loss



$$\|F(G(x)) - x\|_1 \qquad \|G(F(y)) - y\|_1$$

Video



Collection Style Transfer

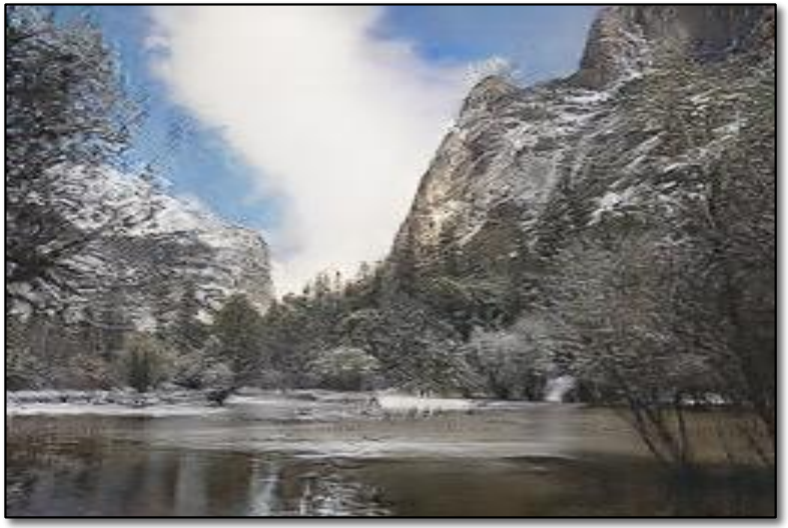
Photograph
© Alexei Efros

Monet

Van Gogh

Cezanne

Ukiyo-e

CG to Real

Grand Theft Auto

▶

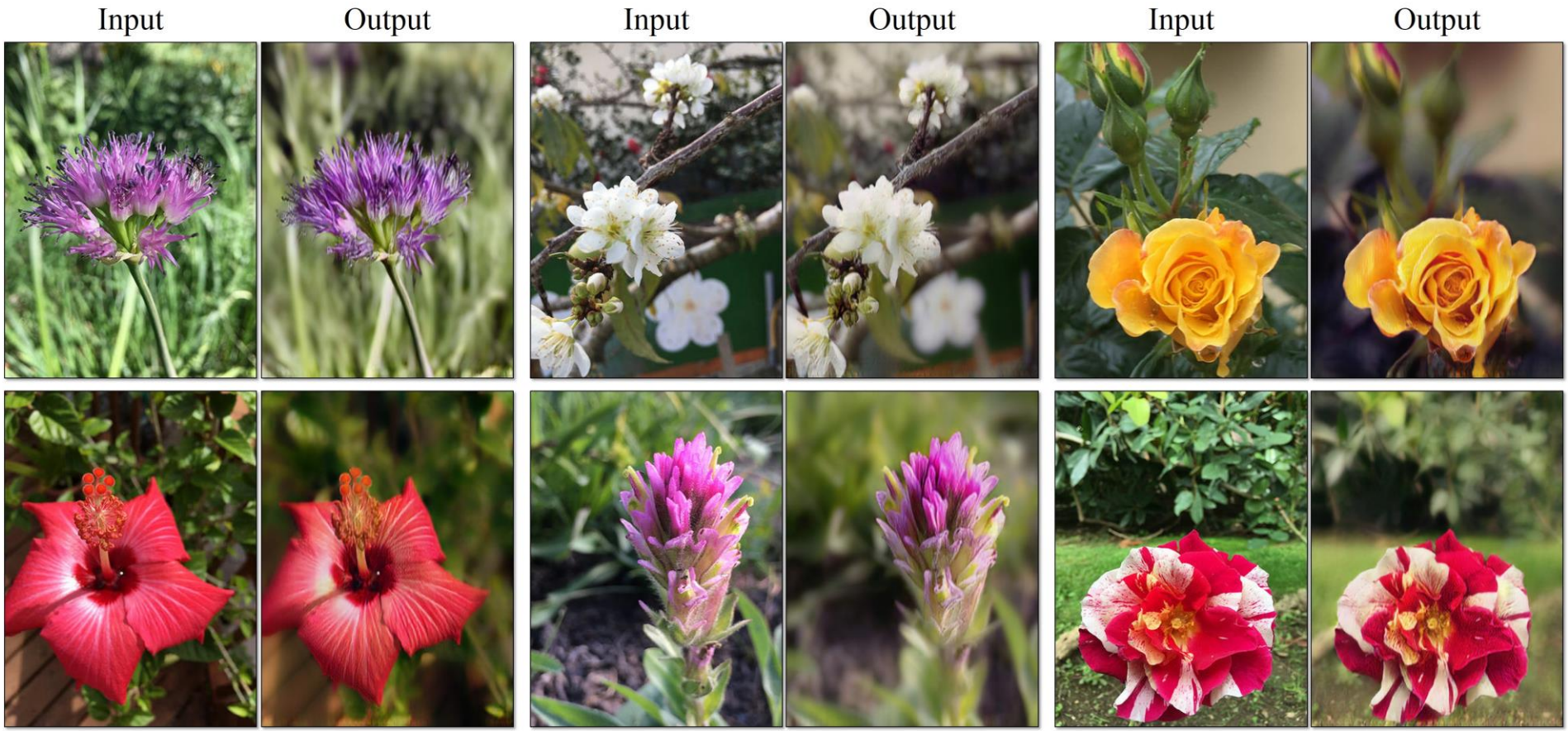
▶

Real to CG

▶

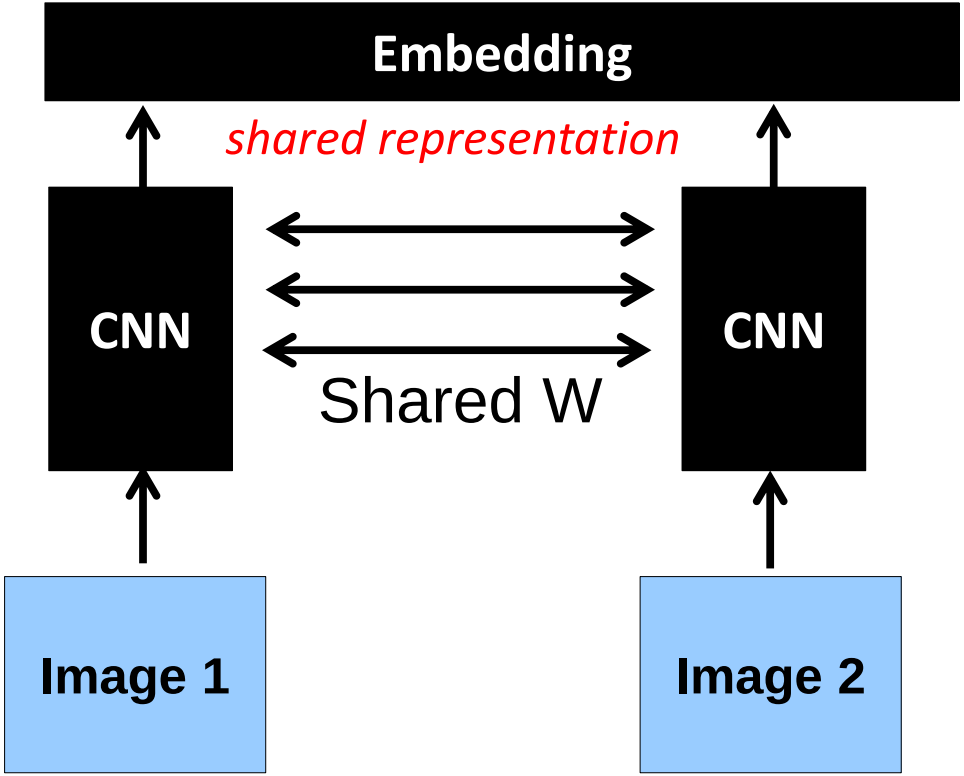
▶

Shallower depth of field

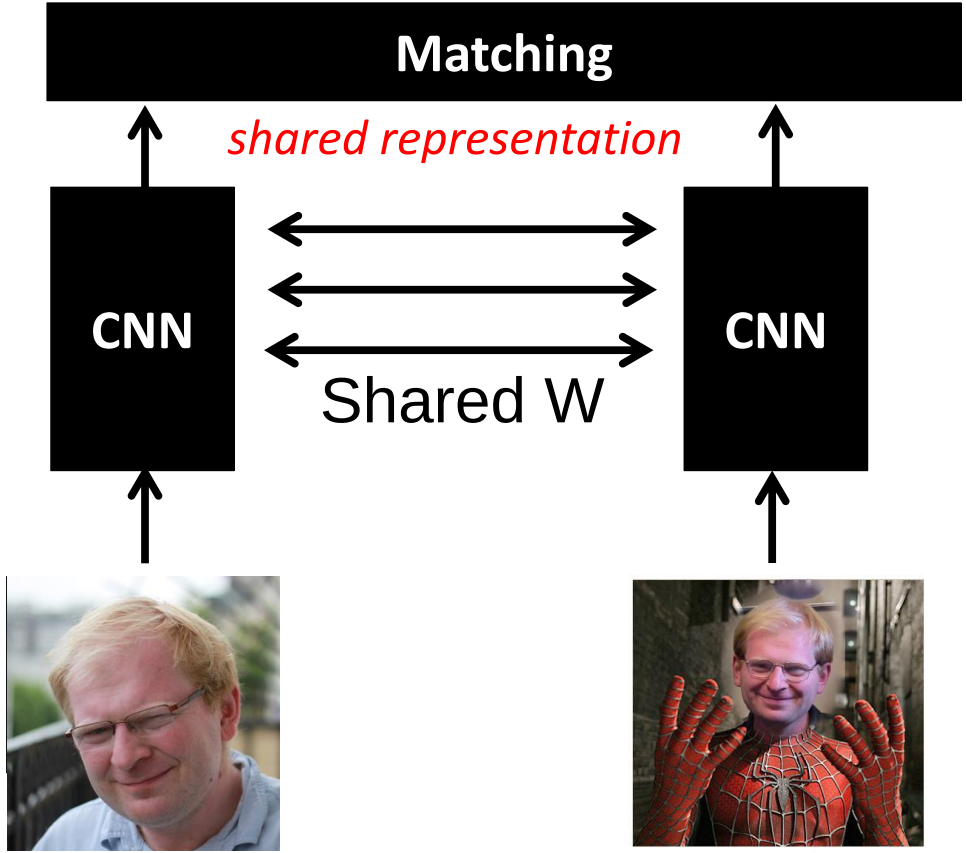


Failure case

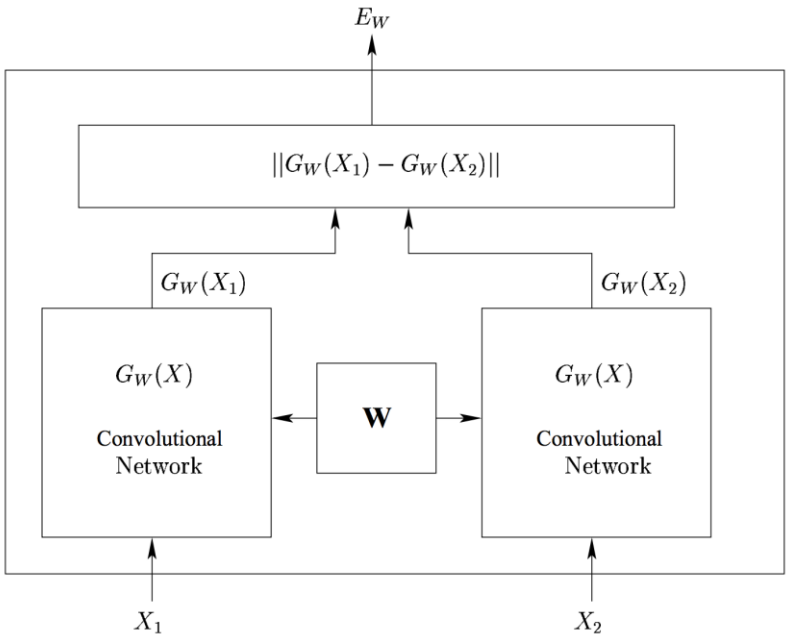
Learning an Embedding

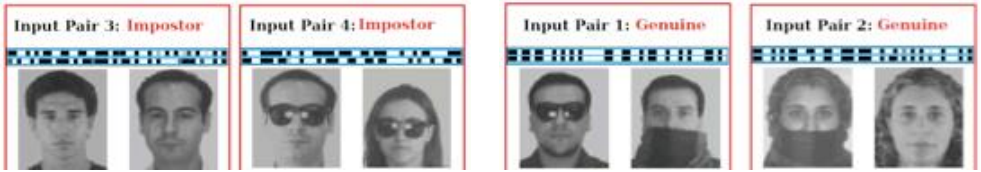


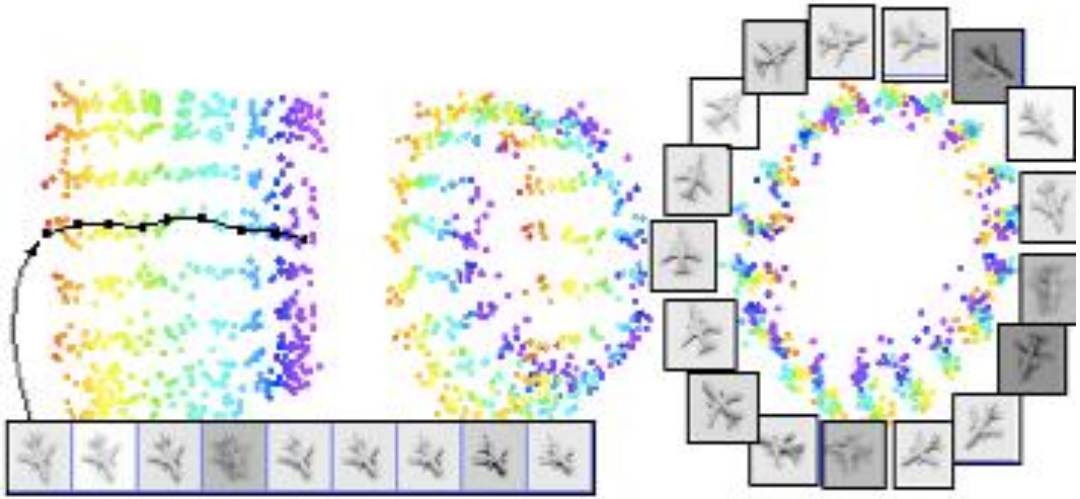
Learning an Embedding



Siamese Network w/ Contrastive Loss







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

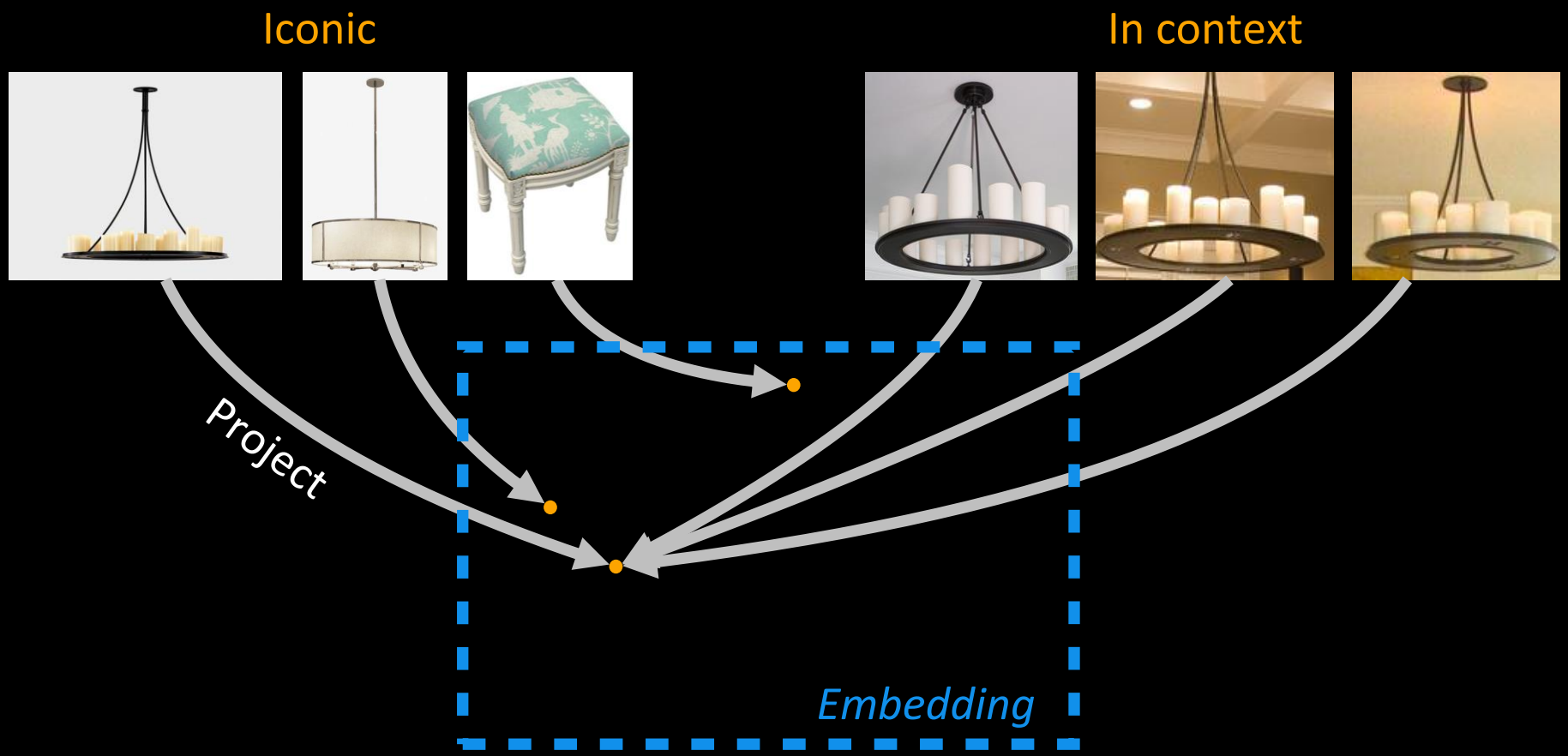
Iconic

(From a product website)

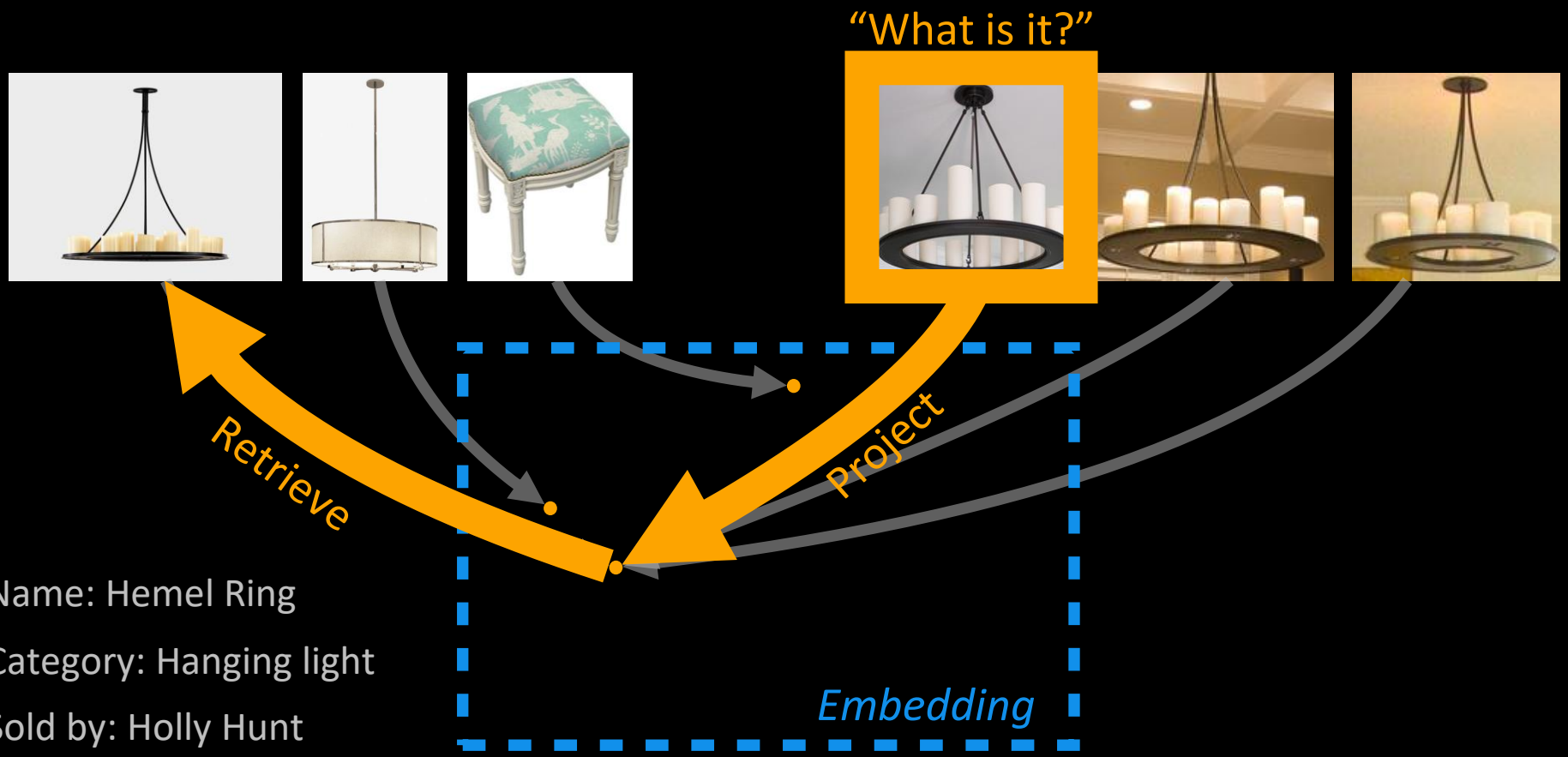
In context

(Cropped from a scene photo)

PROJECTING INTO A JOINT EMBEDDING

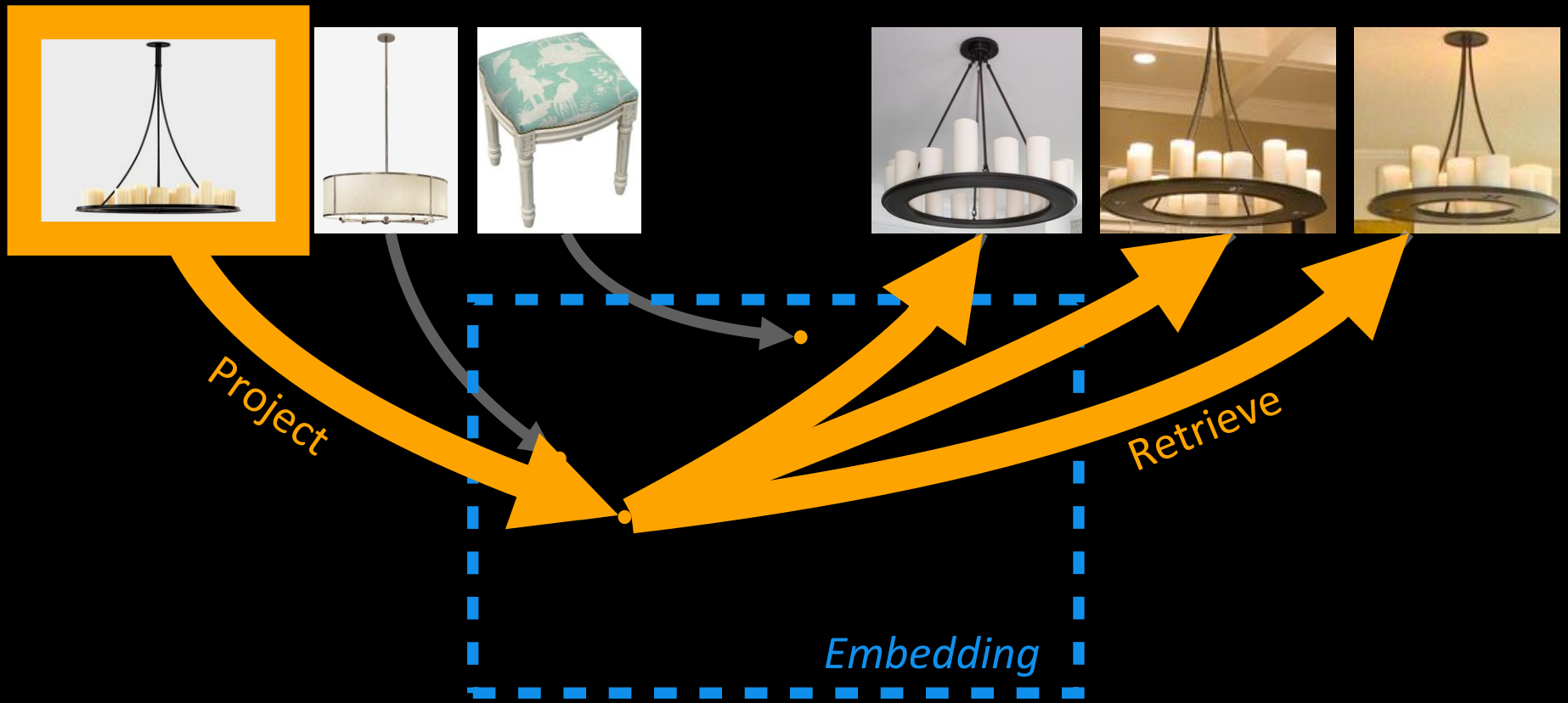


SEARCH USING THE EMBEDDING

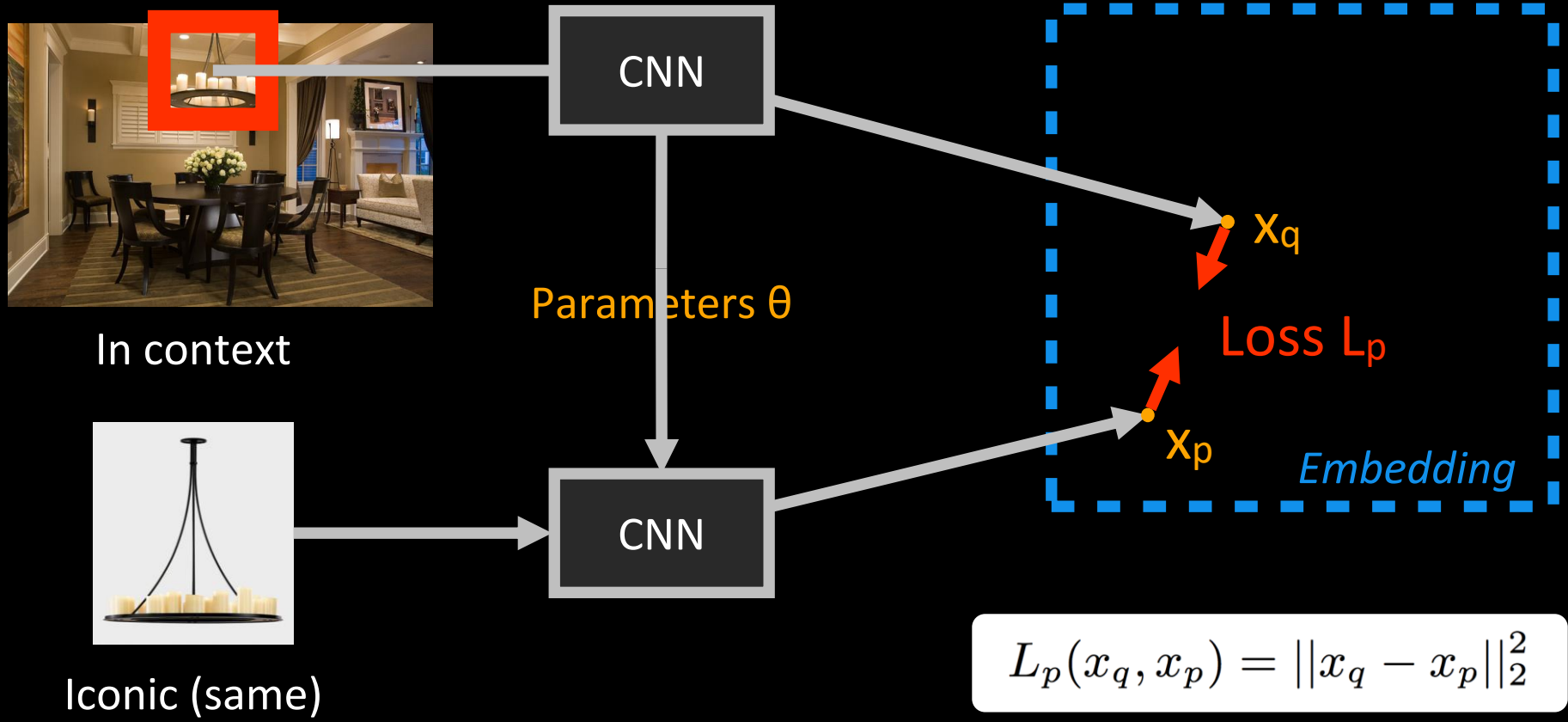


SEARCH USING THE EMBEDDING

"Where is it used?"

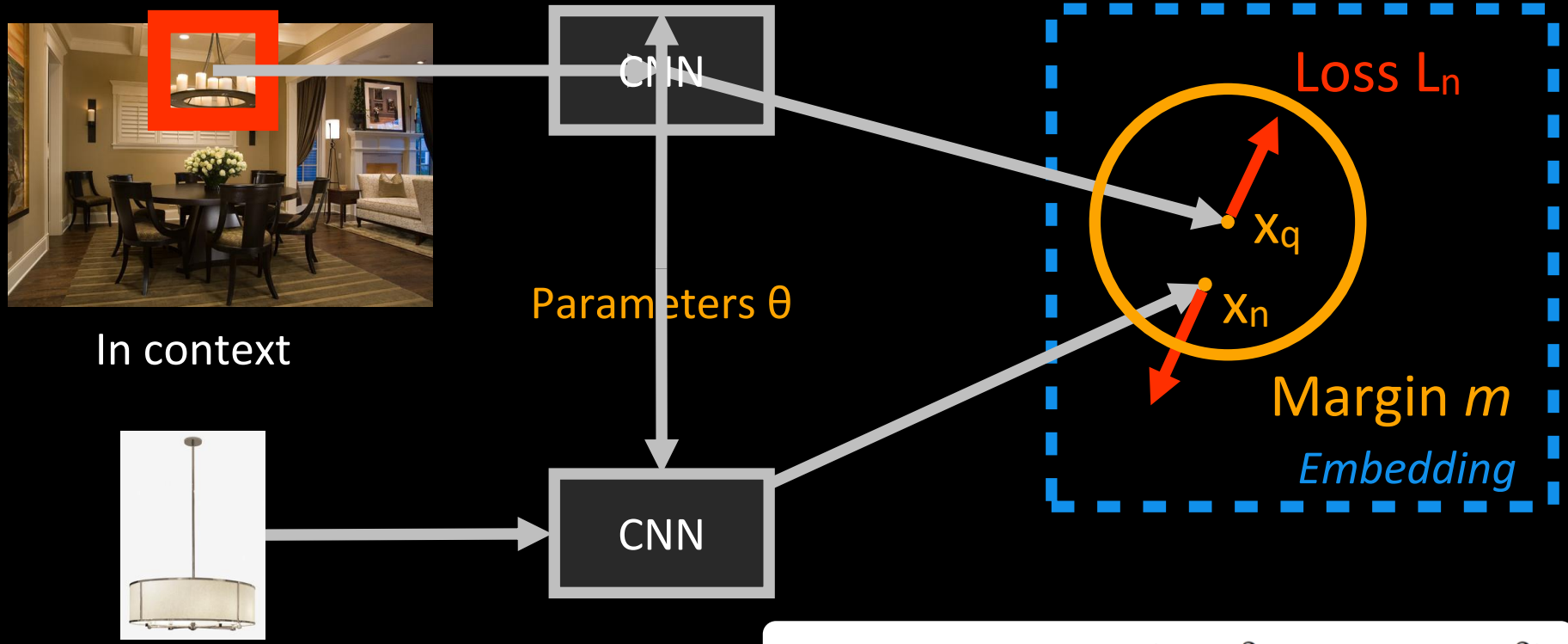


CONTRASTIVE LOSS: POSITIVE EXAMPLE



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

CONTRASTIVE LOSS: NEGATIVE EXAMPLE



In context

Iconic (different)

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

CONTRASTIVE LOSS: ALL TOGETHER

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

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

Margin

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

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

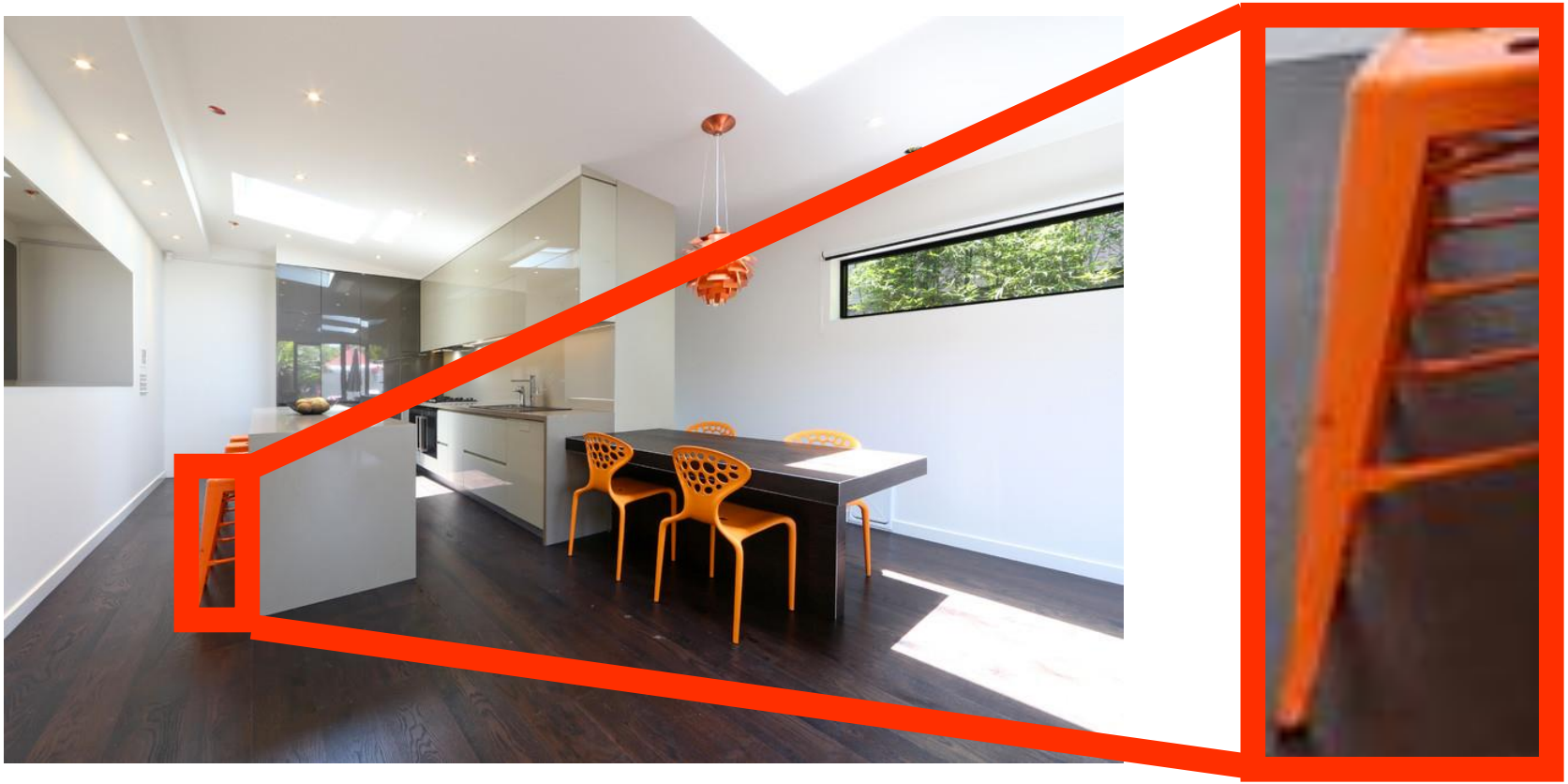
0.024

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

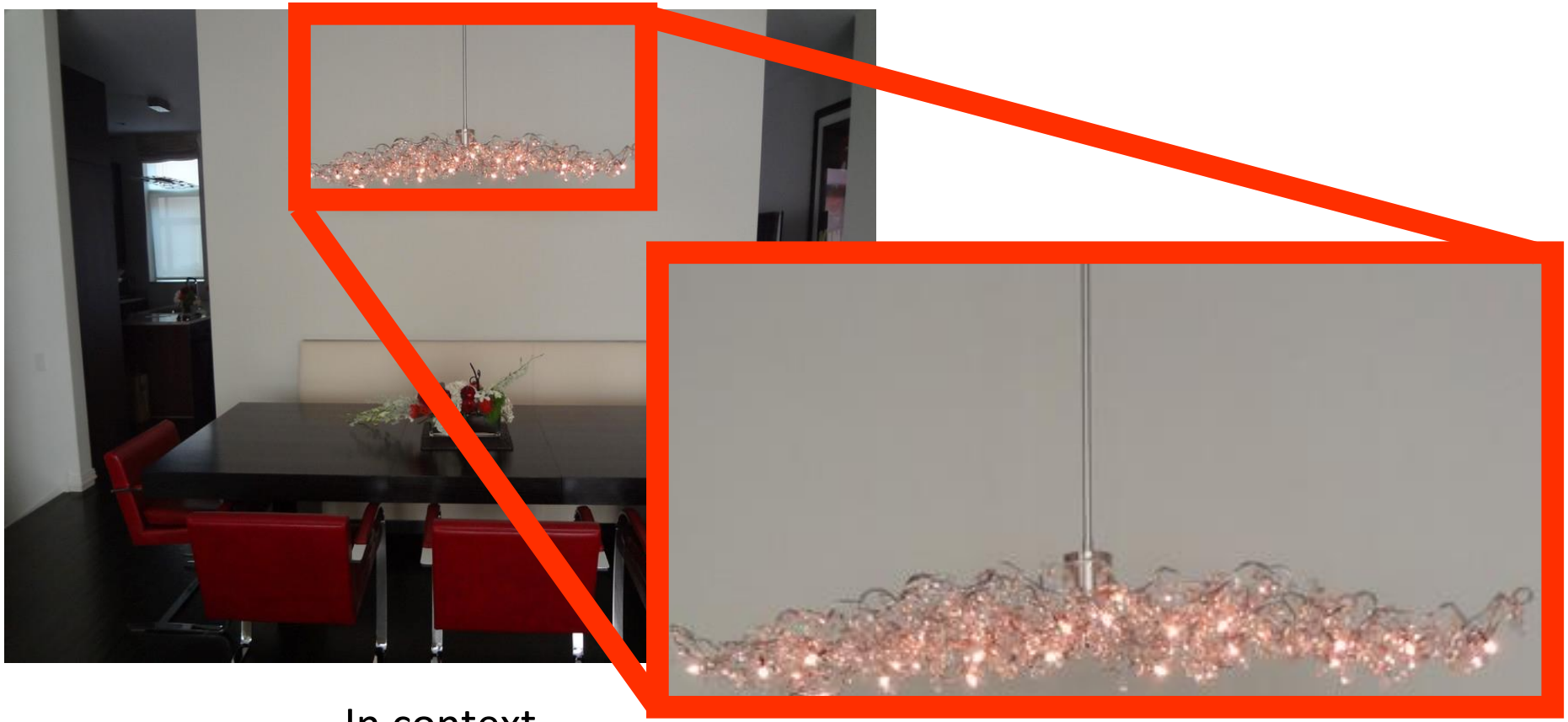
0.050

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?”

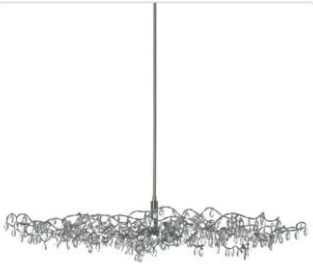


In context

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

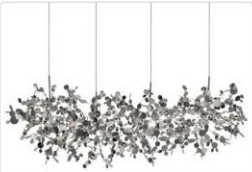
2 boxes

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

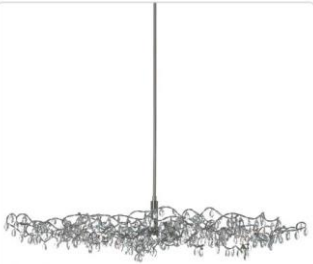
Pendant Lighting | Modern | View jpg | View detail

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

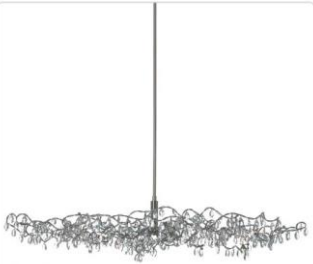
View detail

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

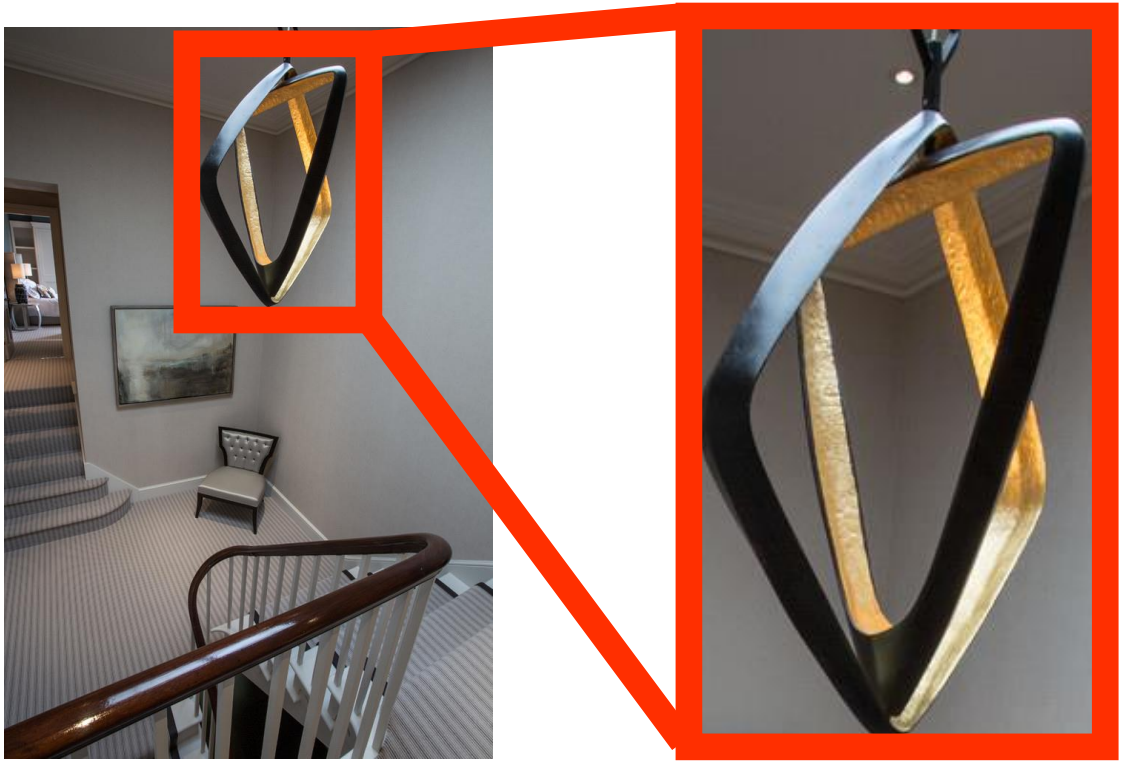
503.232

Mensa Hanging Light Fixture

Chandeliers | Contemporary | View jpg | View detail

514.867

RESULTS: FAILURE CASE

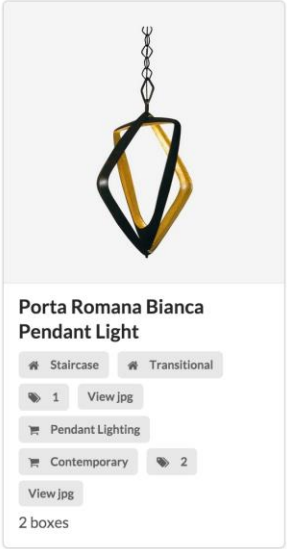


In context

RESULTS: FAILURE CASE

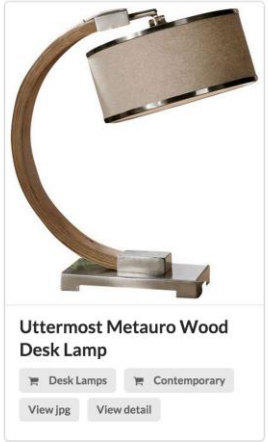
In context

Iconic

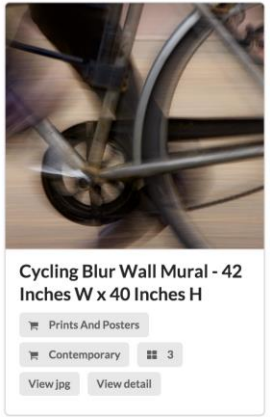


Top 4 results:

0.066



0.067



0.069

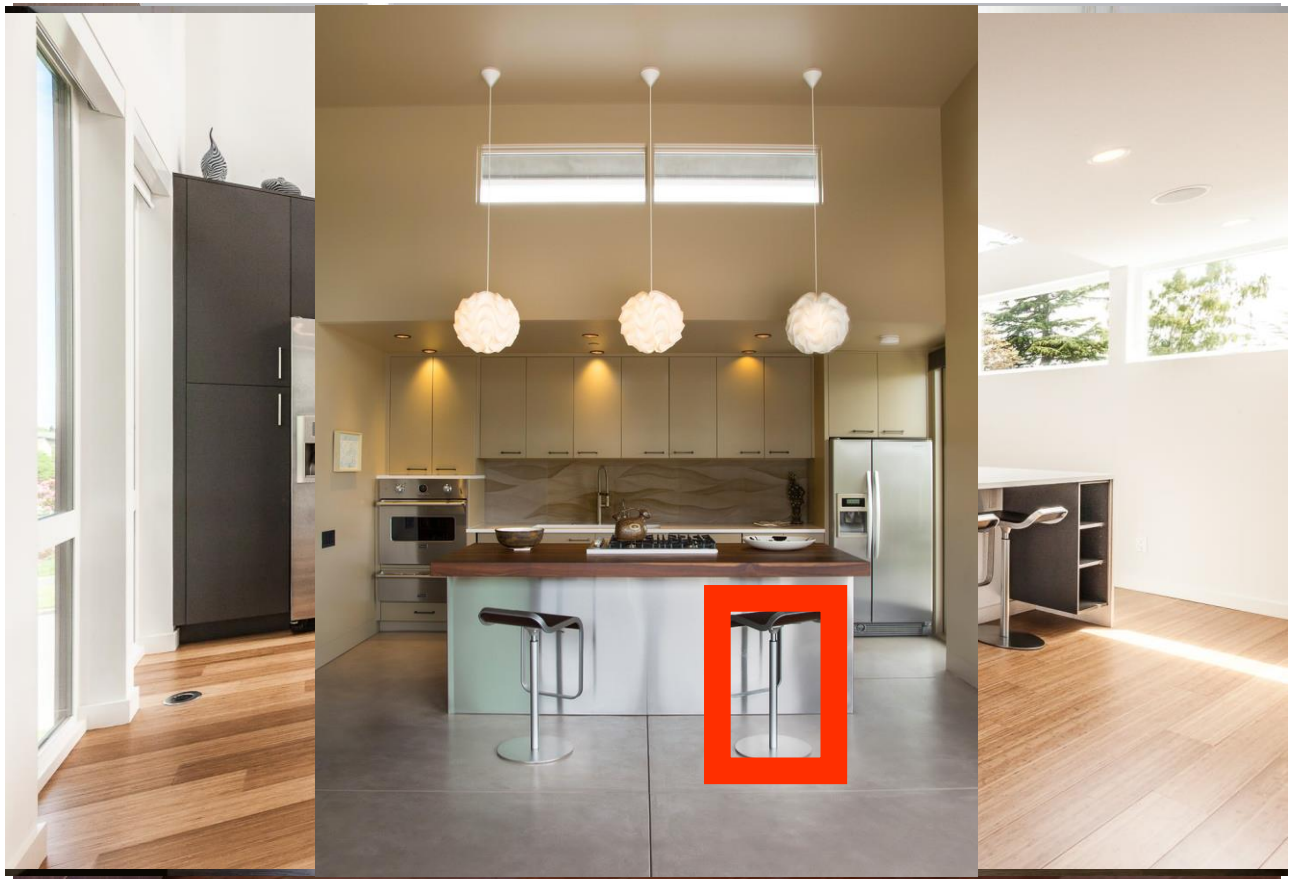
0.069

RESULTS: “WHERE IS IT USED?”

“Maskros Pendant Lamp”

RESULTS: “WHERE IS IT USED?”

"LEM Piston Stool | Design Within Reach”



SEARCHING ACROSS CATEGORIES