CS5670: Computer Vision

Noah Snavely

Synthesizing images with generative adversarial networks (GANs)

Automatically generated celebrities via GANs

Slides from Philipp Isola

Readings

- NIPS 2016 Tutorial on GANs
  - https://arxiv.org/abs/1701.00160

Announcements

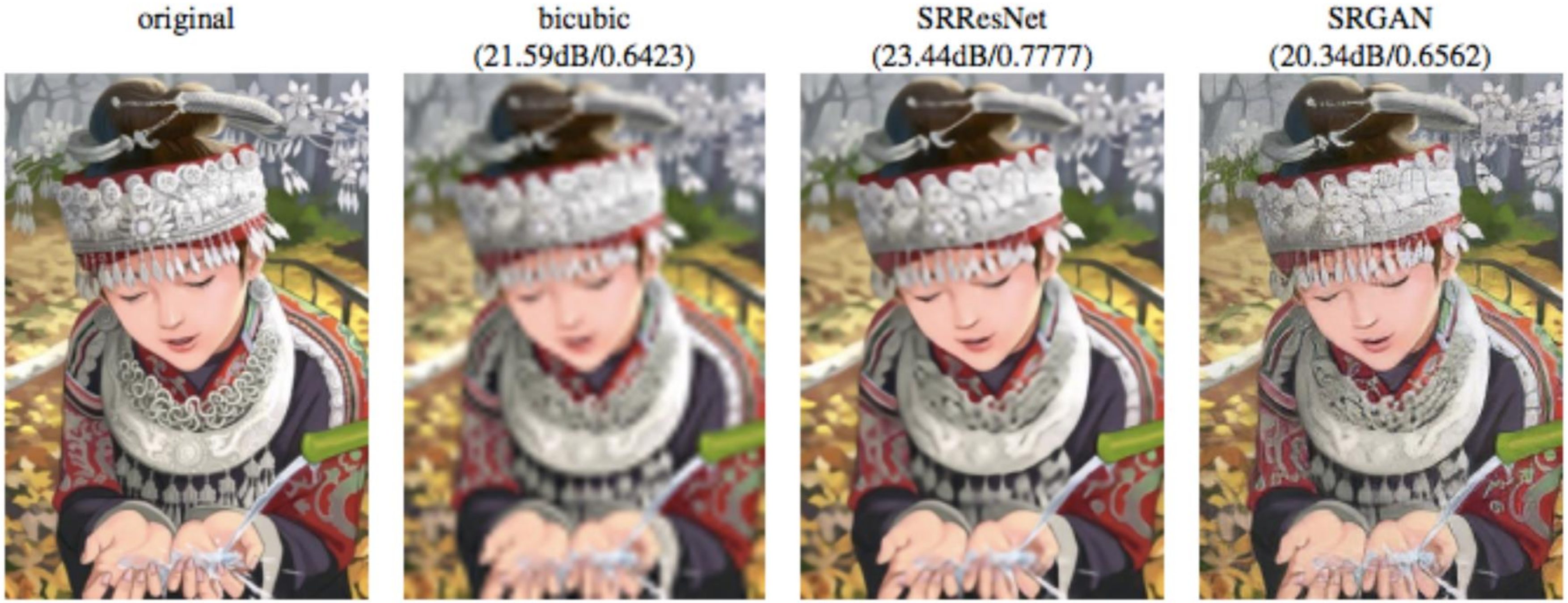
- Project 5 due Monday, 5/14 at 11:59pm
- Reminder: Course evals (5 bonus points)
  - https://apps.engineering.cornell.edu/CourseEval
- Final exam in class on Wednesday
  - Please arrange yourselves with at least one space between you and the closest person in the same row when you arrive

Today

- Generative adversarial networks
- Course review

Motivation: Synthesizing images

Single Image Super-Resolution

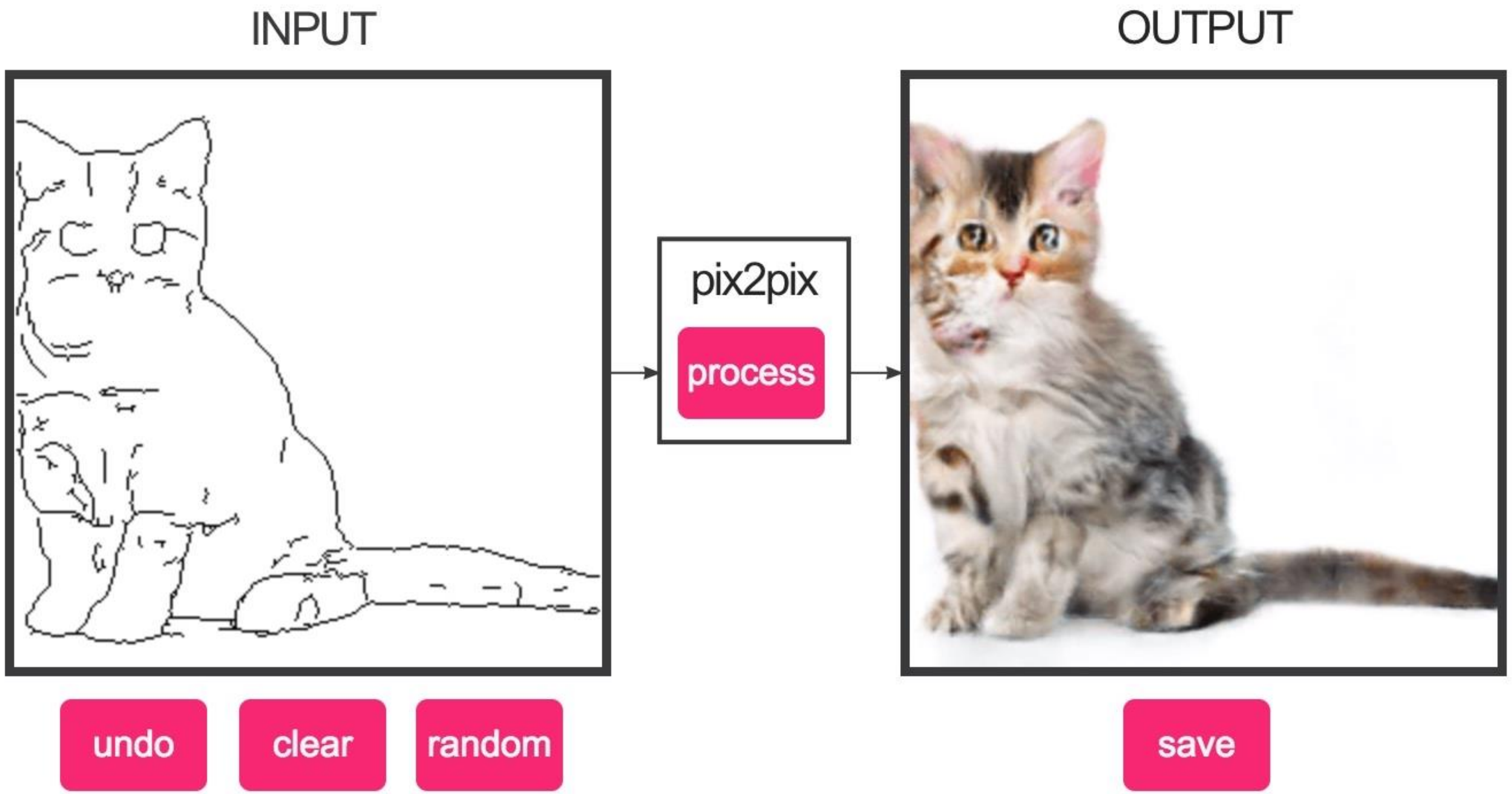


(Ledig et al 2016)

Motivation: Synthesizing images

Image to Image Translation

Demo



https://affinelayer.com/pixsrv/

Image classification

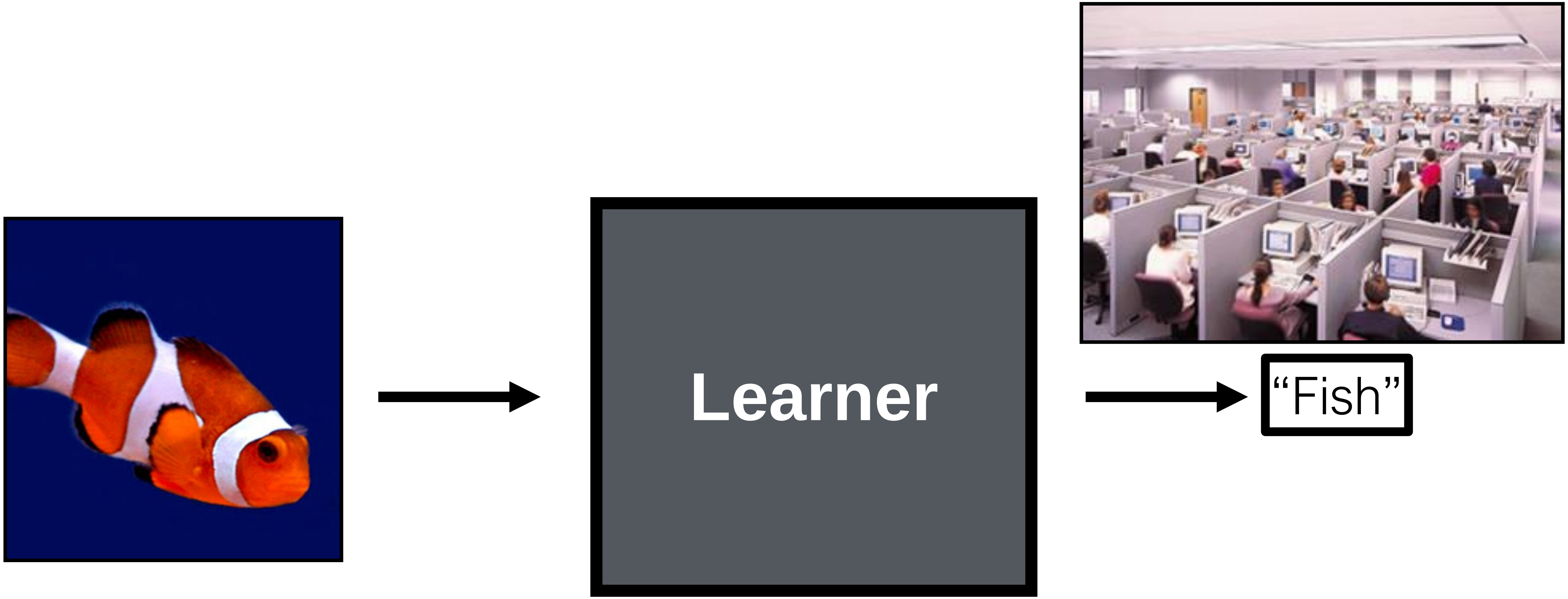


image X

label Y

Image classification

Learner

"Fish"

image X

label Y

Image classification

Learner

"Fish"

image X

label Y

Image classification

⋮

image X

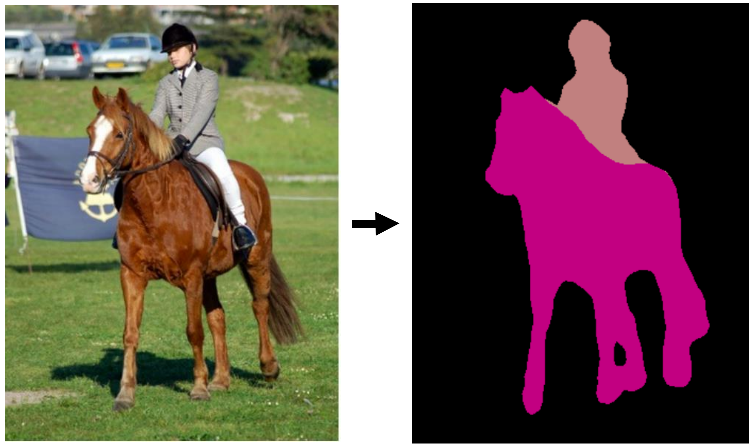
Learner

"Fish"

label Y

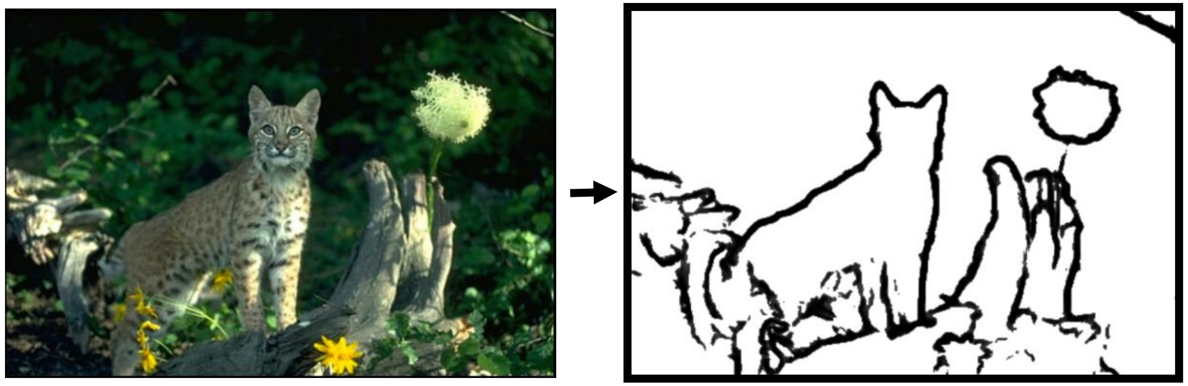
Image prediction ("structured prediction")

Object labeling



[Long et al. 2015, …]

Edge Detection



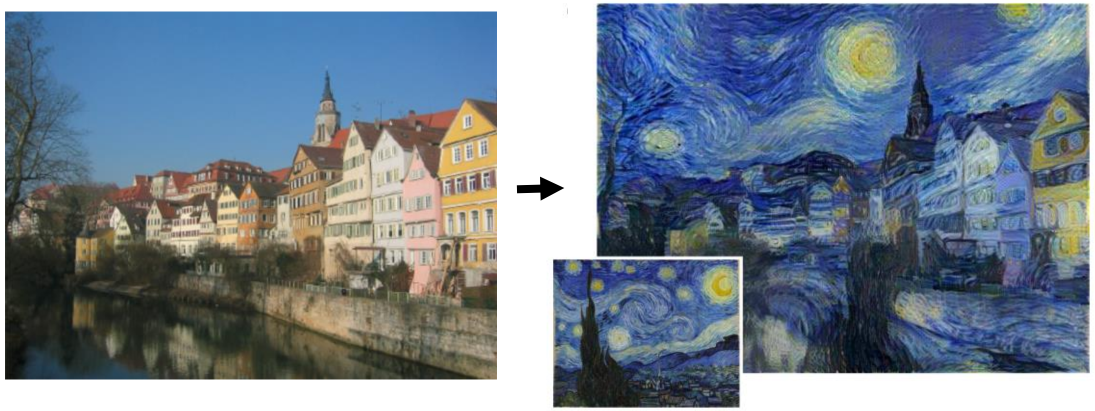
[Xie et al. 2015, …]

Text-to-photo

"this small bird has a pink breast and crown…" →

[Reed et al. 2014, …]

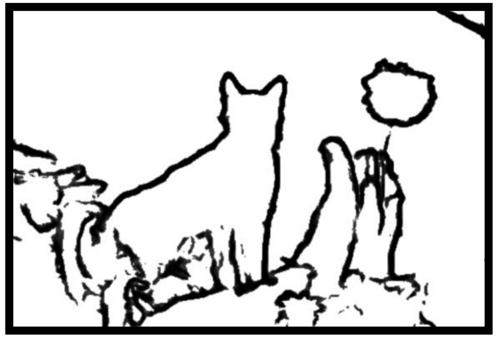
Style transfer



[Gatys et al. 2016, …]

Challenges

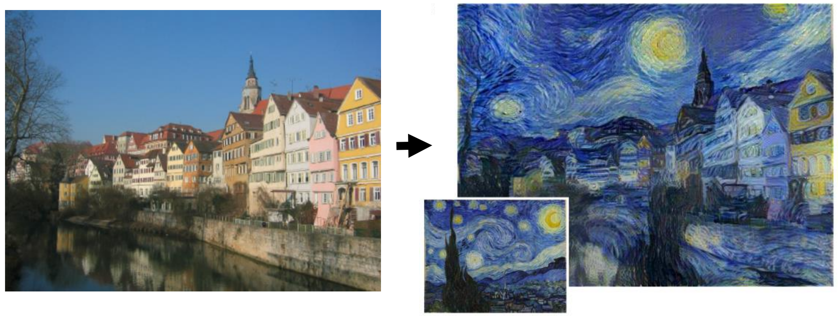
1. Output is high-dimensional, structured object

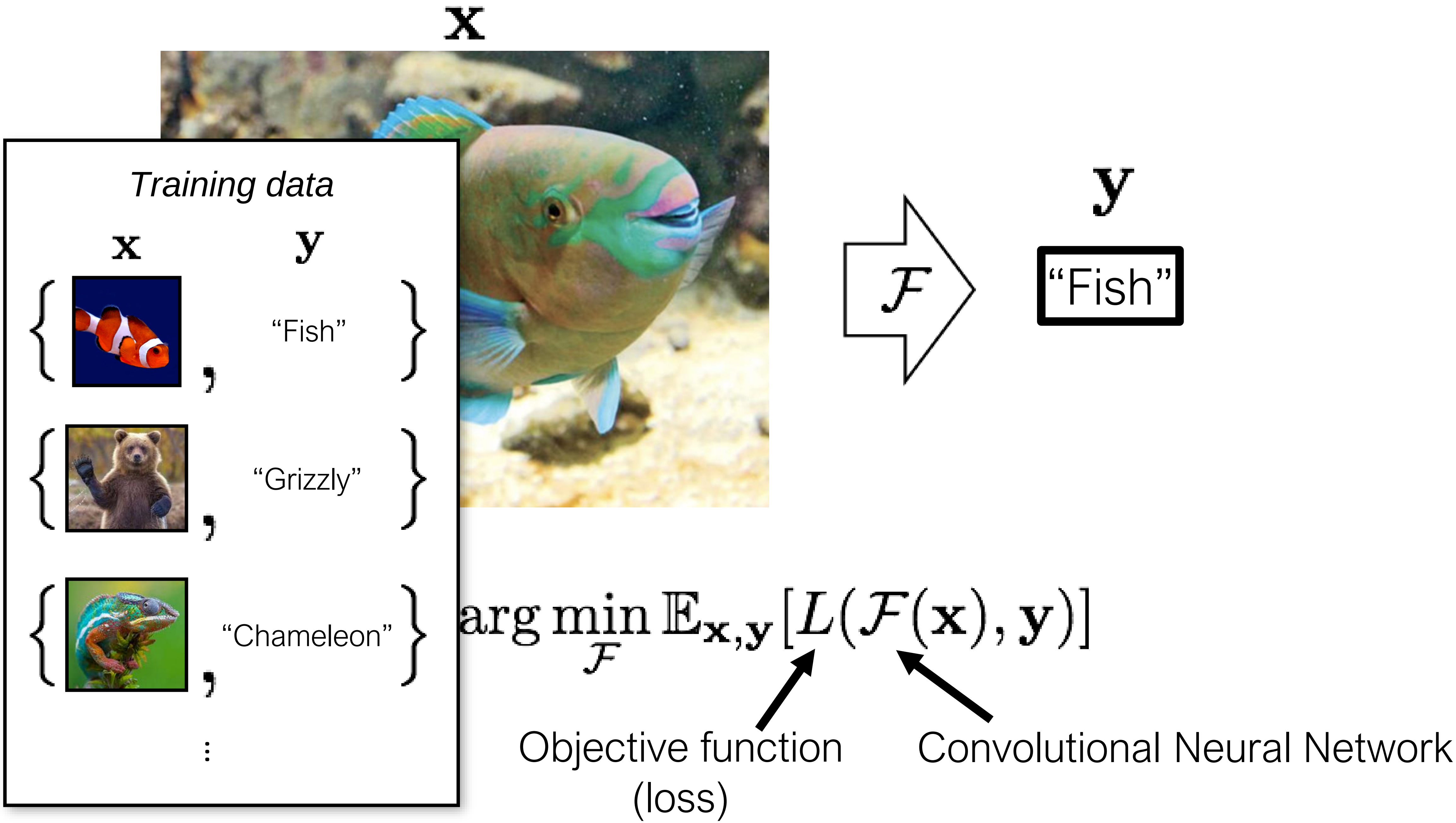


2. Uncertainty in mapping; many plausible outputs

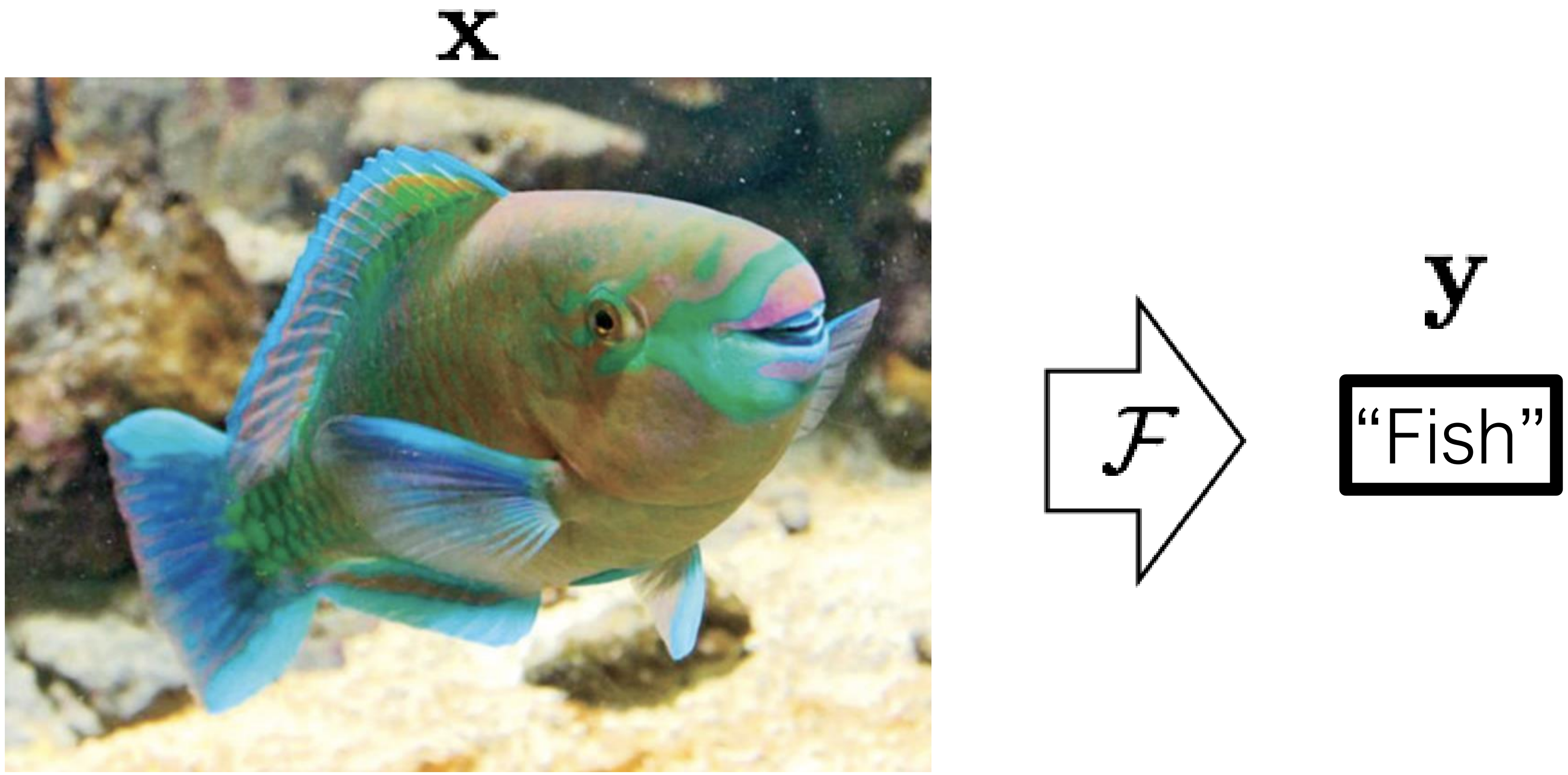
"this small bird has a pink breast and crown…"

3. Lack of supervised training data





$\mathbf{x}$
Training data
$\mathbf{x}$ $\mathbf{y}$
{ , "Fish" }
{ , "Grizzly" }
{ , "Chameleon" }
⋮
$\mathcal{F}$
$\mathbf{y}$
"Fish"
$\arg\min_{\mathcal{F}} \mathbb{E}_{\mathbf{x},\mathbf{y}}[L(\mathcal{F}(\mathbf{x}), \mathbf{y})]$
Objective function (loss)
Convolutional Neural Network



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

“**What** should I do”　　“**How** should I do it?”

Basic loss functions

Prediction: $\hat{\mathbf{y}} = \mathcal{F}(\mathbf{x})$

Truth: $\mathbf{y}$

Classification (cross-entropy):

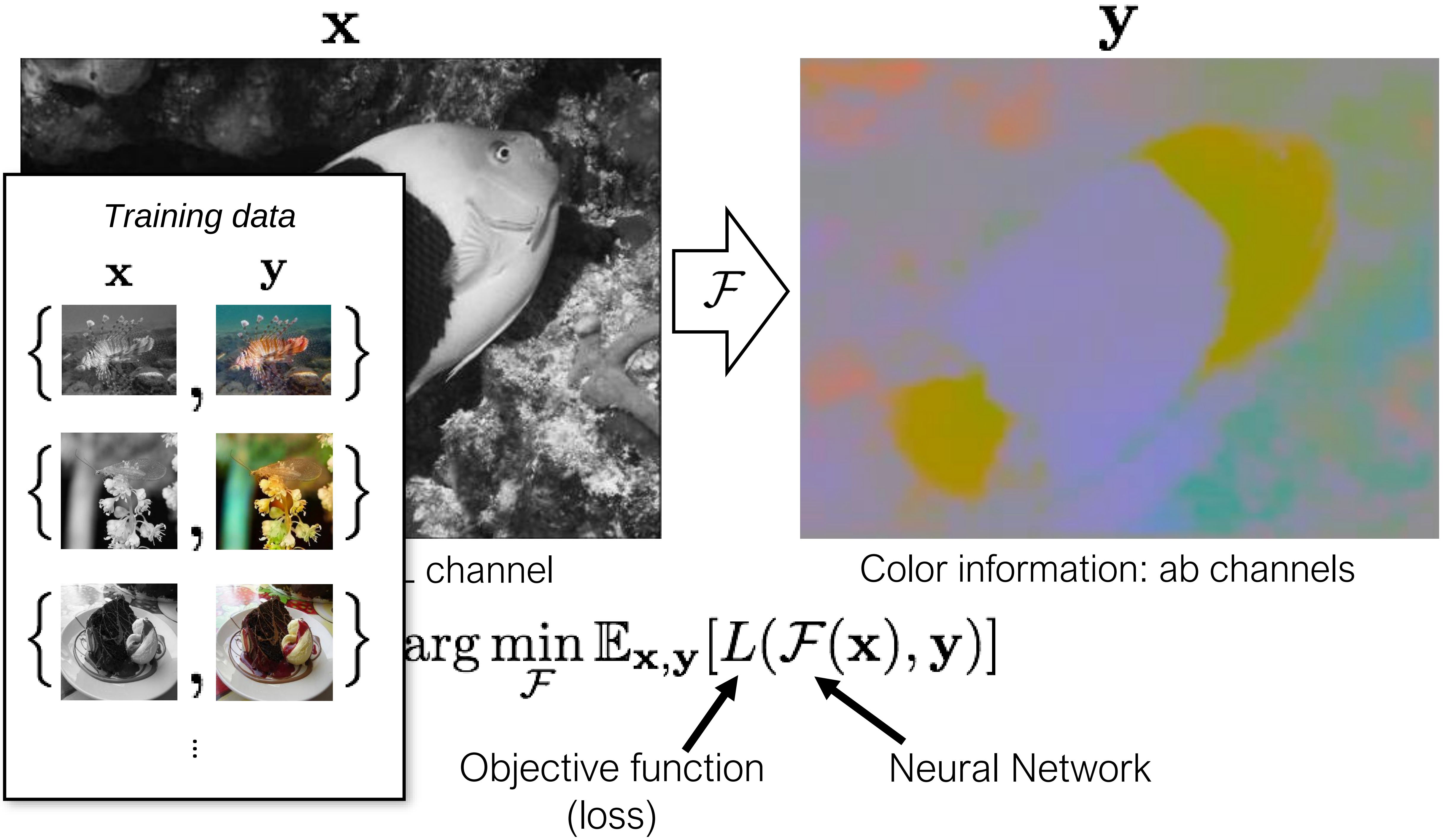
$$L(\hat{\mathbf{y}}, \mathbf{y}) = -\sum_i \hat{\mathbf{y}}_i \log \mathbf{y}_i \longleftarrow$$

How many extra bits it takes to correct the predictions

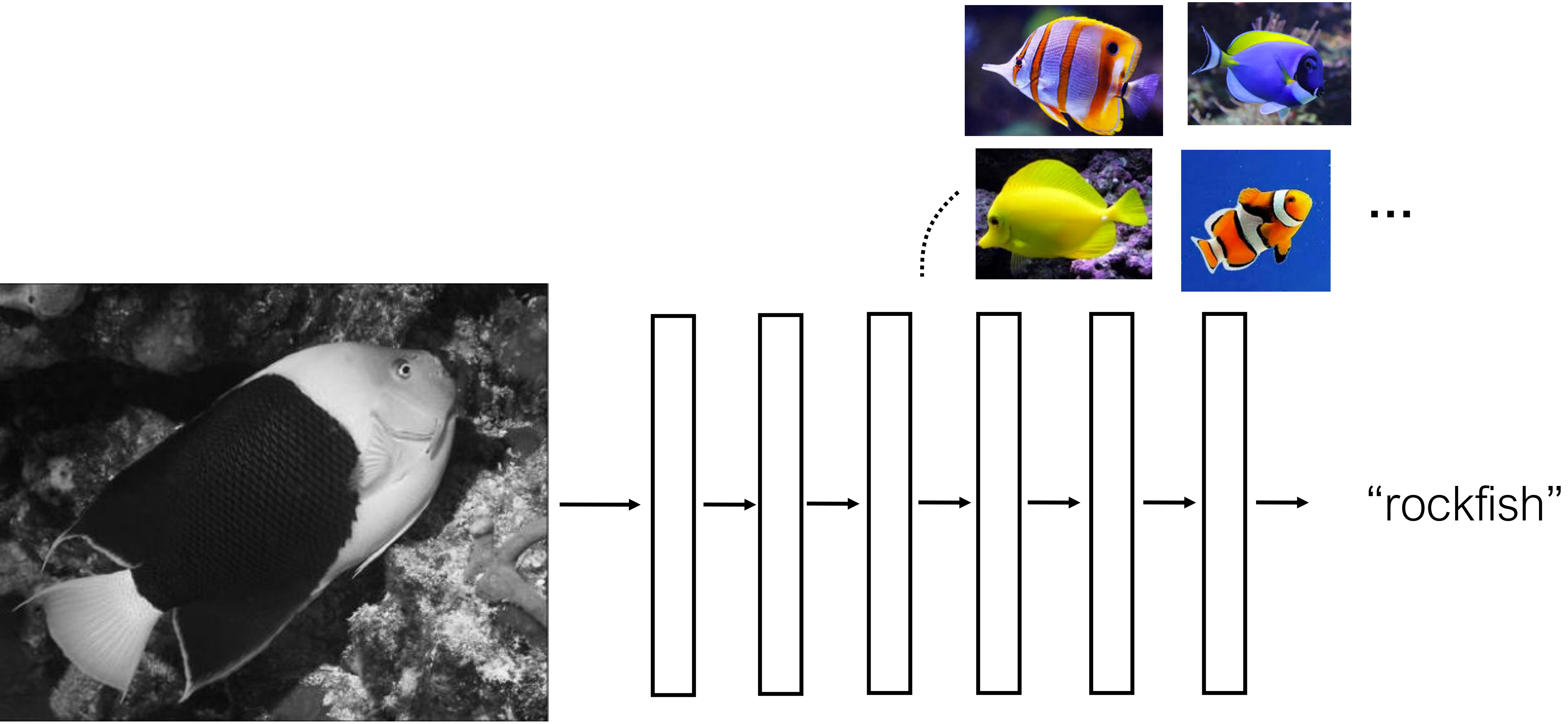
Least-squares regression:

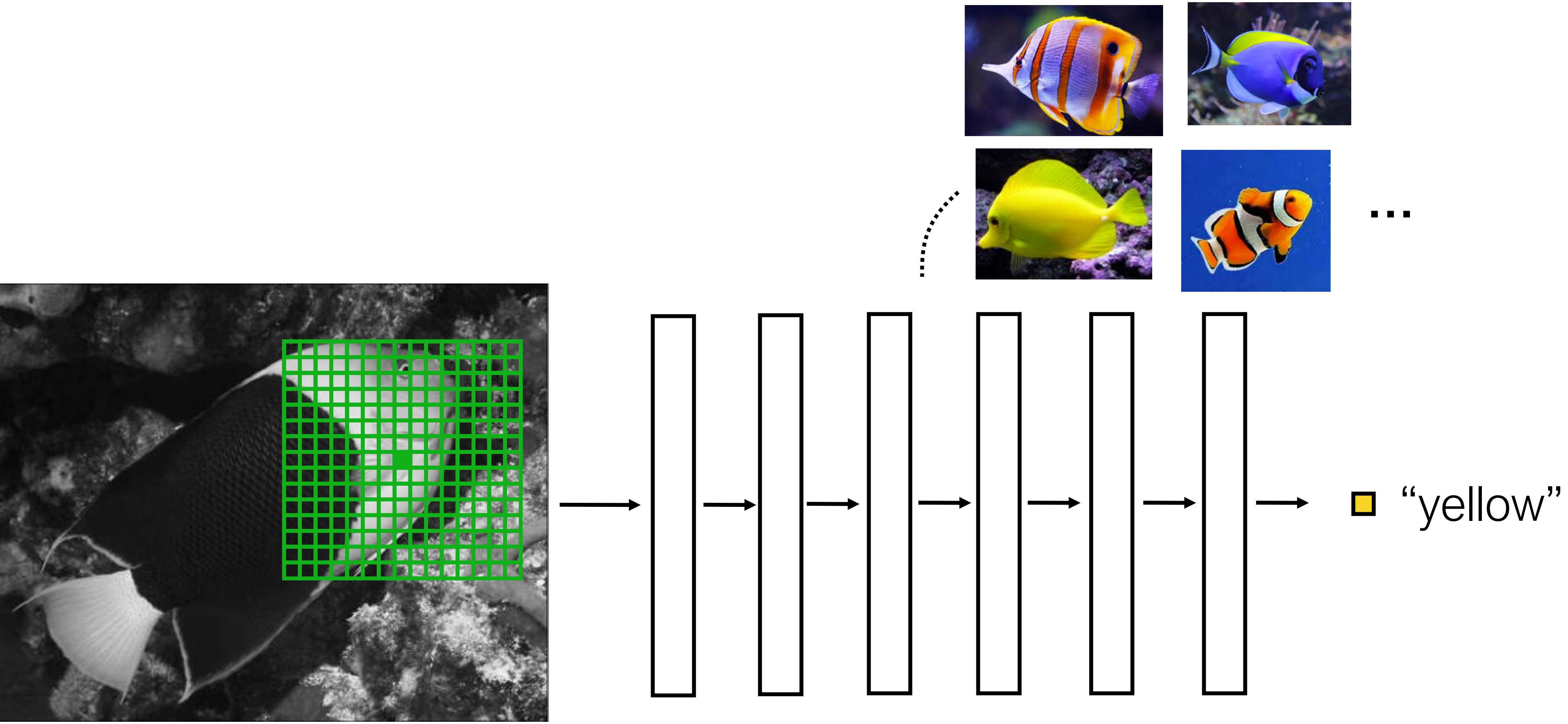
$$L(\hat{\mathbf{y}}, \mathbf{y}) = \|\hat{\mathbf{y}} - \mathbf{y}\|_2 \longleftarrow$$

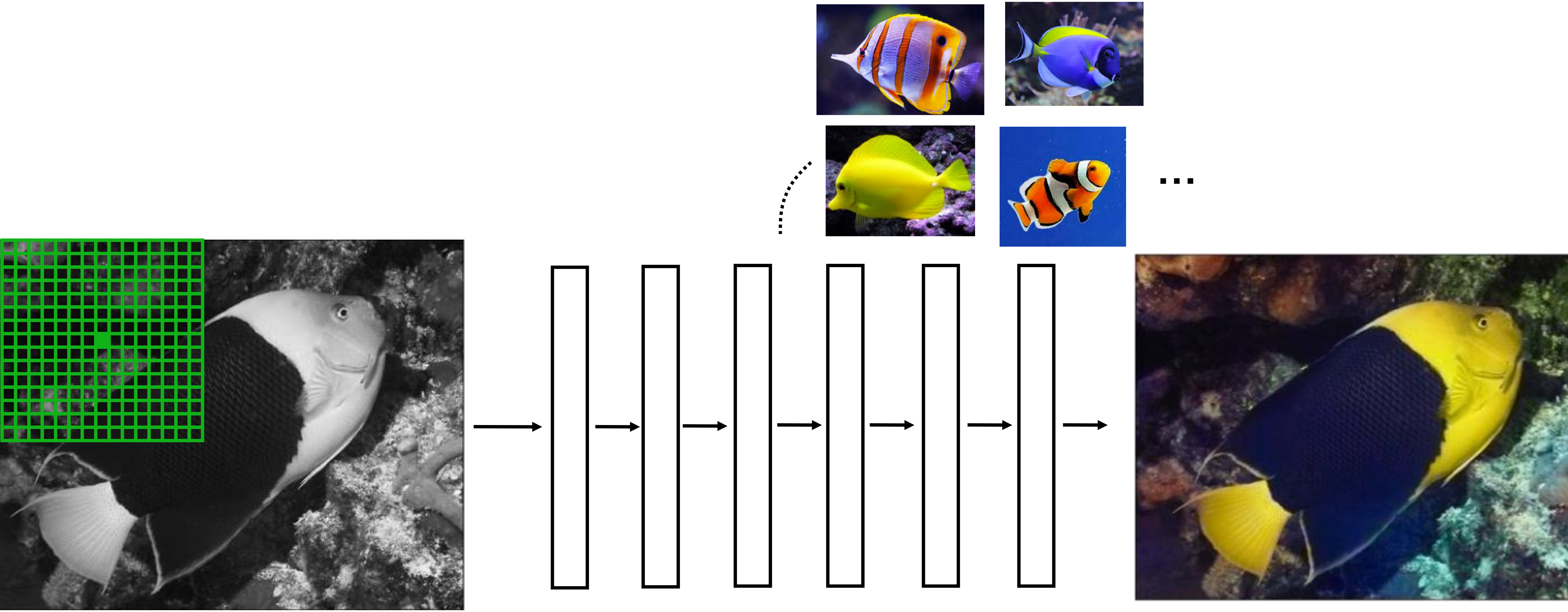
How far off we are in Euclidean distance



x
y
Training data
x
y
$\mathcal{F}$
channel
Color information: ab channels
$\arg\min_{\mathcal{F}} \mathbb{E}_{\mathbf{x},\mathbf{y}}[L(\mathcal{F}(\mathbf{x}), \mathbf{y})]$
Objective function (loss)
Neural Network


...
“rockfish”

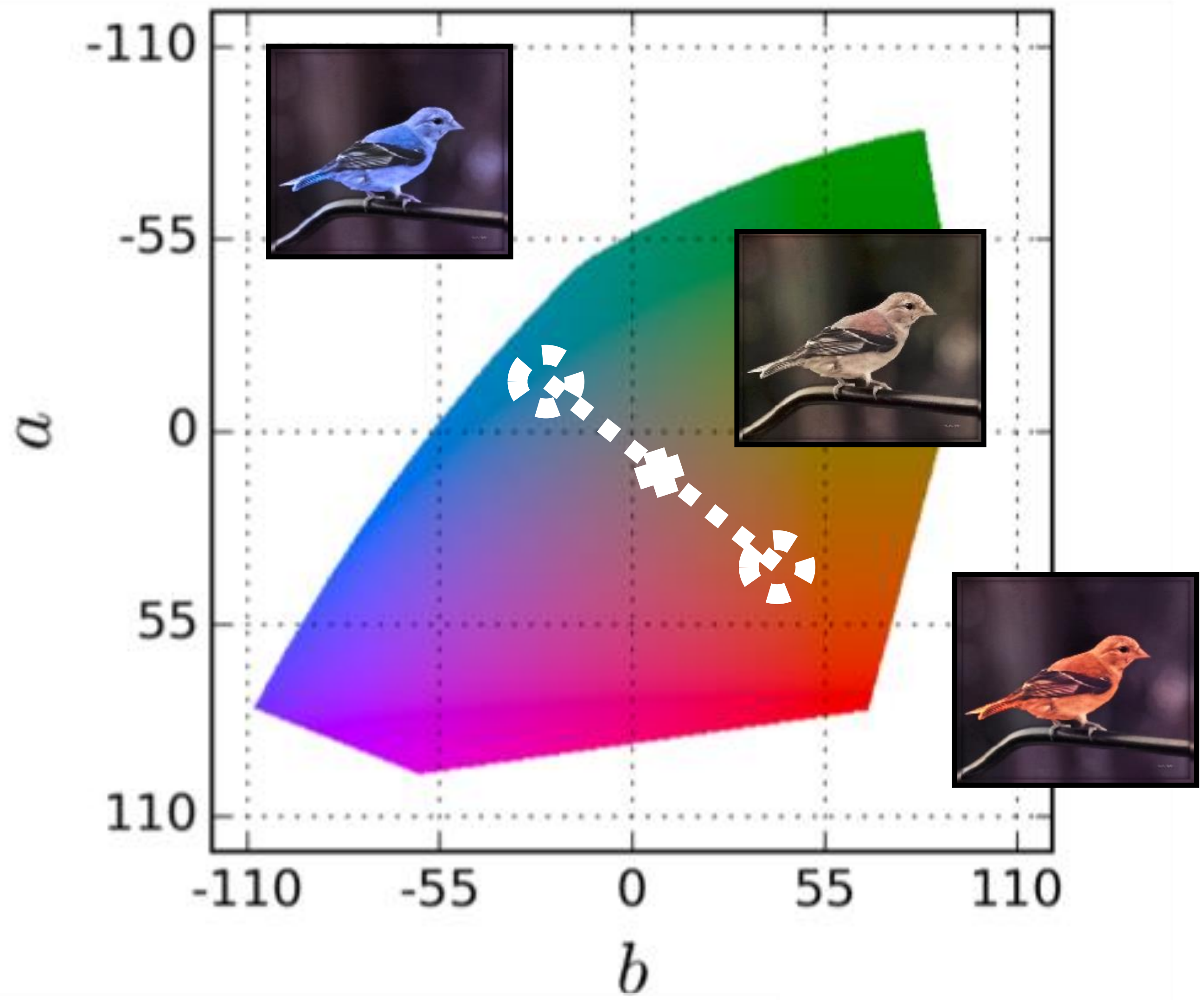

...
"yellow"



Designing loss functions

Input Output Ground truth

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



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

Designing loss functions

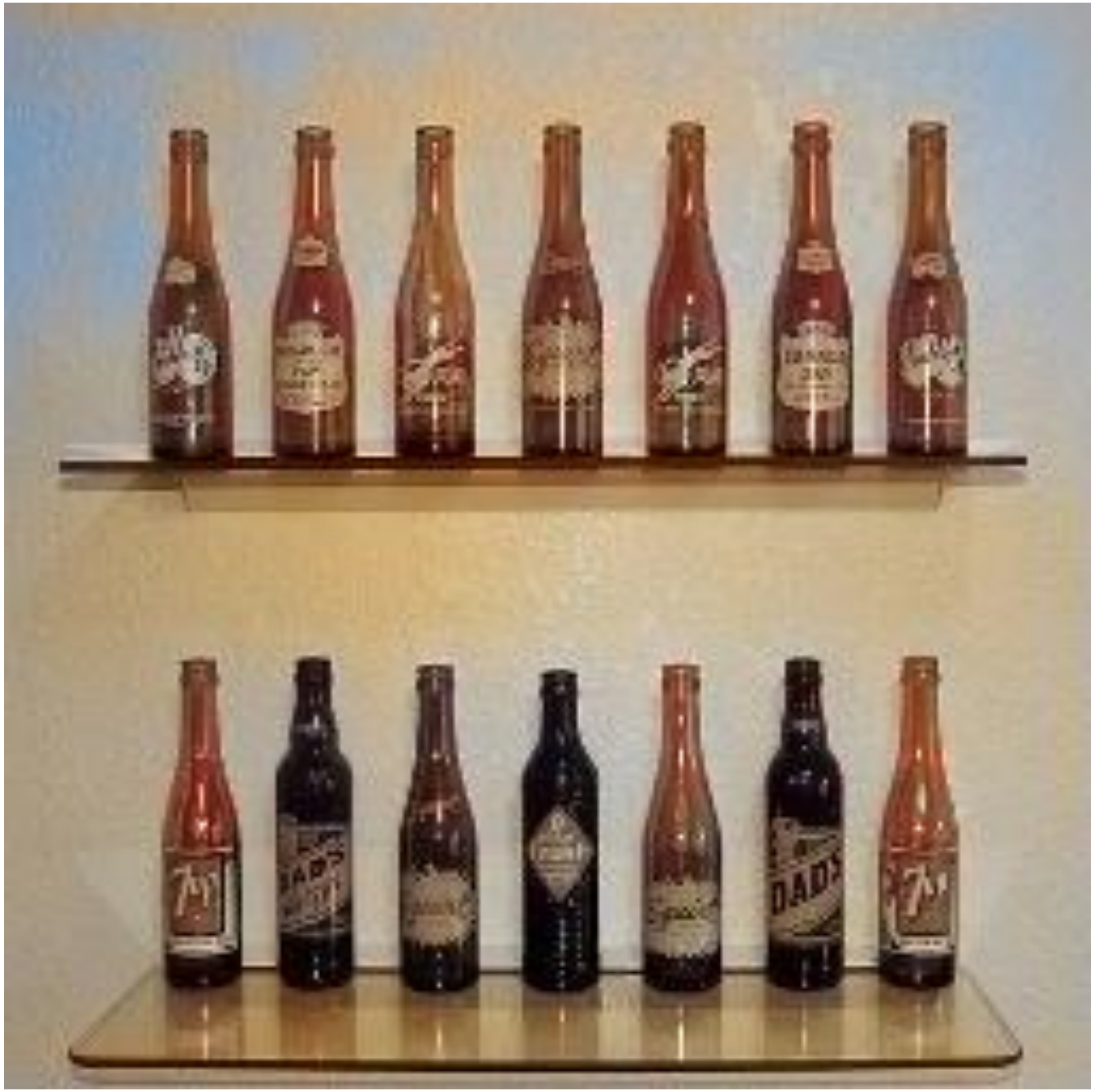
Input

Zhang et al. 2016

Ground truth

Color distribution cross-entropy loss with colorfulness enhancing term.

[Zhang, Isola, Efros, ECCV 2016]

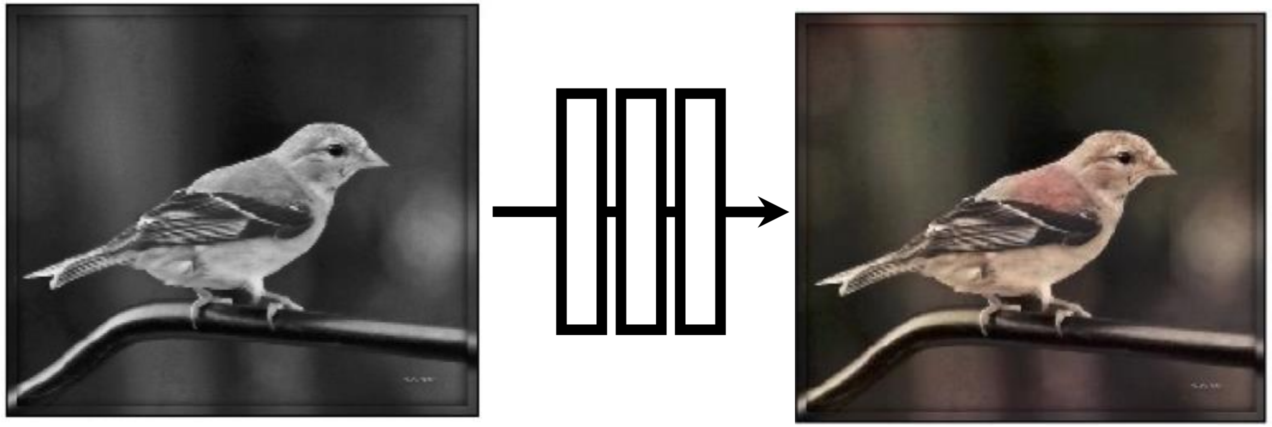

DADS

Designing loss functions

Be careful what you wish for!

Designing loss functions

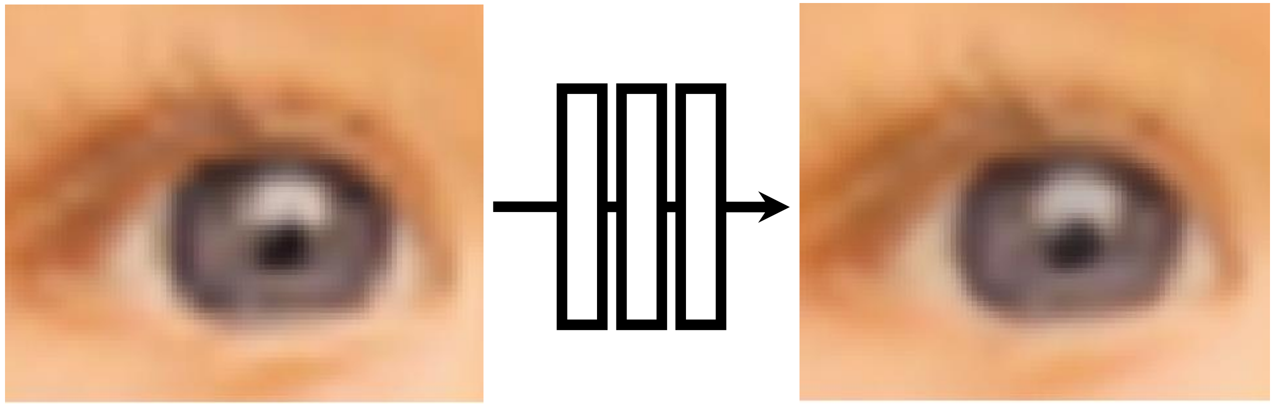
Image colorization



L2 regression

[Zhang, Isola, Efros, ECCV 2016]

Super-resolution

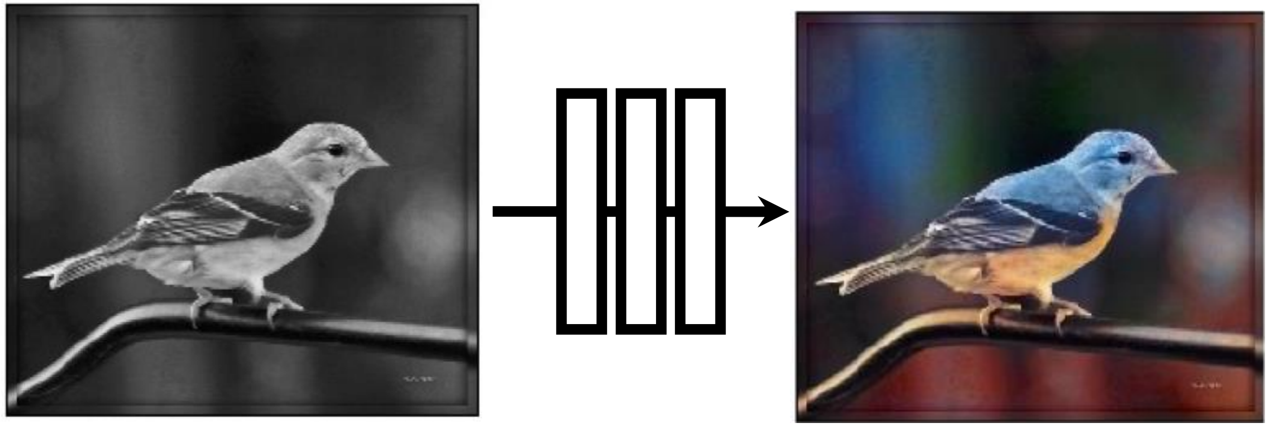


L2 regression

[Johnson, Alahi, Li, ECCV 2016]

Designing loss functions

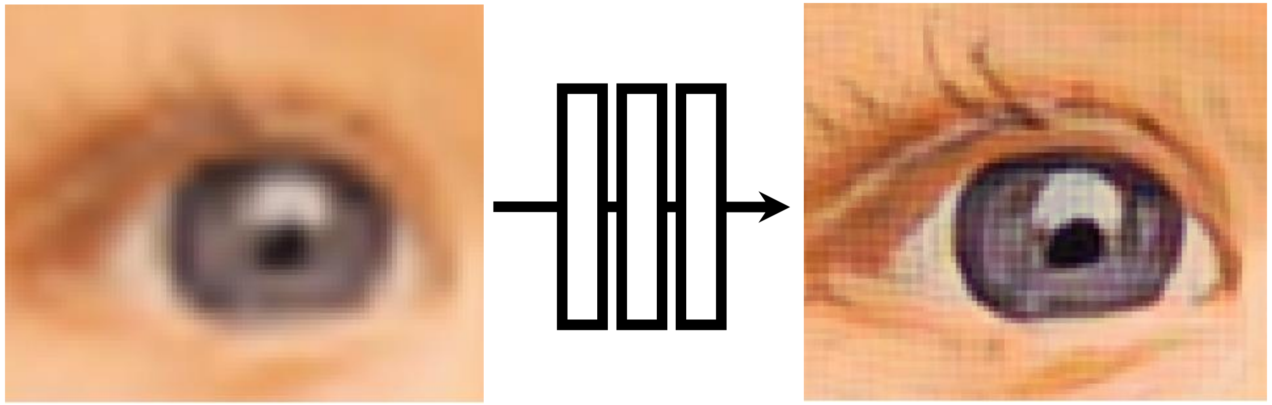
Image colorization



[Zhang, Isola, Efros, ECCV 2016]

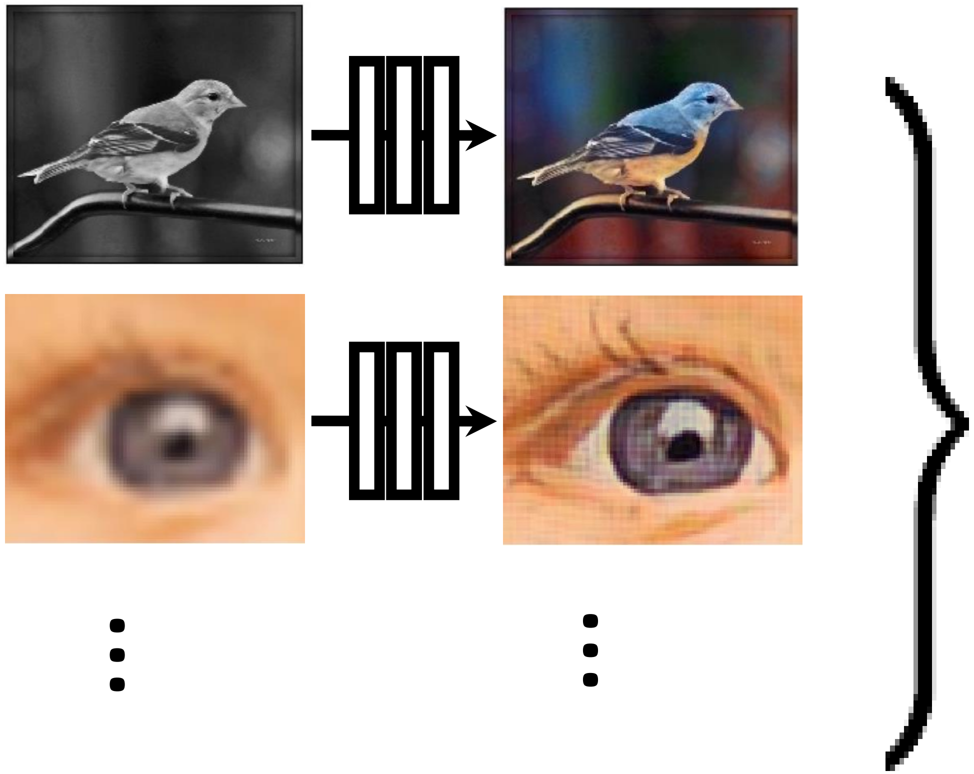
Cross entropy objective, with colorfulness term

Super-resolution



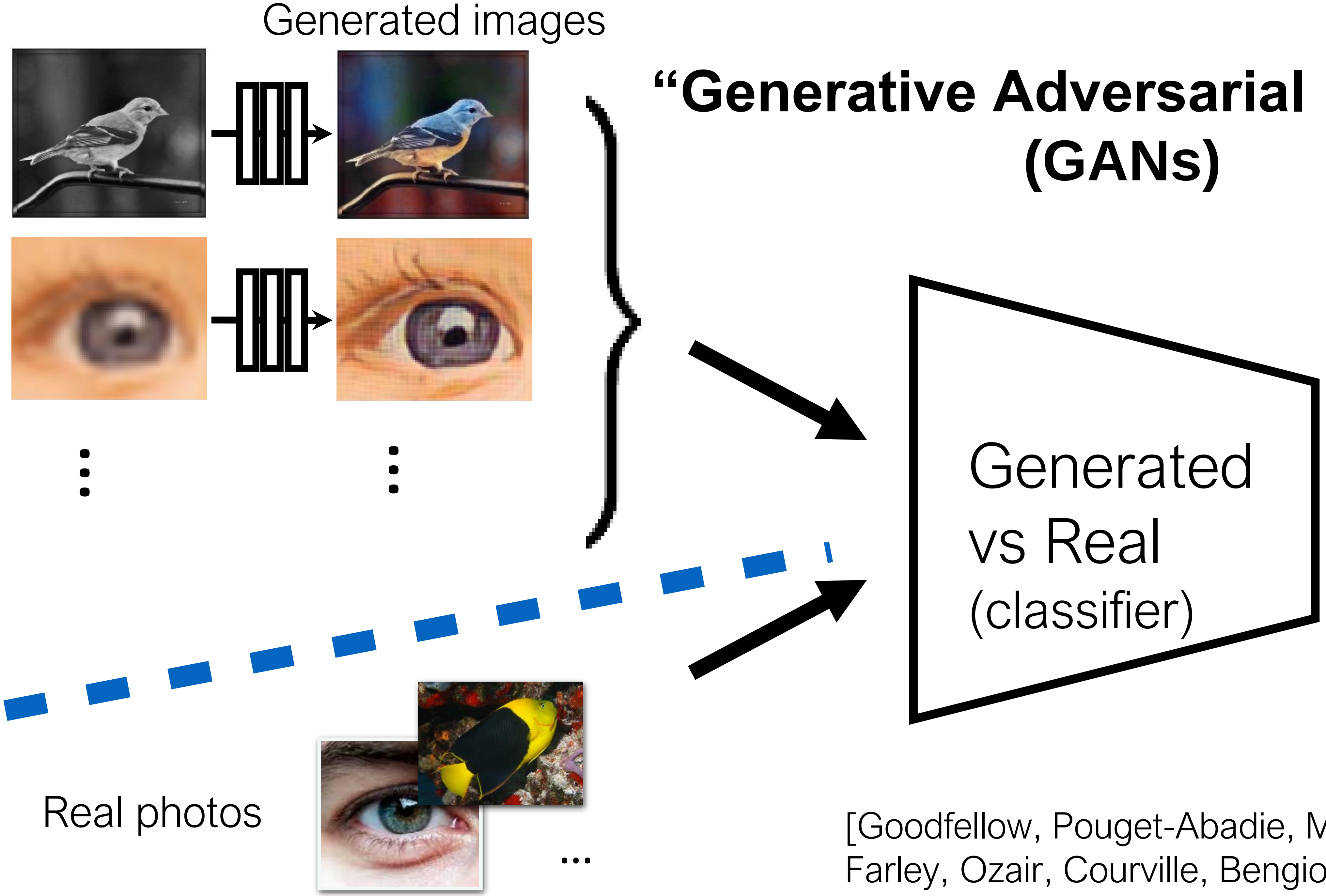
[Johnson, Alahi, Li, ECCV 2016]

Deep feature covariance matching objective



Universal loss?

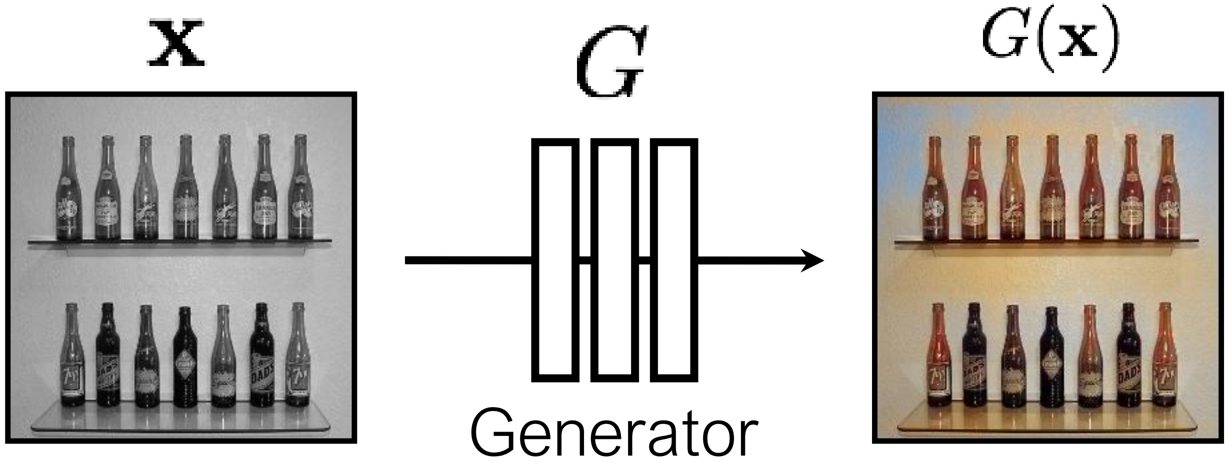
“Generative Adversarial Network” (GANs)



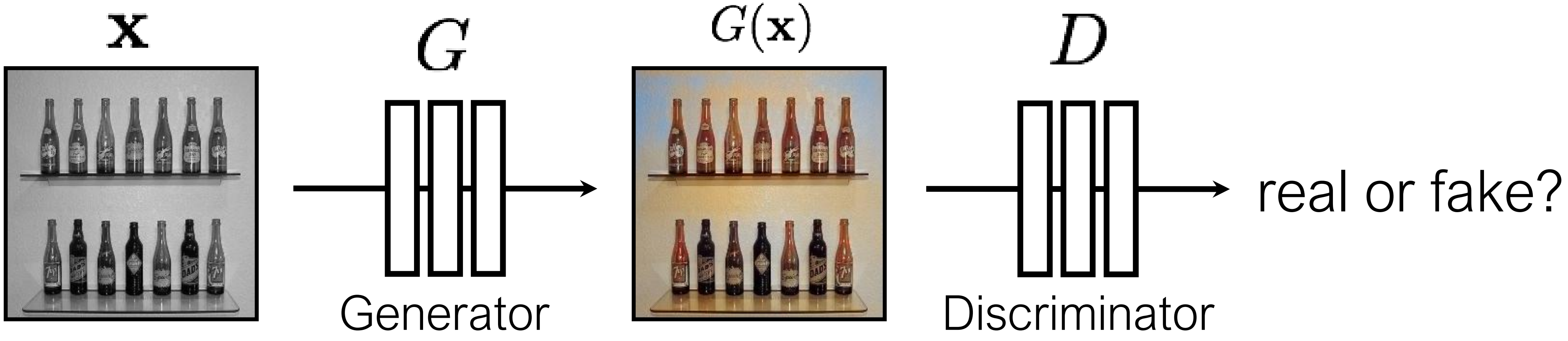
[Goodfellow, Pouget-Abadie, Mirza, Xu, Warde-Farley, Ozair, Courville, Bengio 2014]

Conditional GANs

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



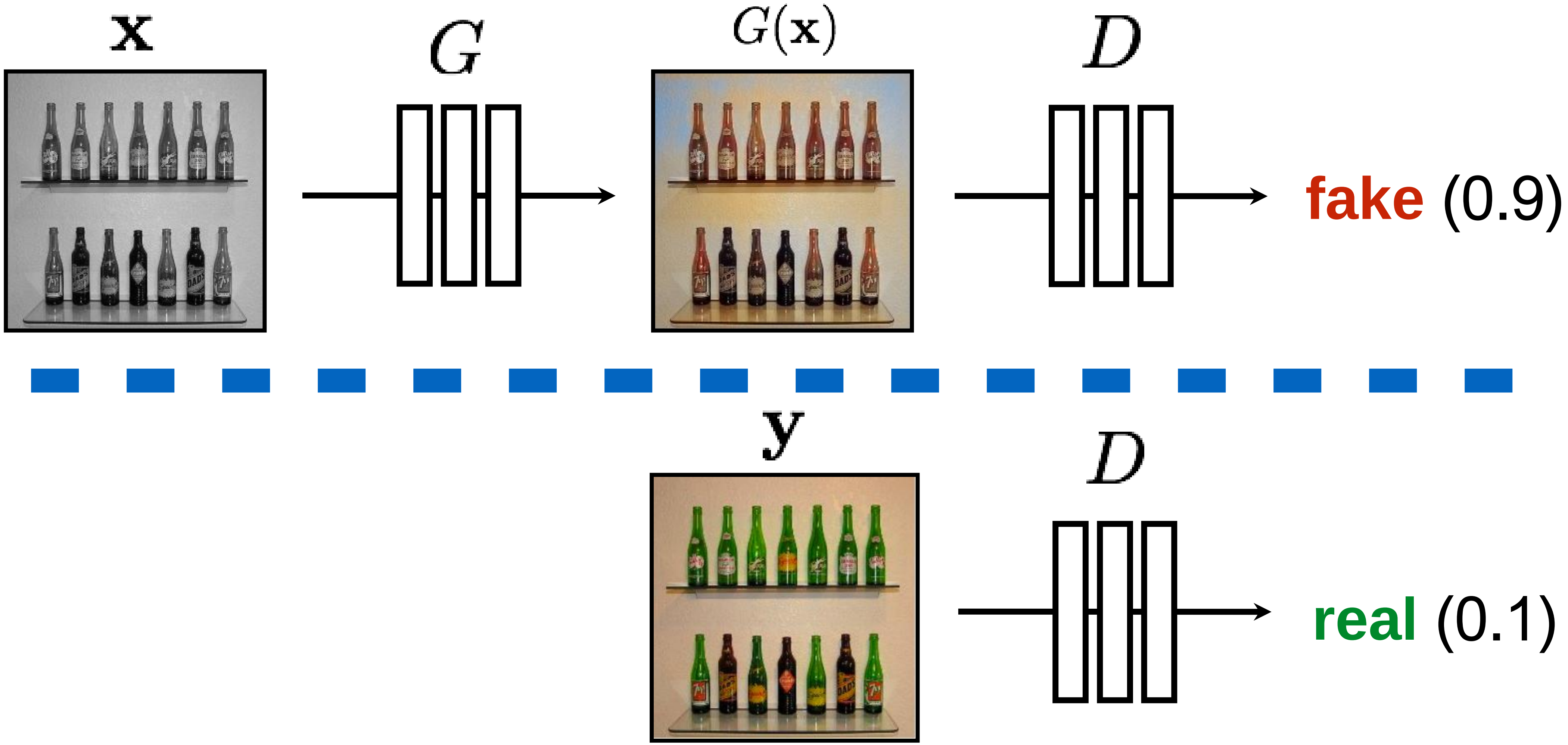
[Goodfellow et al., 2014]



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

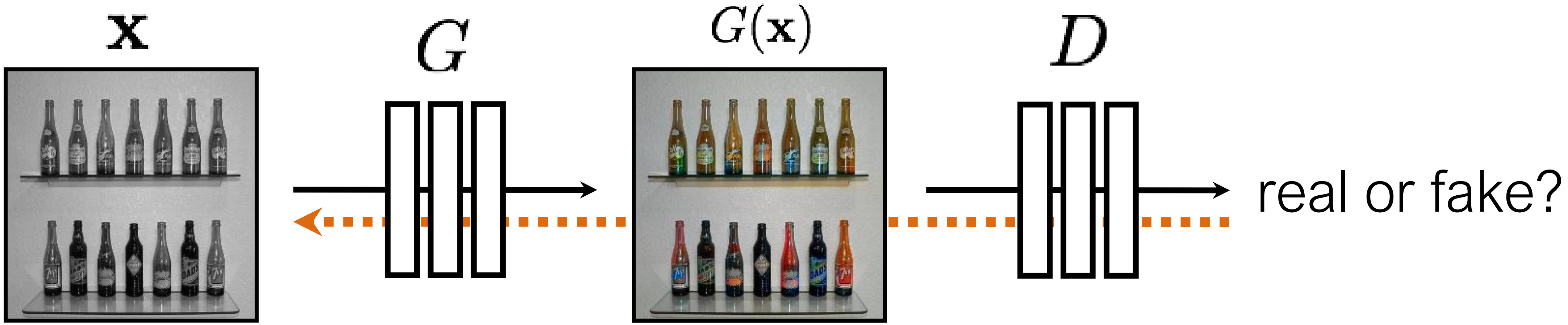
D tries to identify the fakes

[Goodfellow et al., 2014]



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

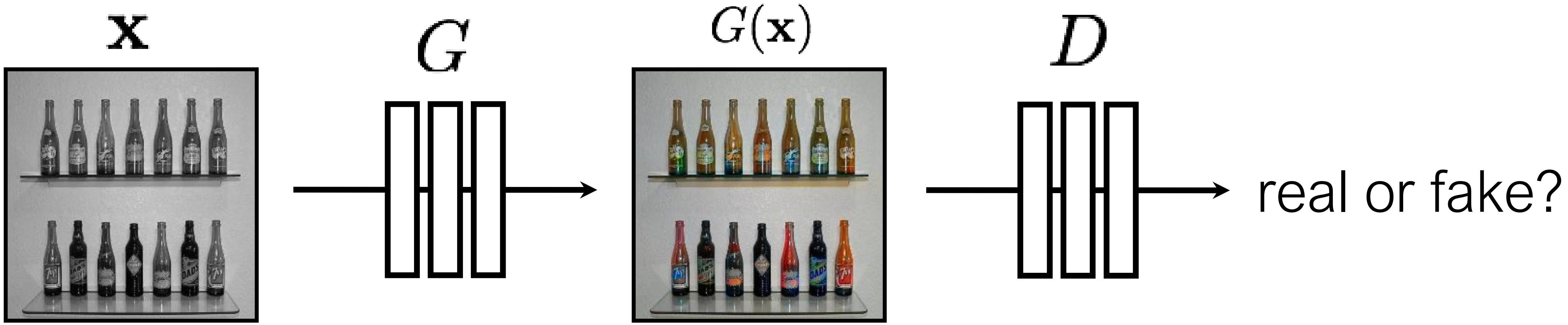
[Goodfellow et al., 2014]



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

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

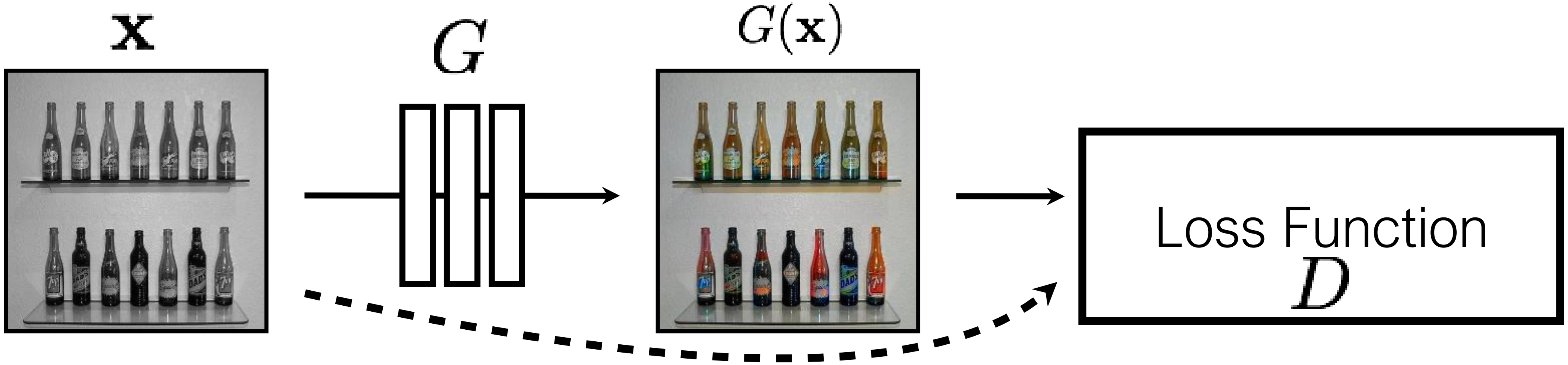
[Goodfellow et al., 2014]



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

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

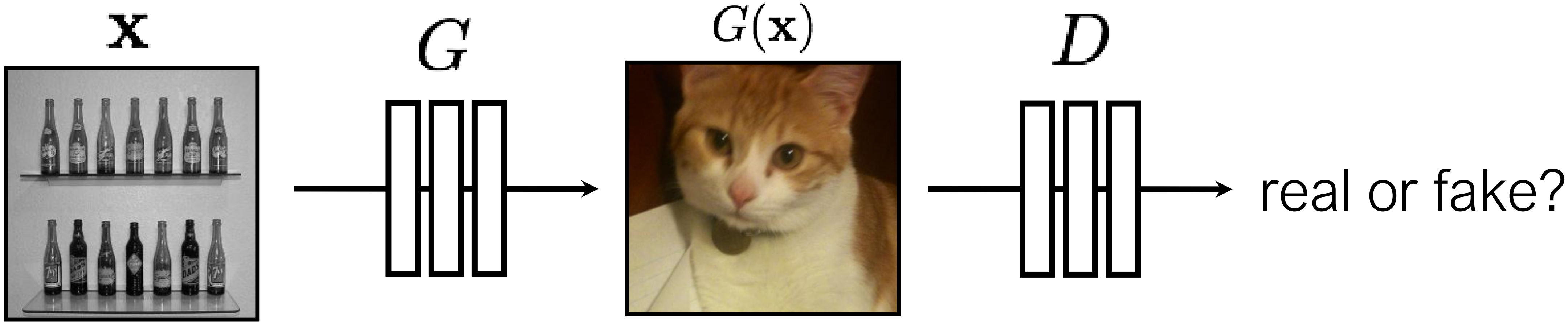
[Goodfellow et al., 2014]



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

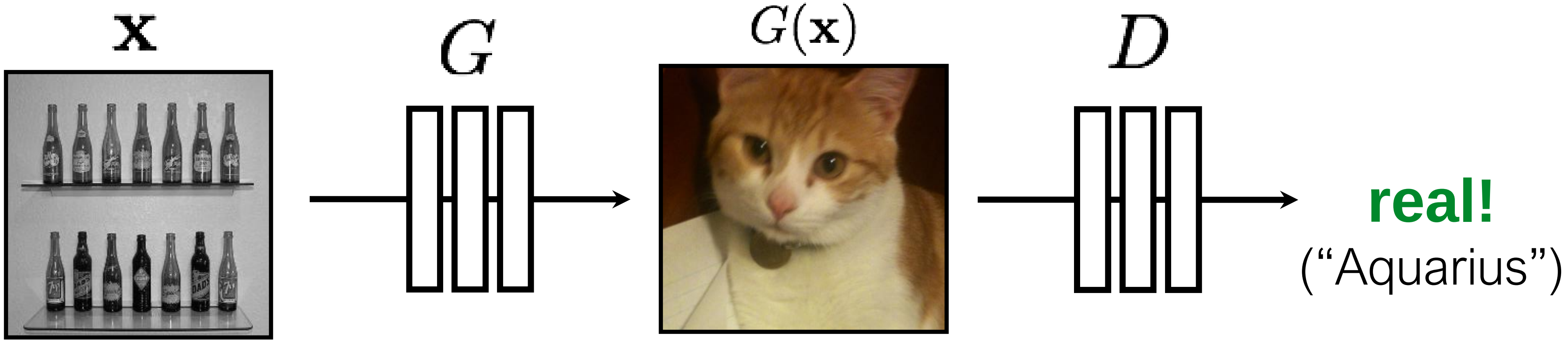
Rather than being hand-designed, it is *learned*.

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



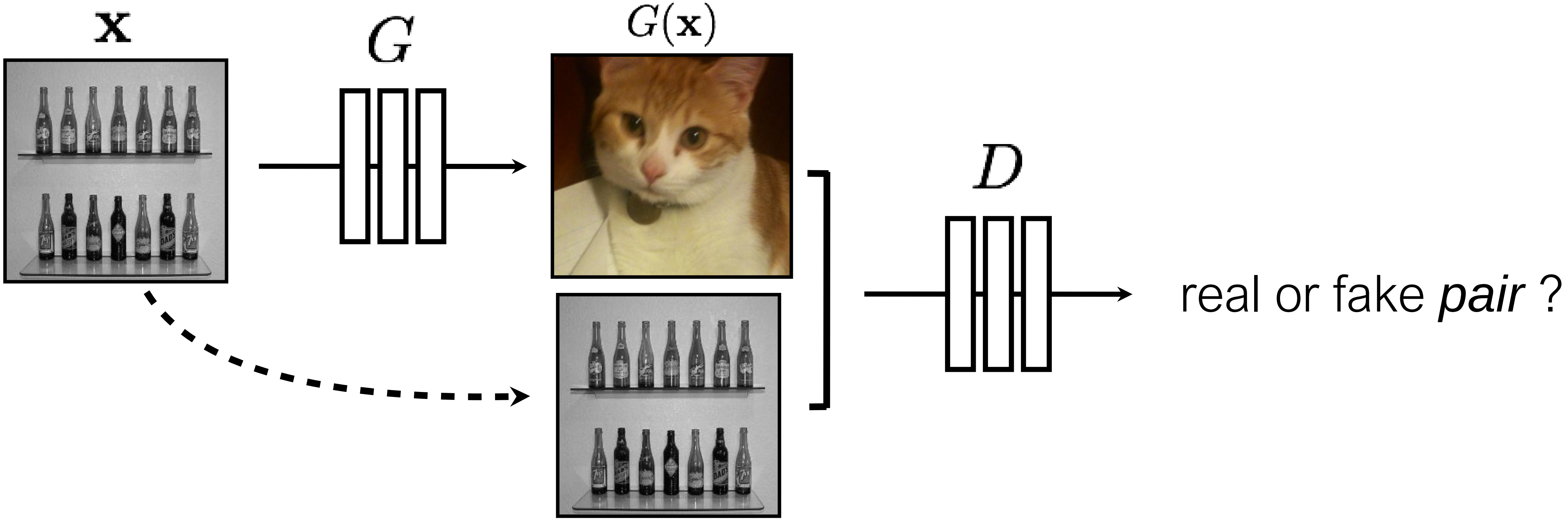
$$\arg\min_{G}\max_{D}\ \mathbb{E}_{\mathbf{x},\mathbf{y}}[\ \log D(G(\mathbf{x})) \quad + \quad \log(1 - D(\mathbf{y}))\ ]$$

[Goodfellow et al., 2014]



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

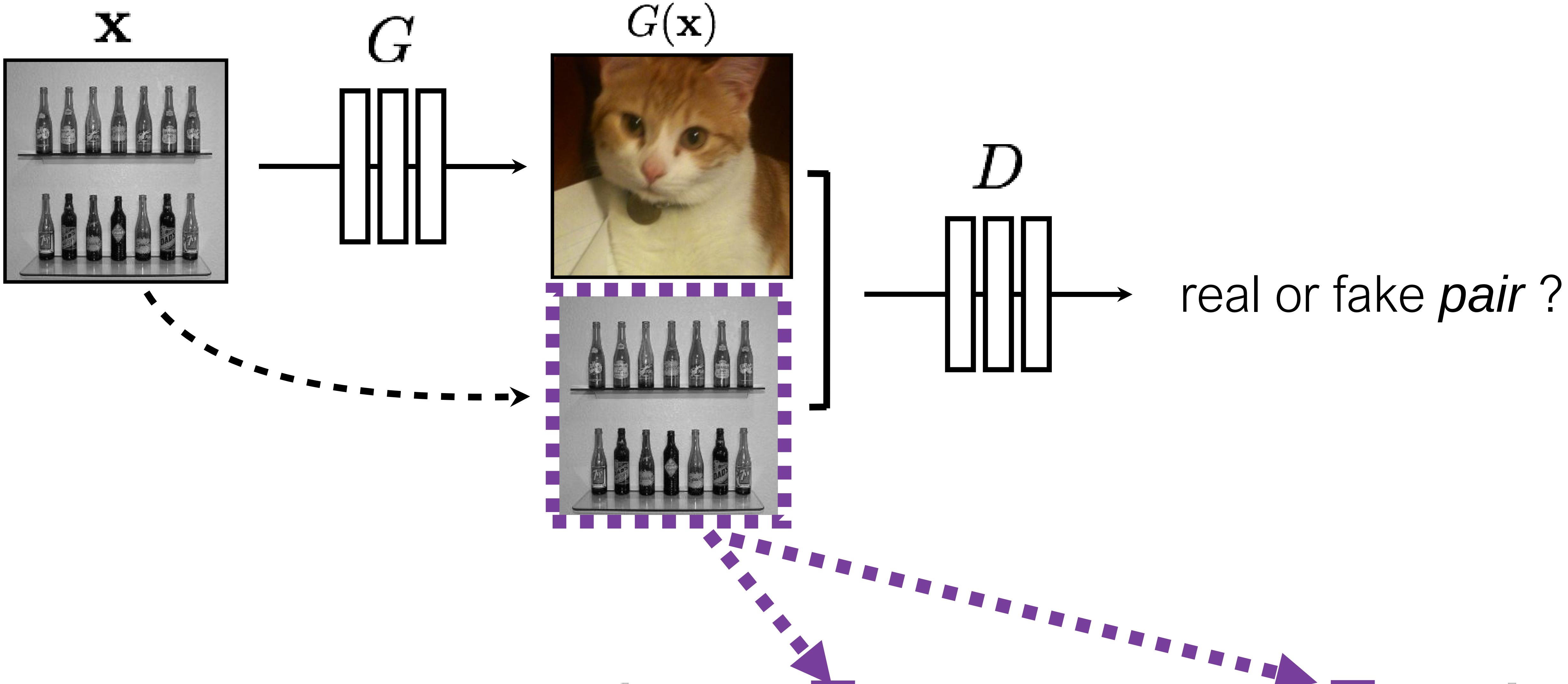
[Goodfellow et al., 2014]



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

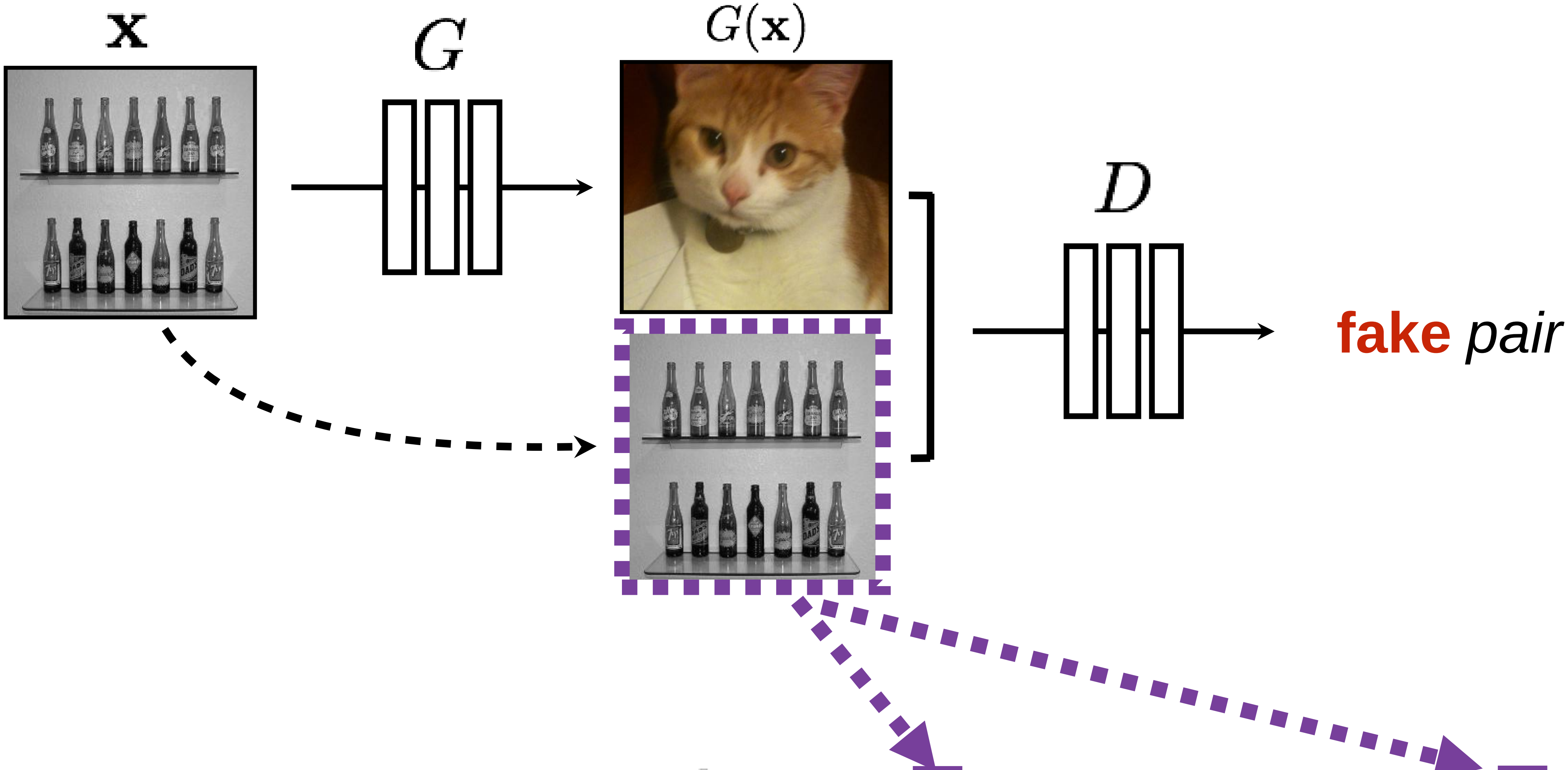
[Goodfellow et al., 2014]

[Isola et al., 2017]



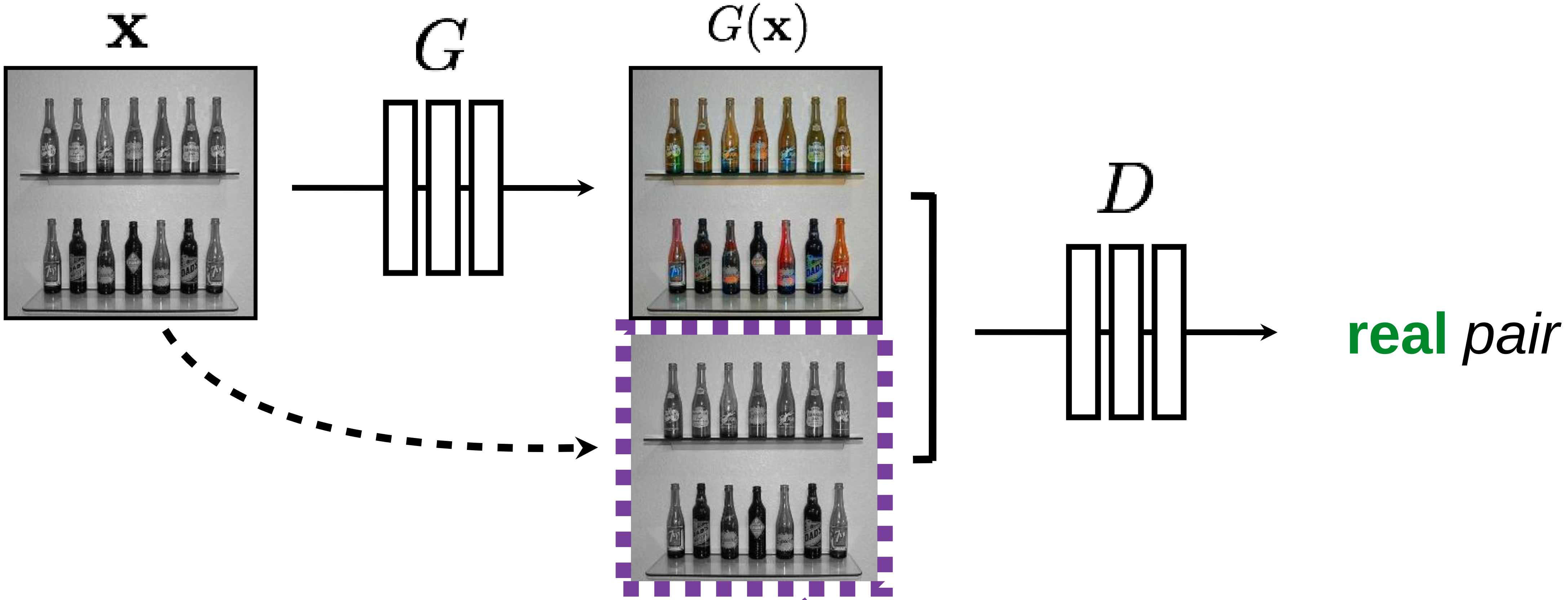
$$\arg\min_{G}\max_{D}\ \mathbb{E}_{\mathbf{x},\mathbf{y}}\left[\ \log D(\boxed{\mathbf{x}}, G(\mathbf{x})) + \log(1 - D(\boxed{\mathbf{x}}, \mathbf{y}))\ \right]$$

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



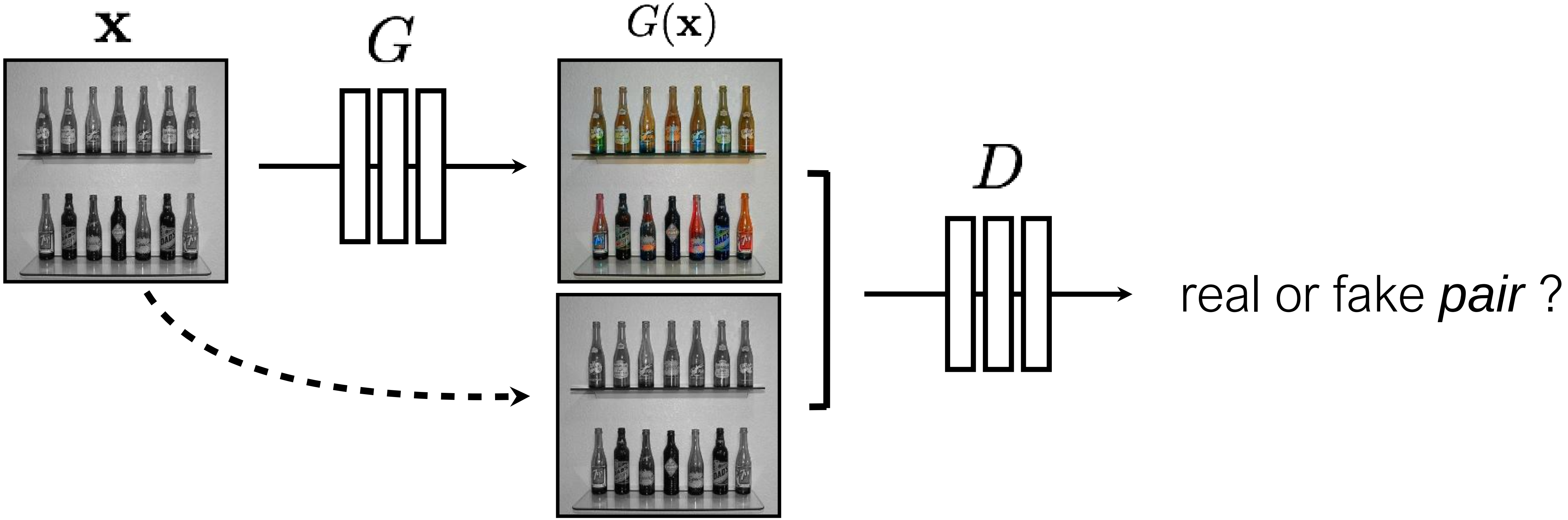
$$\arg\min_{G}\max_{D}\ \mathbb{E}_{\mathbf{x},\mathbf{y}}[\ \log D(\boxed{\mathbf{x}}, G(\mathbf{x})) + \log(1 - D(\boxed{\mathbf{x}}, \mathbf{y}))\ ]$$

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



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

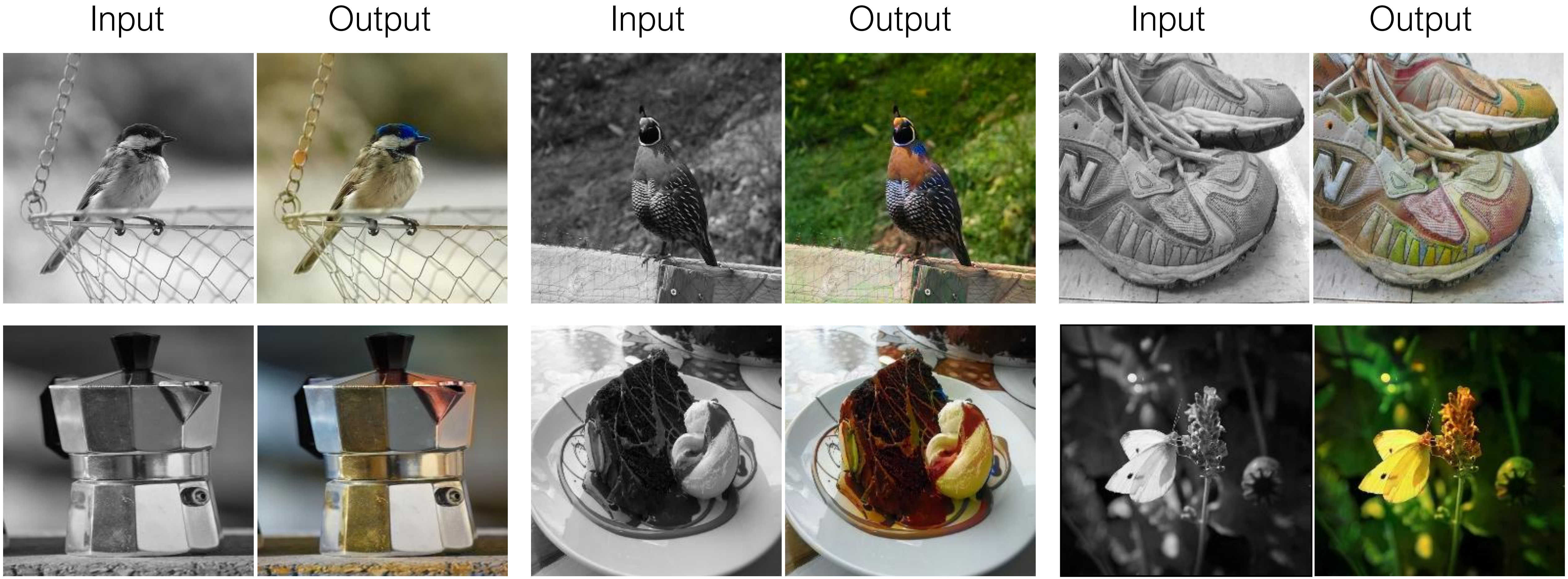
[Goodfellow et al., 2014]
[Isola et al., 2017]



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

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

BW → Color



Data from [Russakovsky et al. 2015]

Input

Output

Groundtruth

Data from
[maps.google.com]

Input

Output

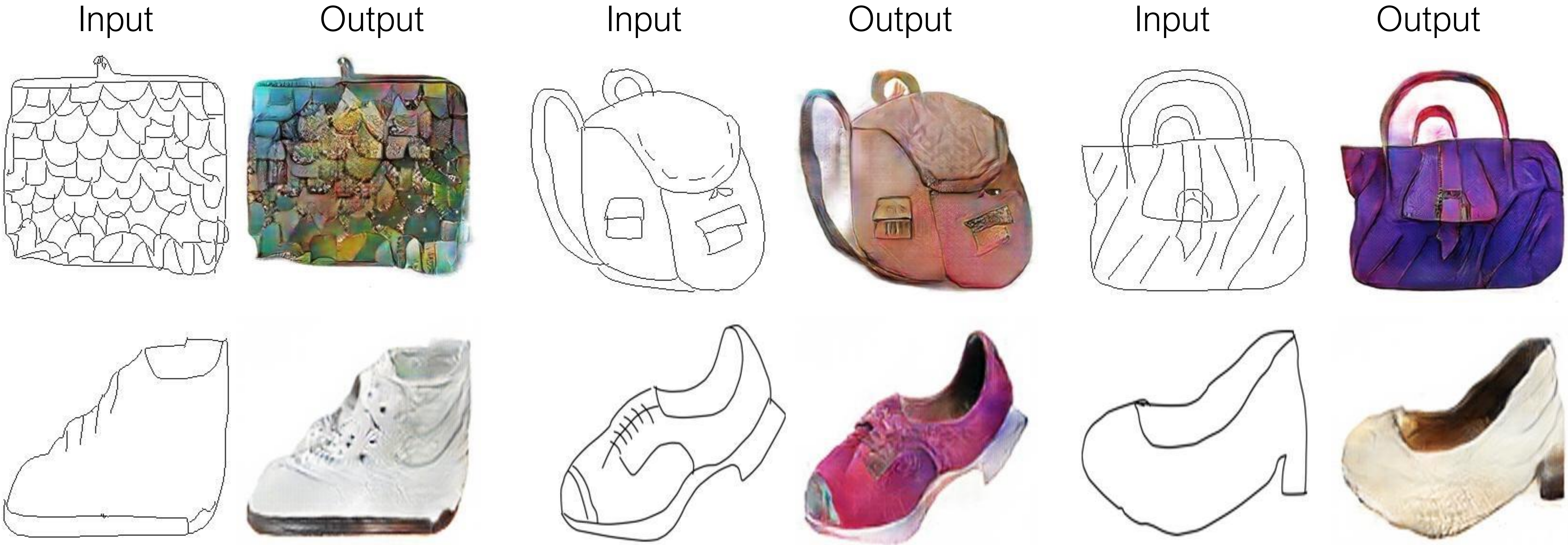
Groundtruth

Data from [maps.google.

Edges → Images

Edges from [Xie & Tu, 2015]

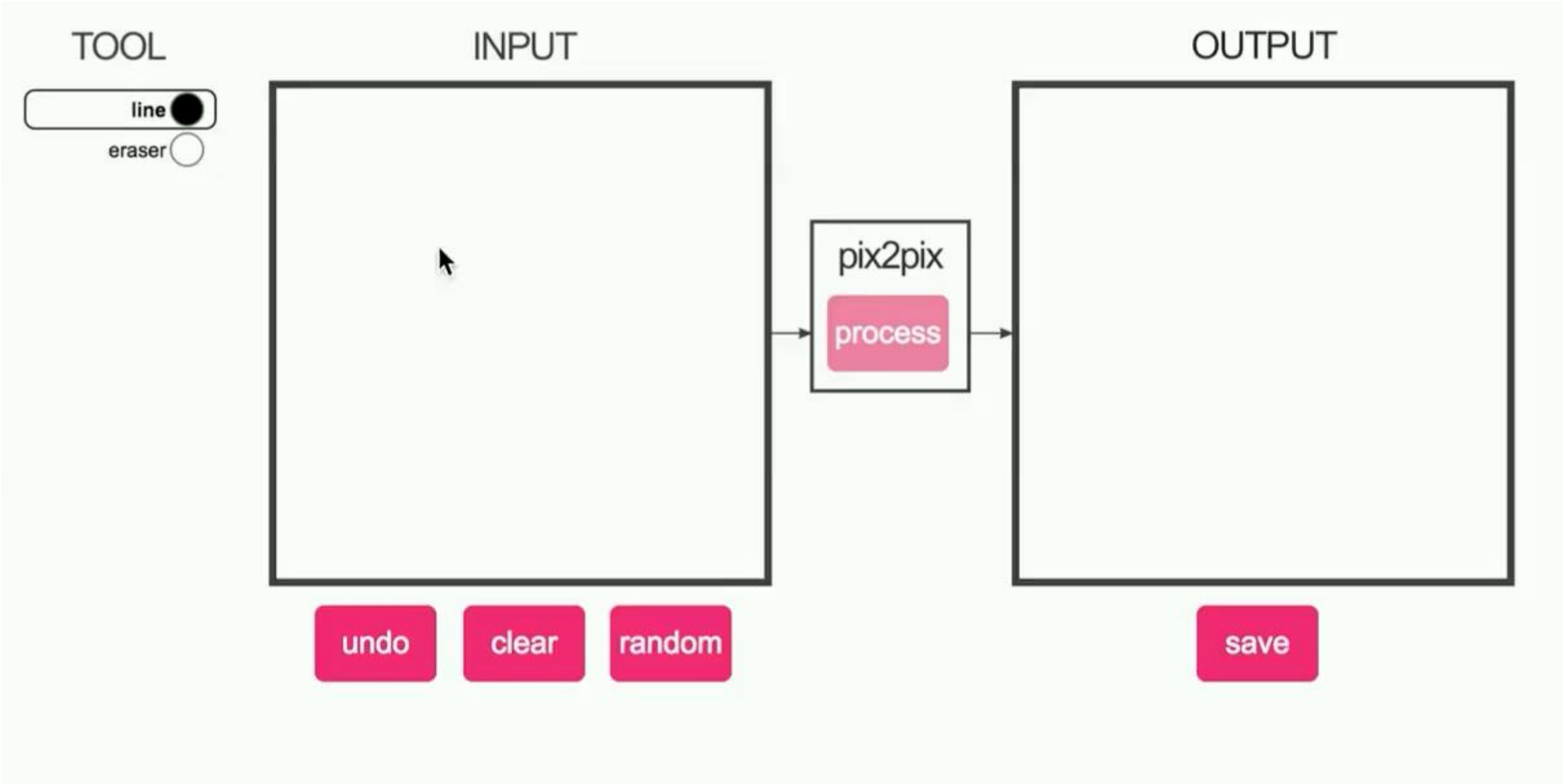
Sketches → Images

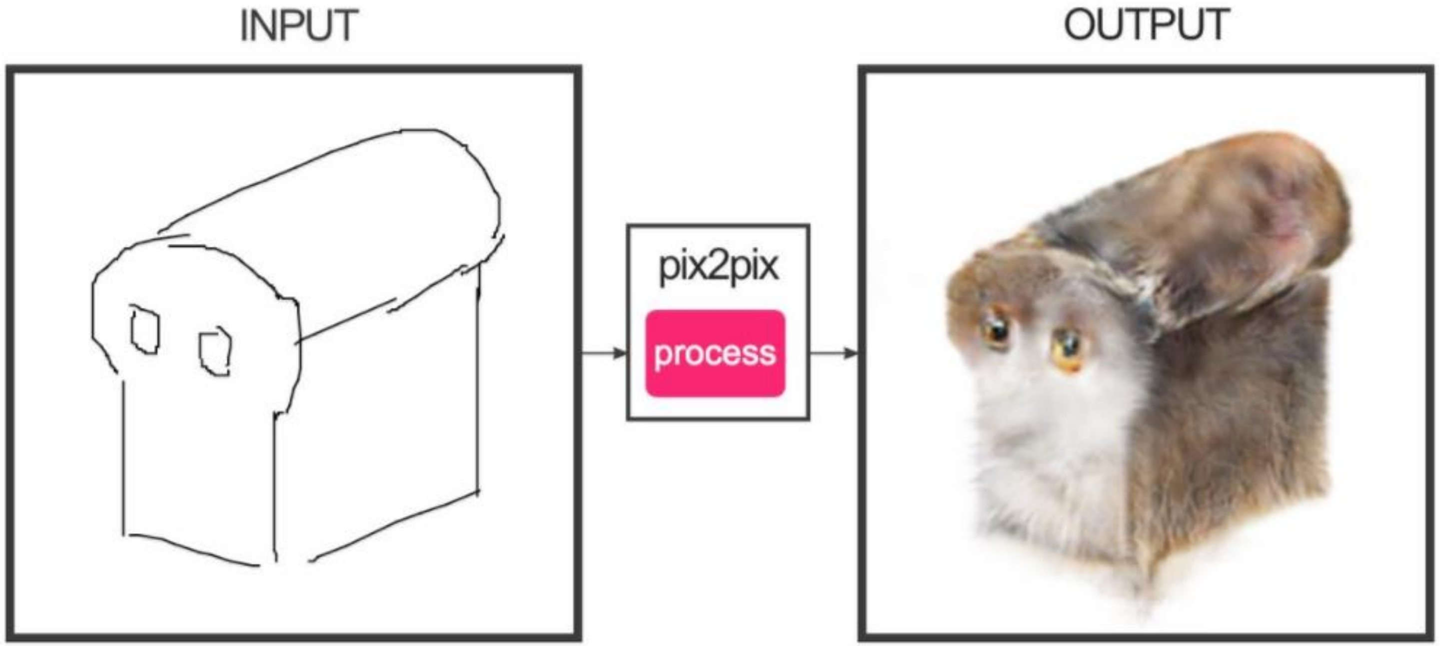


Trained on Edges → Images

Data from [Eitz, Hays, Alexa, 2012]

#edges2cats [Chris Hesse]



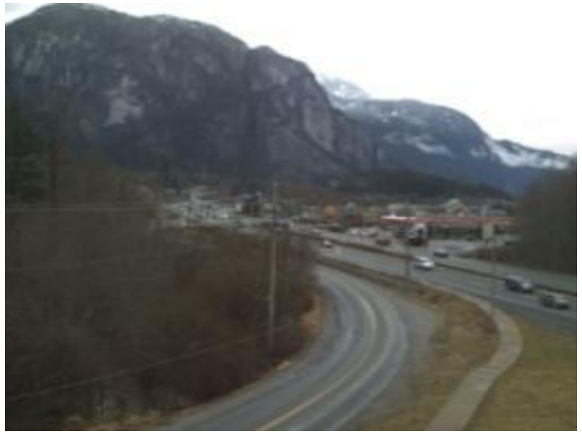


Ivy Tasi @ivymyt

Vitaly Vidmirov @vvid

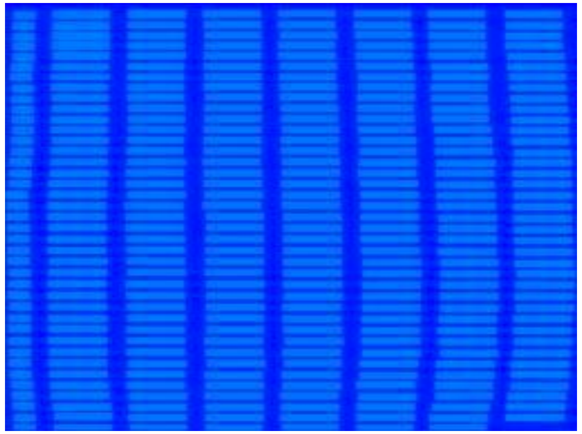
Hallucinations

Input

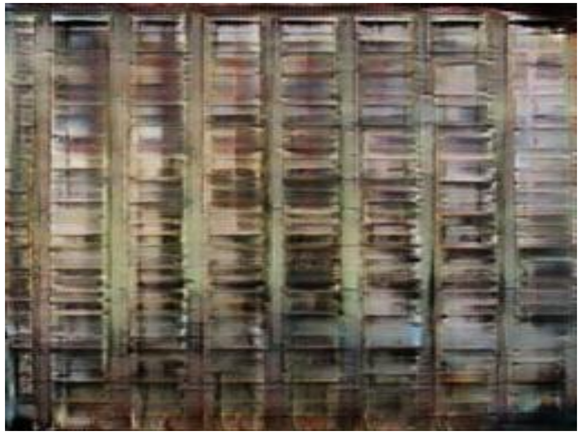


Output

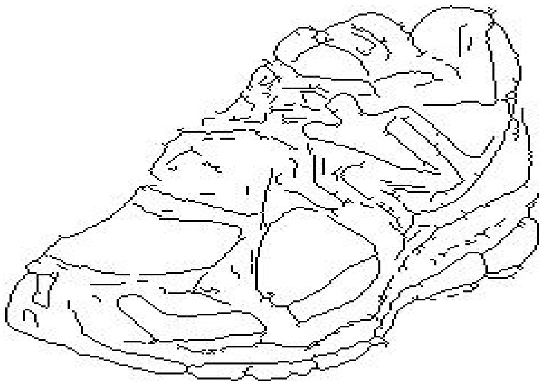
Input



Output

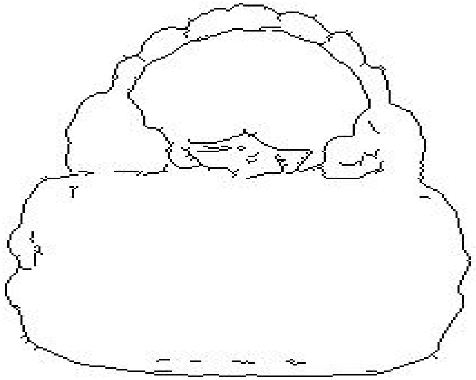


Input



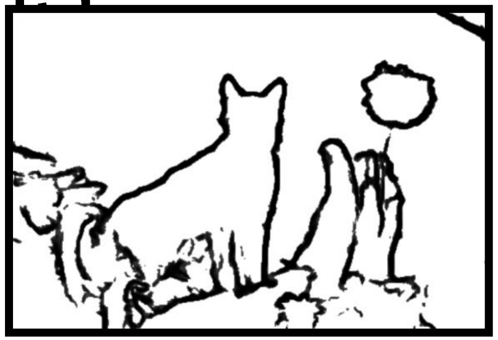
Output

Input



Output

Challenges —> Solutions



1. Output is high-dimensional, structured object

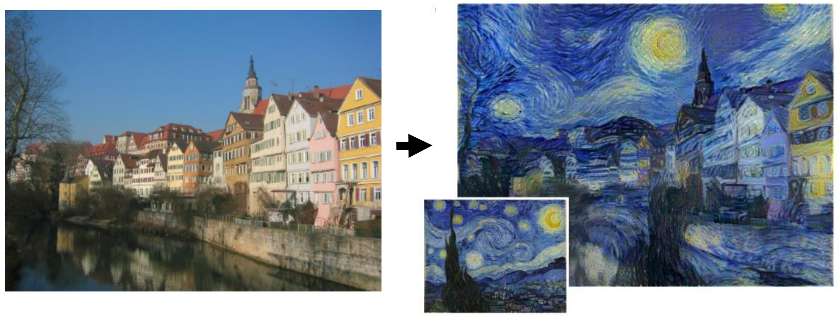
   —> Use a deep net, D, to analyze output!

“this small bird has a pink breast and crown…”

2. Uncertainty in mapping; many plausible outputs

   —> D only cares about “plausibility”, doesn’t hedge

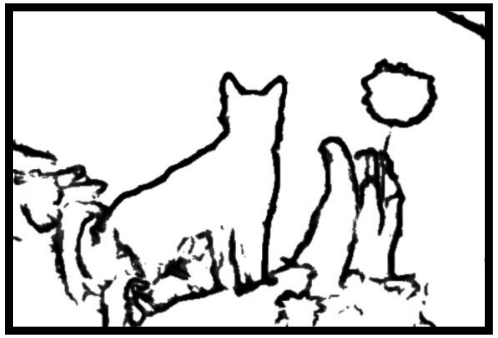
3. Lack of supervised training data



Challenges —> Solutions

1. Output is high-dimensional, structured object

   —> Use a deep net, D, to analyze output!

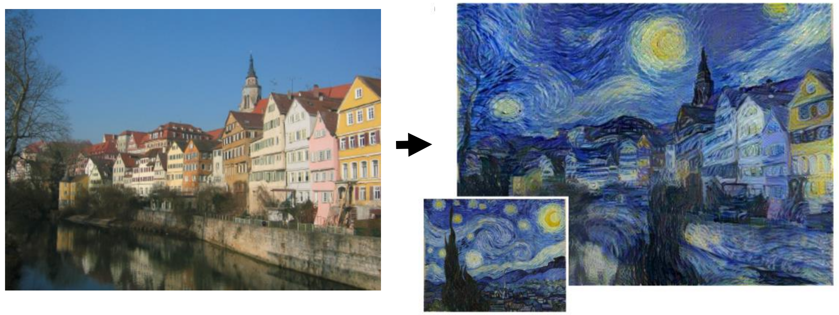


“this small bird has a pink breast and crown…”

2. Uncertainty in mapping; many plausible outputs

   —> D only cares about “plausibility”, doesn’t hedge

3. **Lack of supervised training data**



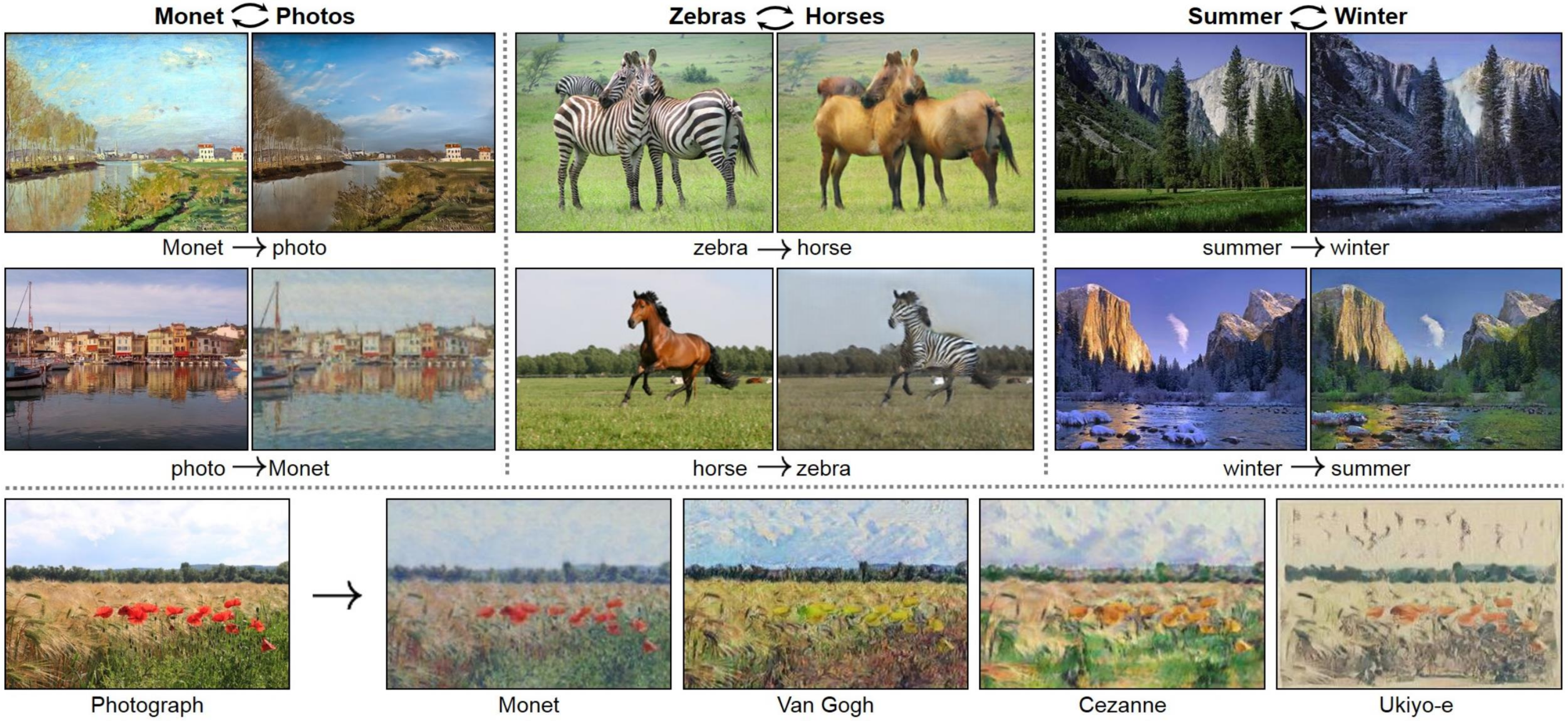
Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

Jun-Yan Zhu* **Taesung Park*** **Phillip Isola** **Alexei A. Efros**

UC Berkeley

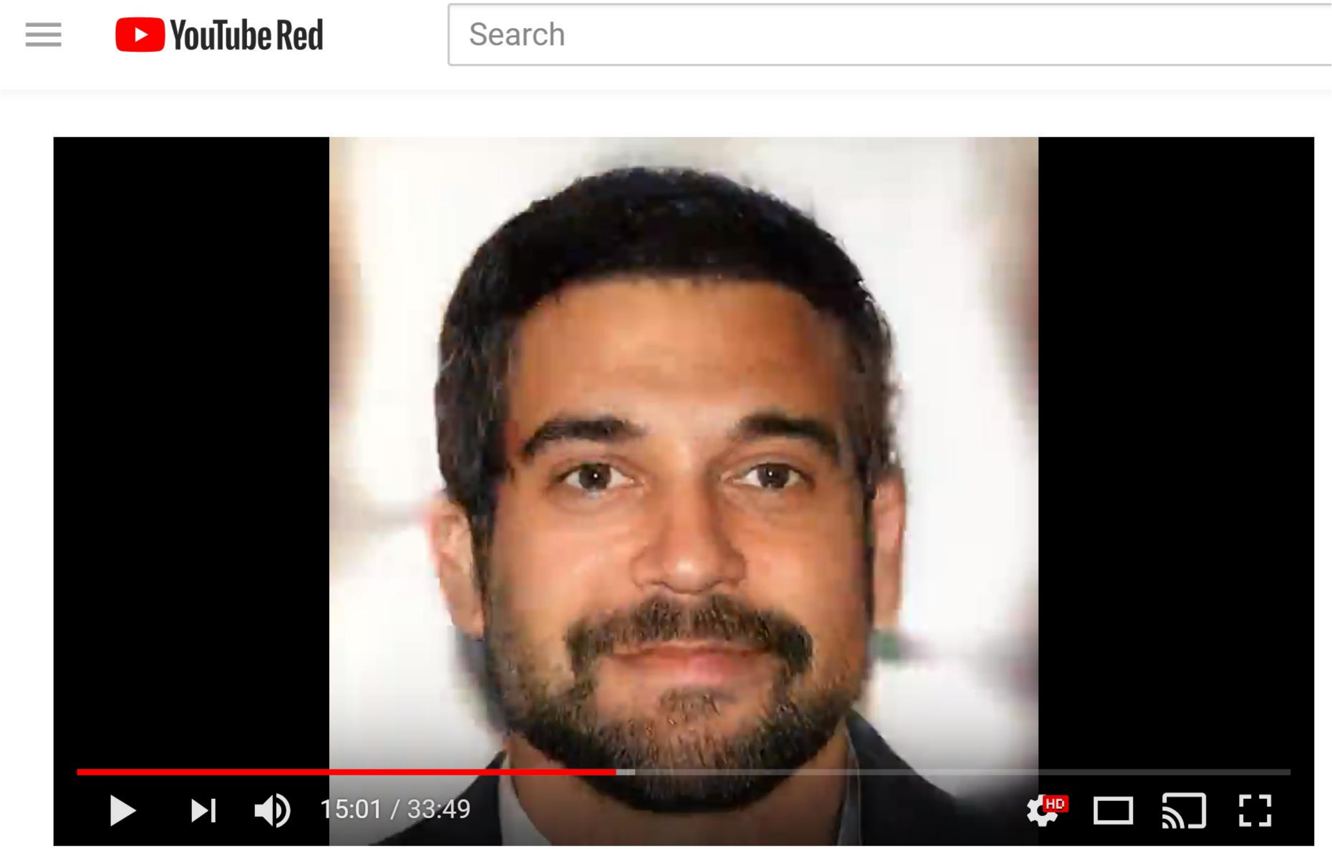
In ICCV 2017

[Paper] [Code (Torch)] [Code (PyTorch)]



https://junyanz.github.io/CycleGAN/

Progressive GANs



Fake Celebs with Progressive Growing of GANs

30 minutes of fake celebrities

https://www.youtube.com/watch?v=f8xSD4HO_8k

Questions?