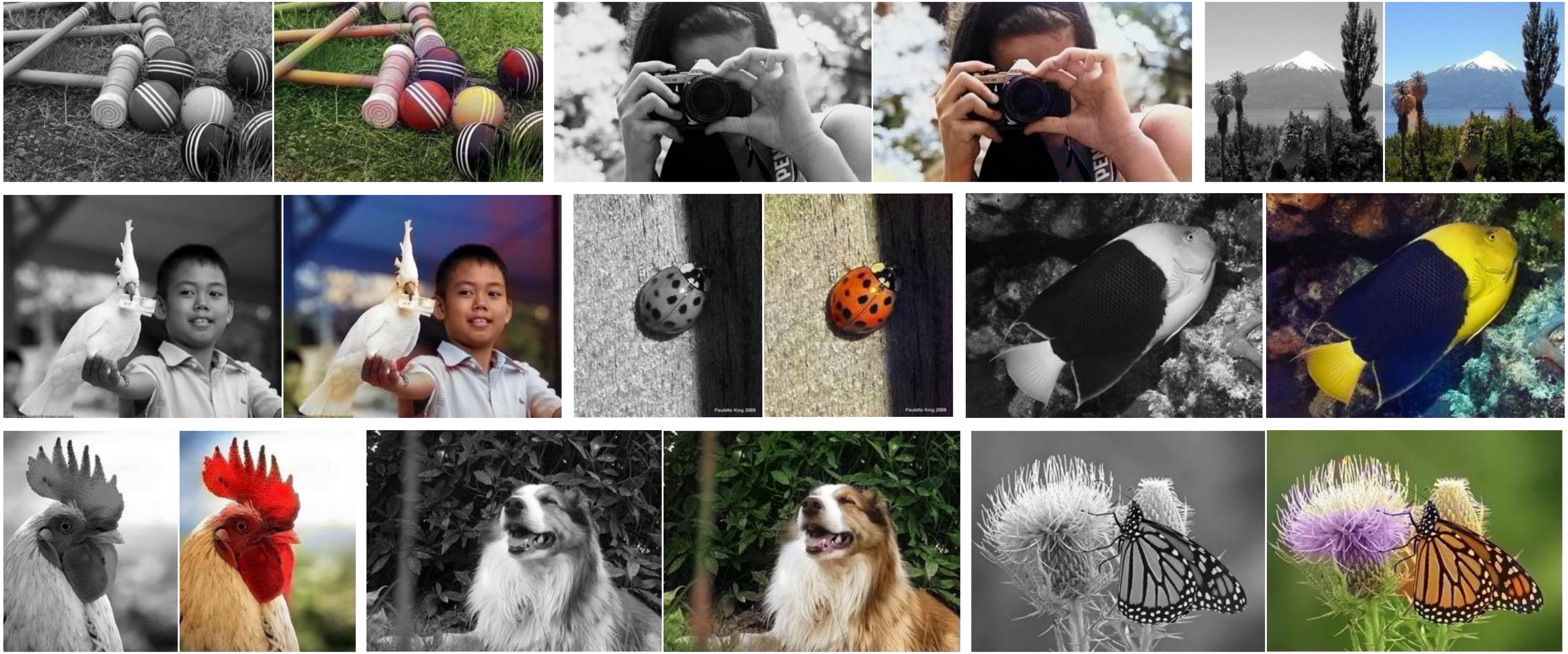
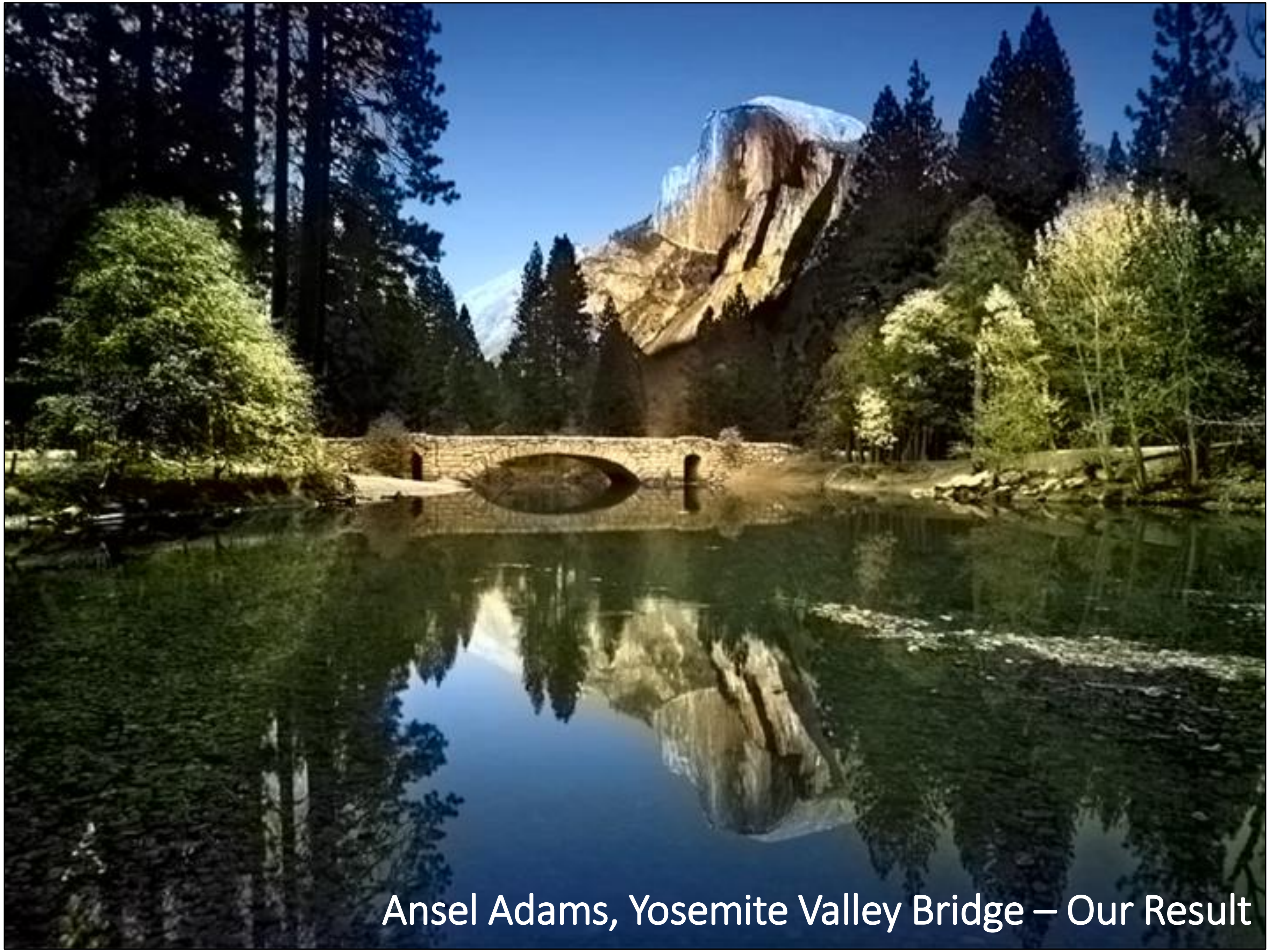


Colorful Image Colorization

Richard Zhang, Phillip Isola, Alexei (Alyosha) Efros

richzhang.github.io/colorization


Ansel Adams, Yosemite Valley Bridge


Ansel Adams, Yosemite Valley Bridge – Our Result

Grayscale image: *L* channel
$\mathbf{X} \in \mathbb{R}^{H \times W \times 1}$

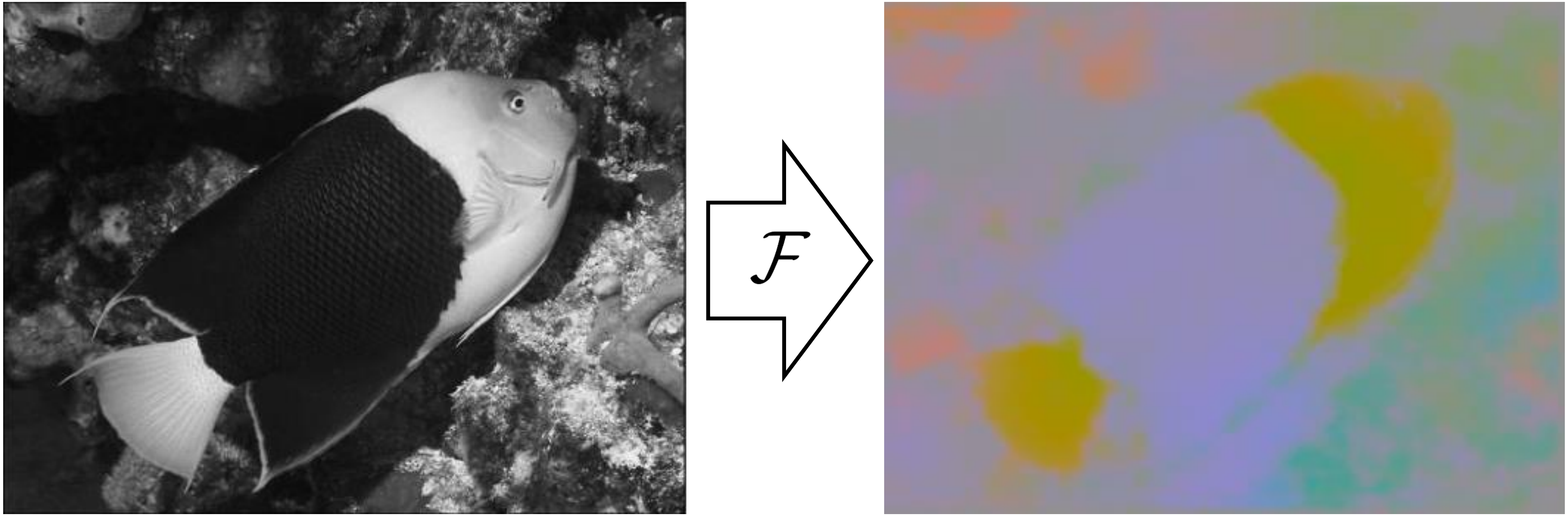
$\boxed{\boldsymbol{L}}$

Grayscale image: *L* channel
$\mathbf{X} \in \mathbb{R}^{H \times W \times 1}$

L

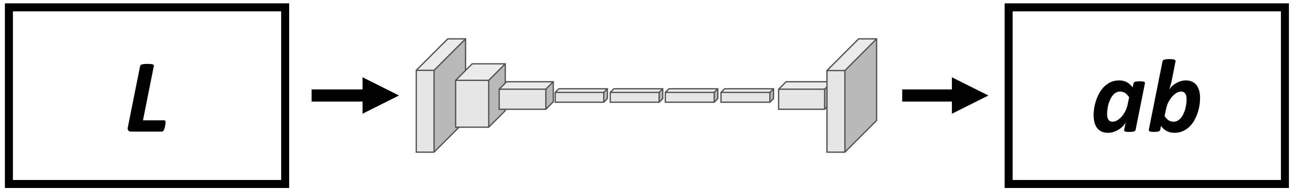
Color information: *ab* channels
$\widehat{\mathbf{Y}} \in \mathbb{R}^{H \times W \times 2}$

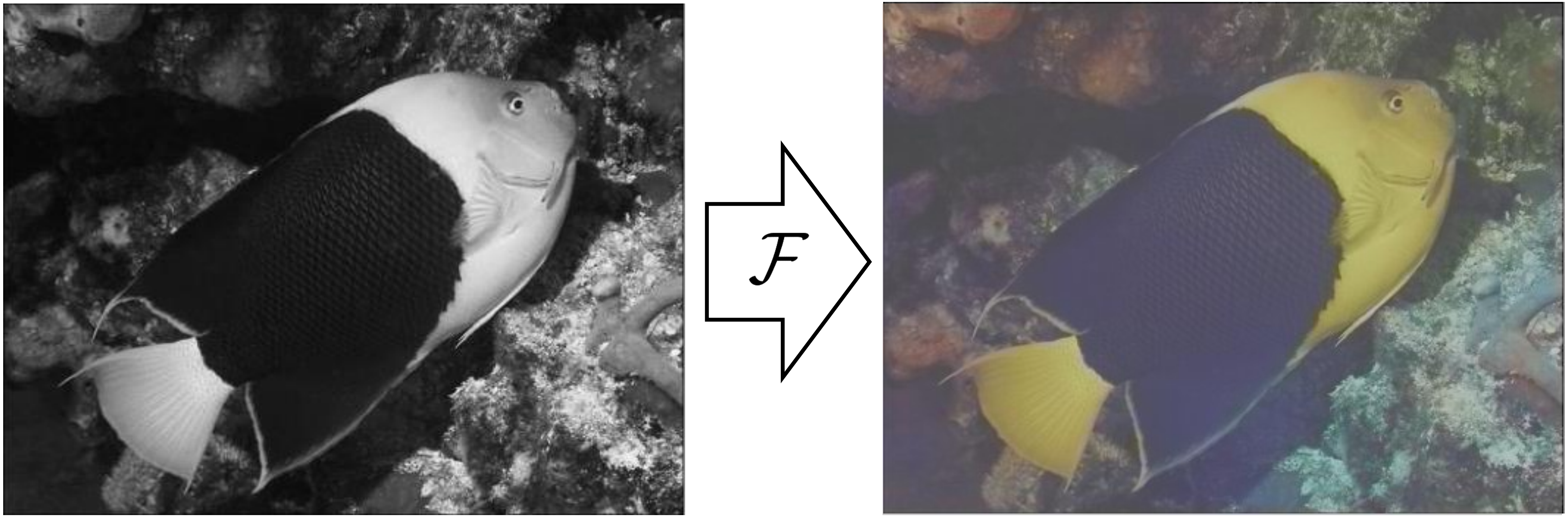
ab



Grayscale image: *L* channel
$$\mathbf{X} \in \mathbb{R}^{H \times W \times 1}$$

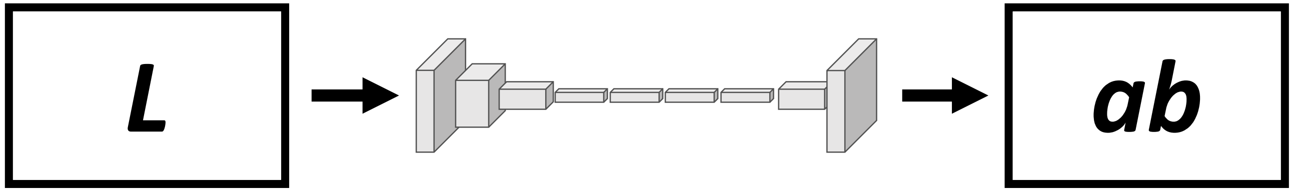
Color information: *ab* channels
$$\widehat{\mathbf{Y}} \in \mathbb{R}^{H \times W \times 2}$$

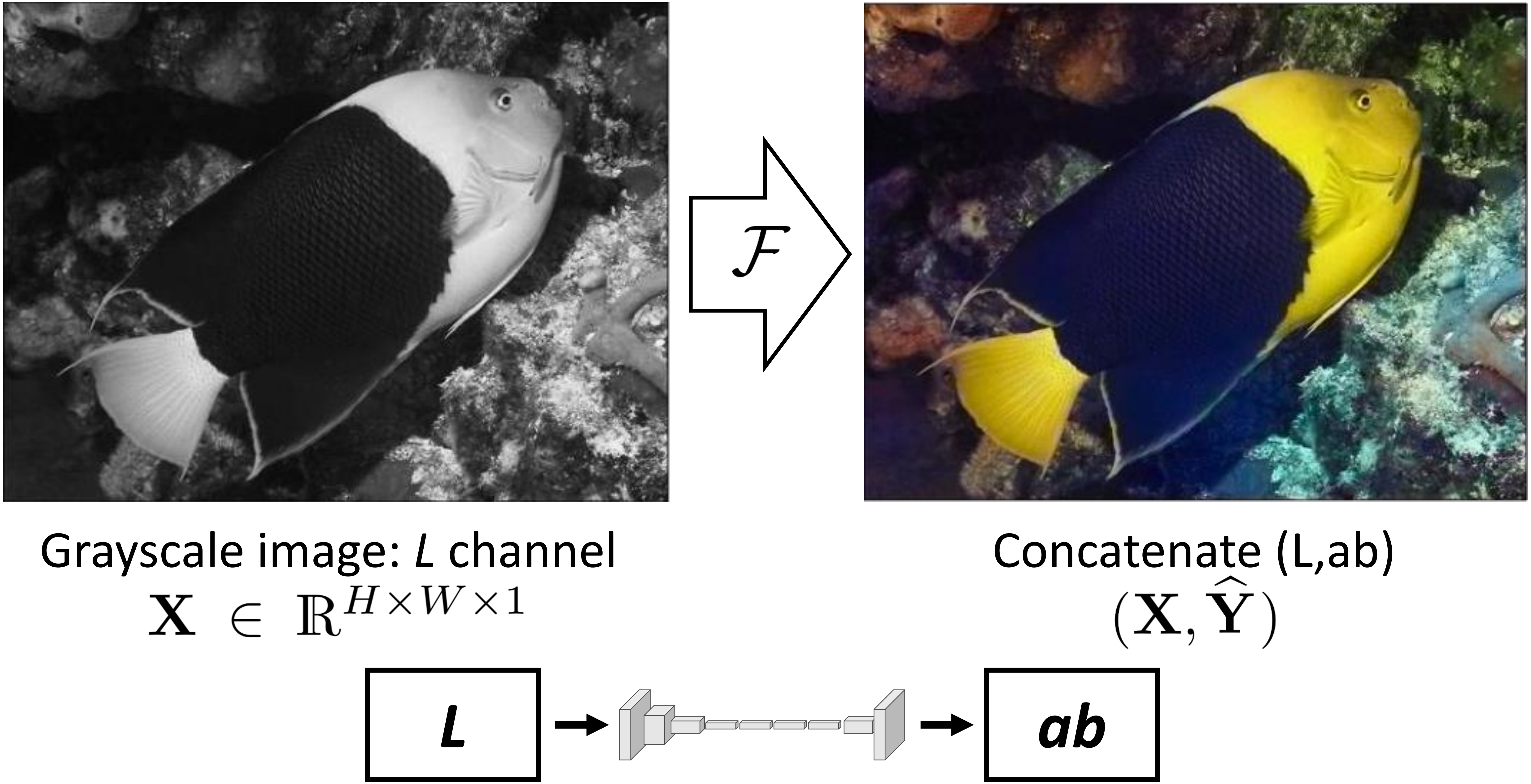


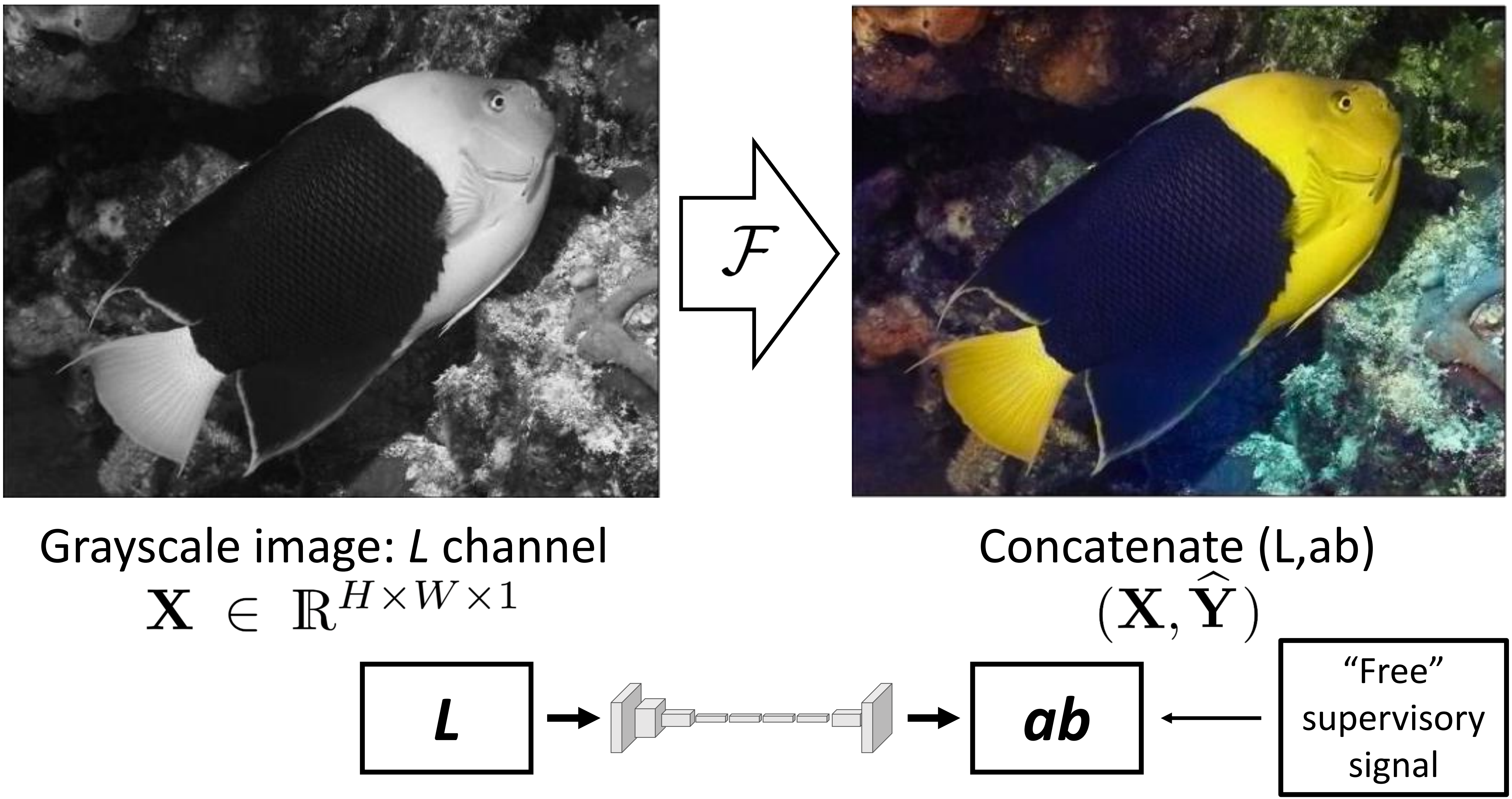


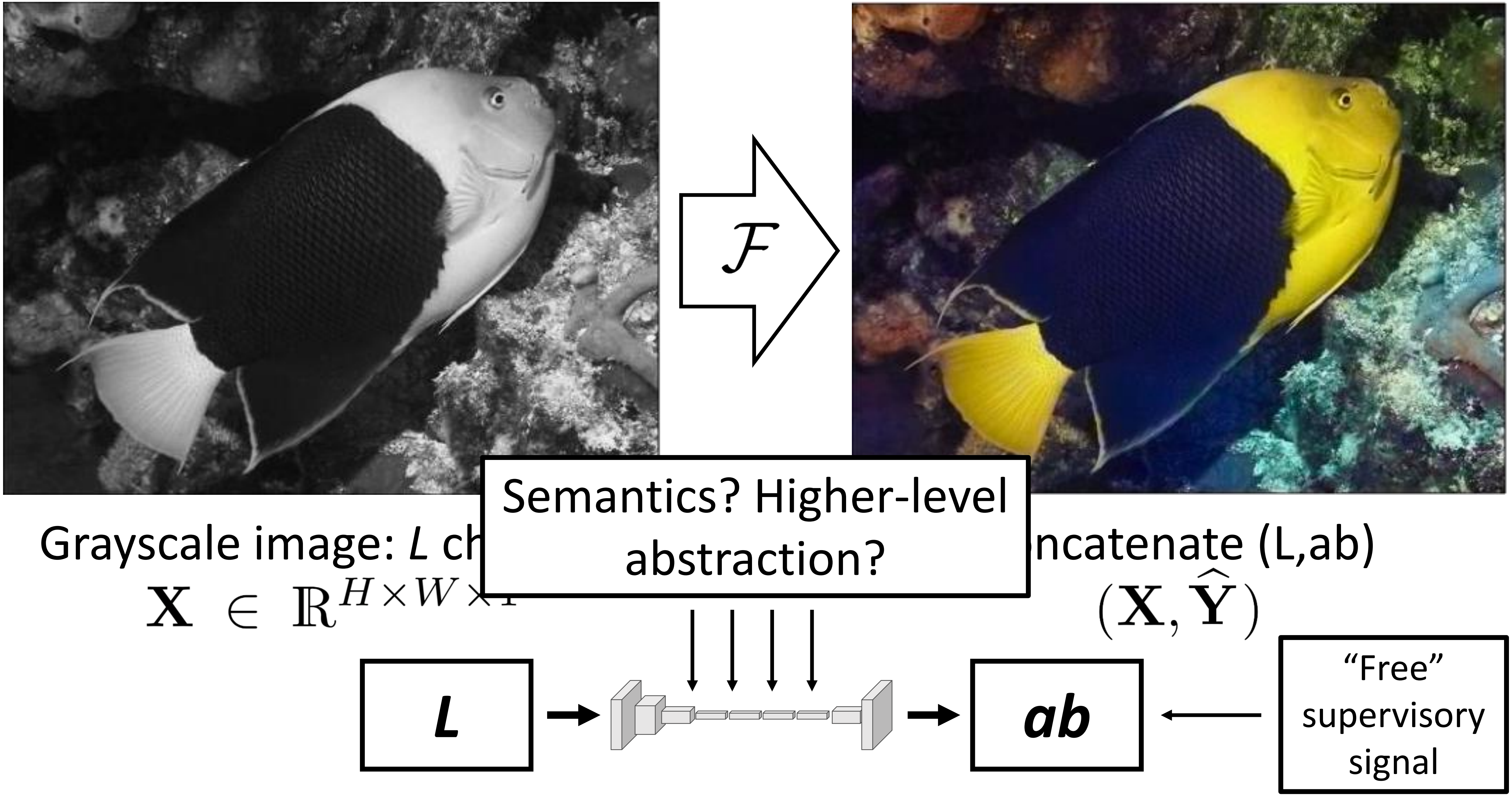
Grayscale image: *L* channel

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




$\mathcal{F}$
Grayscale image: *L* channel
$\mathbf{X} \in \mathbb{R}^{H \times W \times 1}$
Concatenate (L,ab)
$(\mathbf{X}, \widehat{\mathbf{Y}})$
L
ab


$\mathcal{F}$
Grayscale image: *L* channel
$\mathbf{X} \in \mathbb{R}^{H \times W \times 1}$
Concatenate (L,ab)
$(\mathbf{X}, \widehat{\mathbf{Y}})$
L
ab
"Free" supervisory signal



$\mathcal{F}$
Semantics? Higher-level abstraction?
Grayscale image: *L* ch
$\mathbf{X} \in \mathbb{R}^{H \times W \times}$
ncatenate (L,ab)
$(\mathbf{X}, \widehat{\mathbf{Y}})$
L
ab
"Free" supervisory signal

Inherent Ambiguity

Grayscale

Inherent Ambiguity

Our Output

Inherent Ambiguity

Our Output

Ground Truth

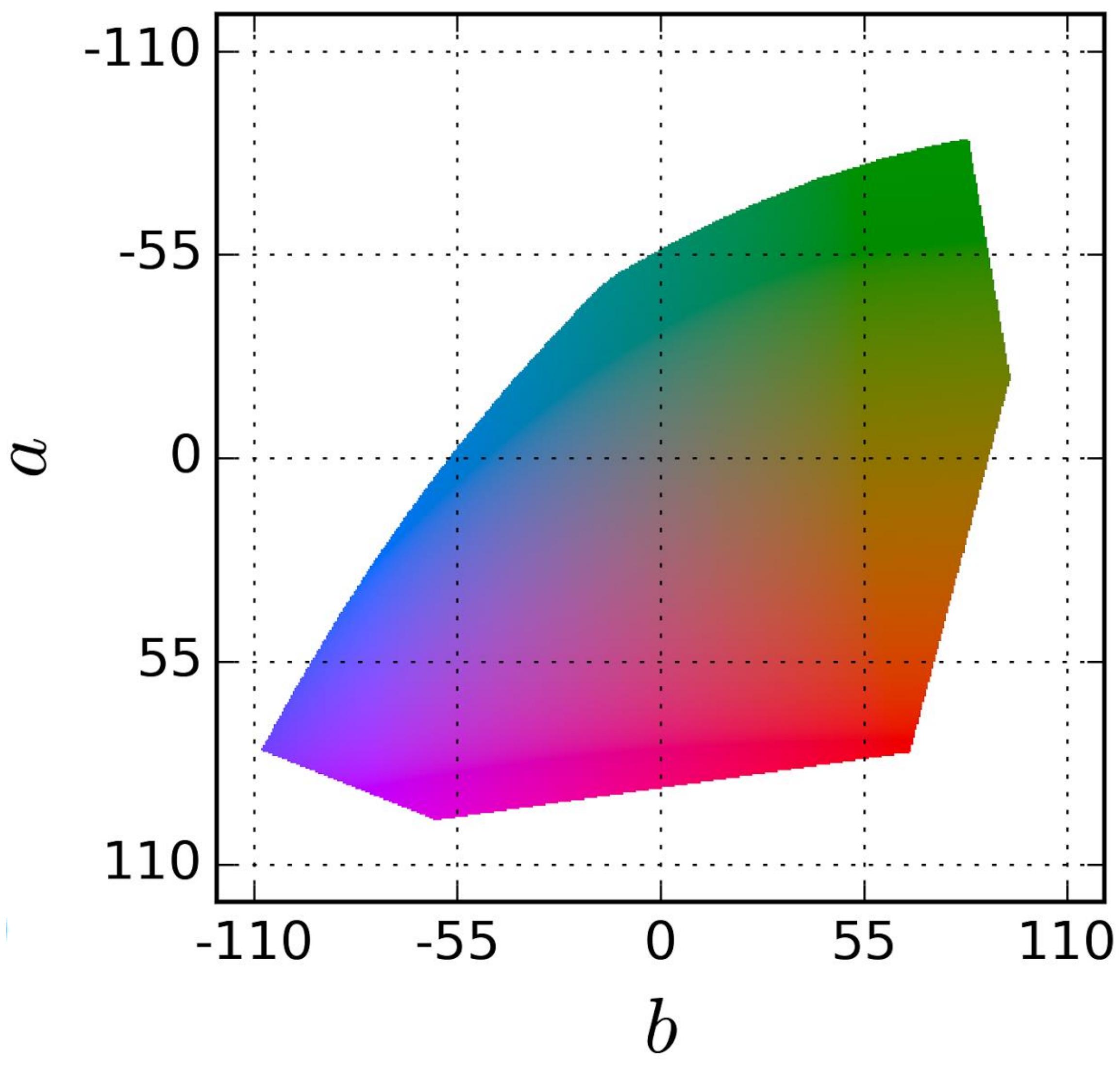
Better Loss Function

- Regression with L2 loss inadequate

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

Colors in *ab* space

(continuous)



Better Loss Function

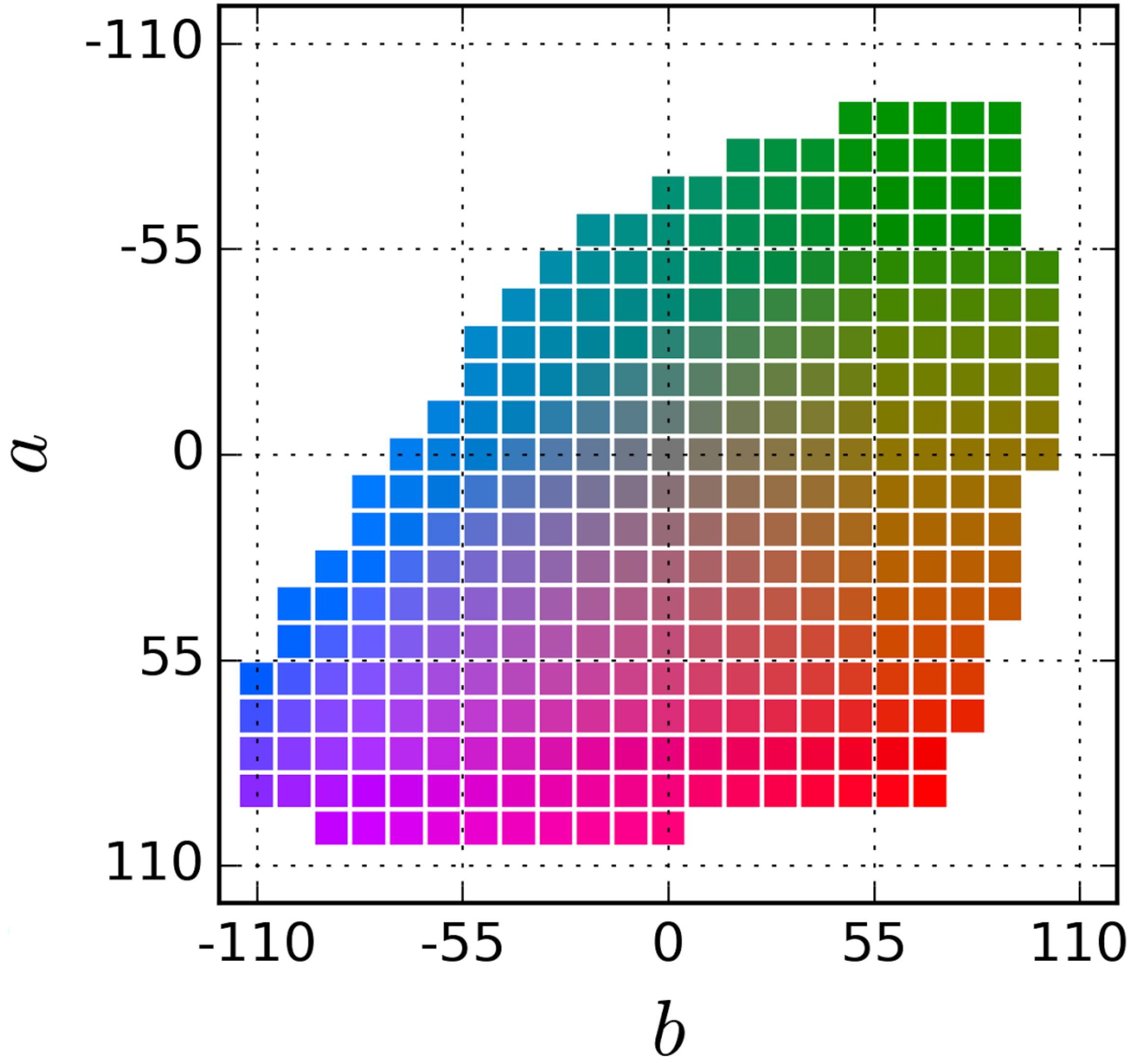
- Regression with L2 loss inadequate

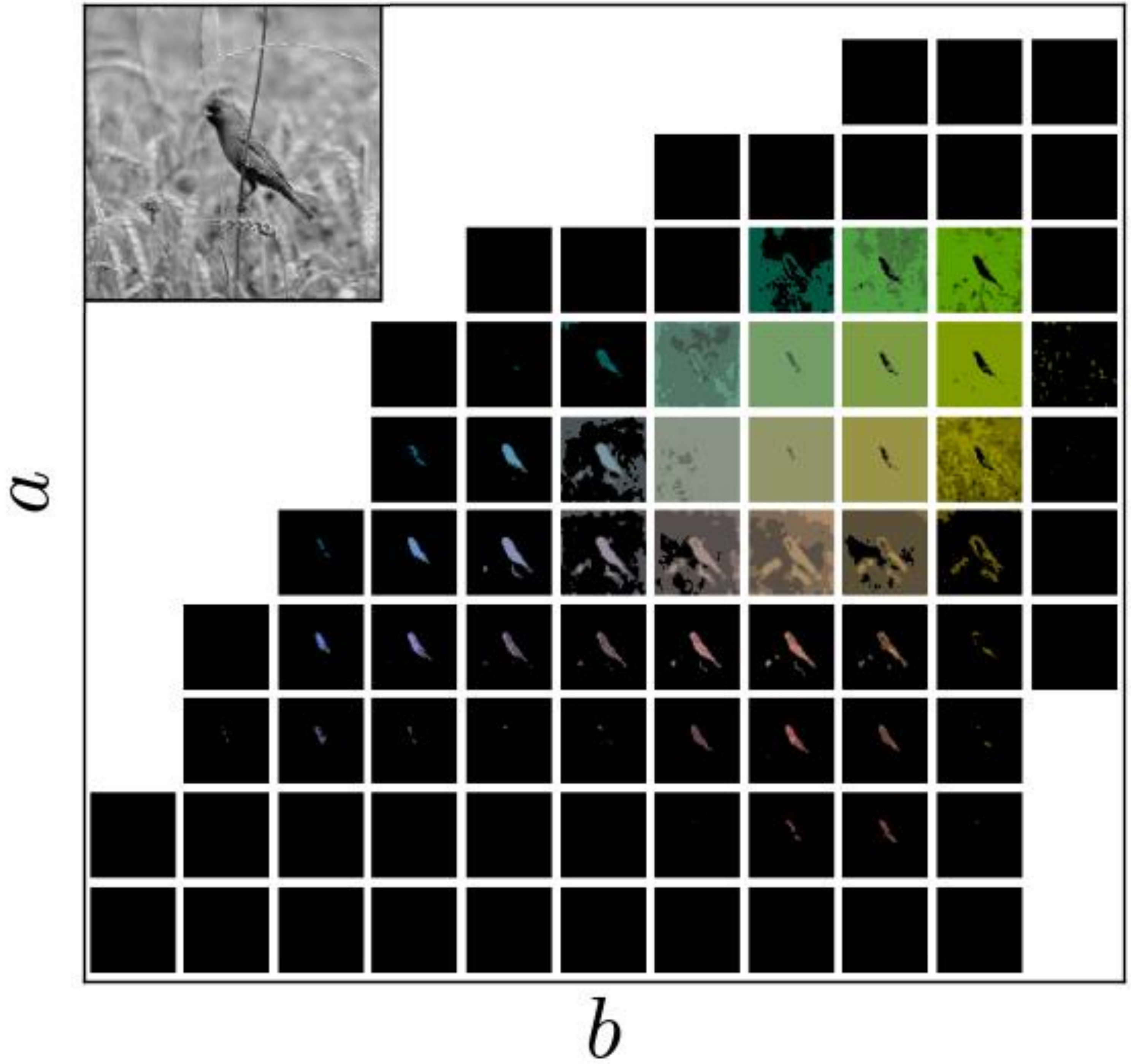
$$\mathrm{L}_2(\widehat{\mathbf{Y}}, \mathbf{Y}) = \frac{1}{2} \sum_{h,w} \|\mathbf{Y}_{h,w} - \widehat{\mathbf{Y}}_{h,w}\|_2^2$$

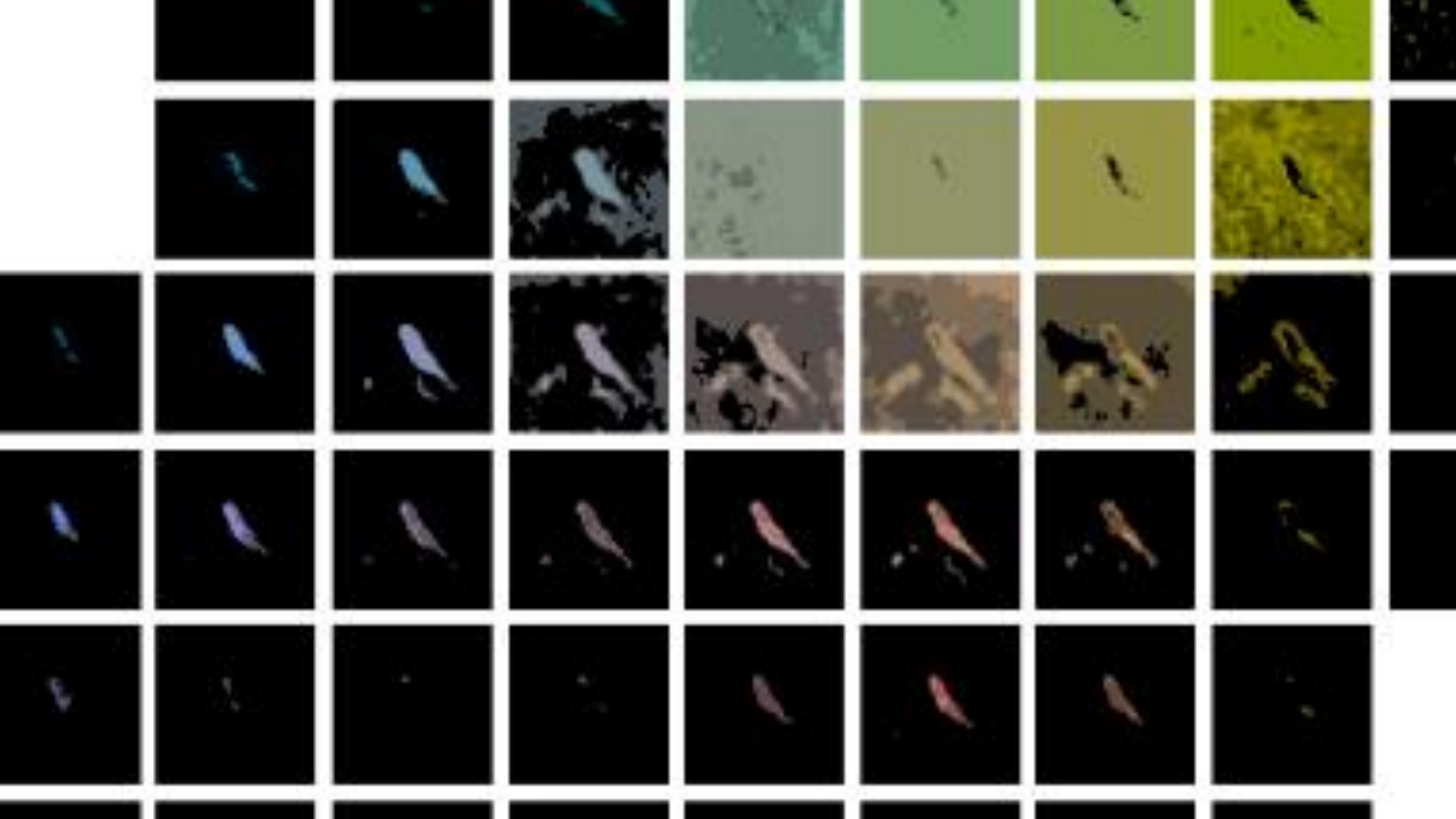
- Use **multinomial classification**

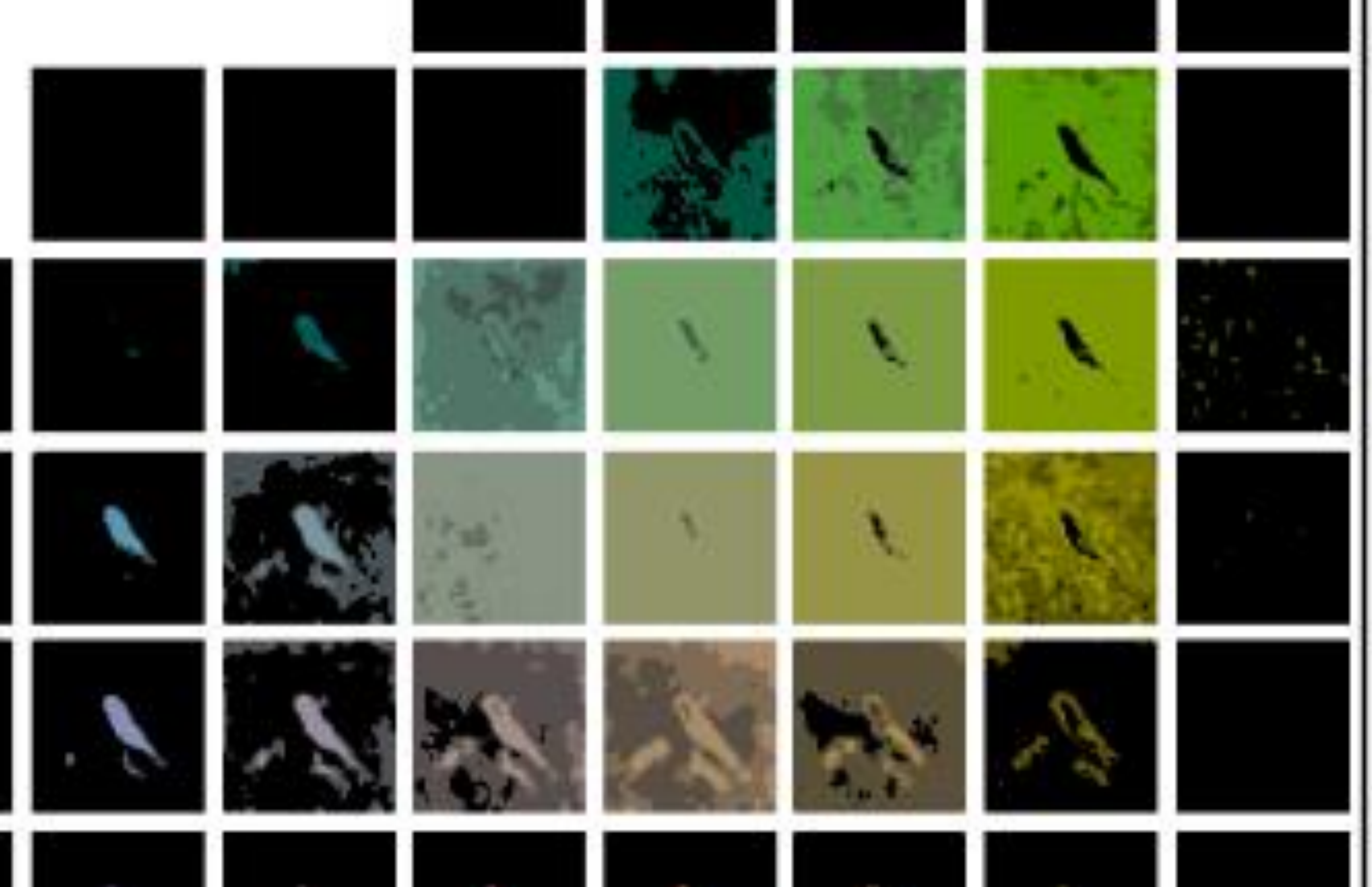
$$\mathrm{L}(\widehat{\mathbf{Z}}, \mathbf{Z}) = -\frac{1}{HW} \sum_{h,w} \sum_{q} \mathbf{Z}_{h,w,q} \log(\widehat{\mathbf{Z}}_{h,w,q})$$

Colors in *ab* space
(discrete)




a
b





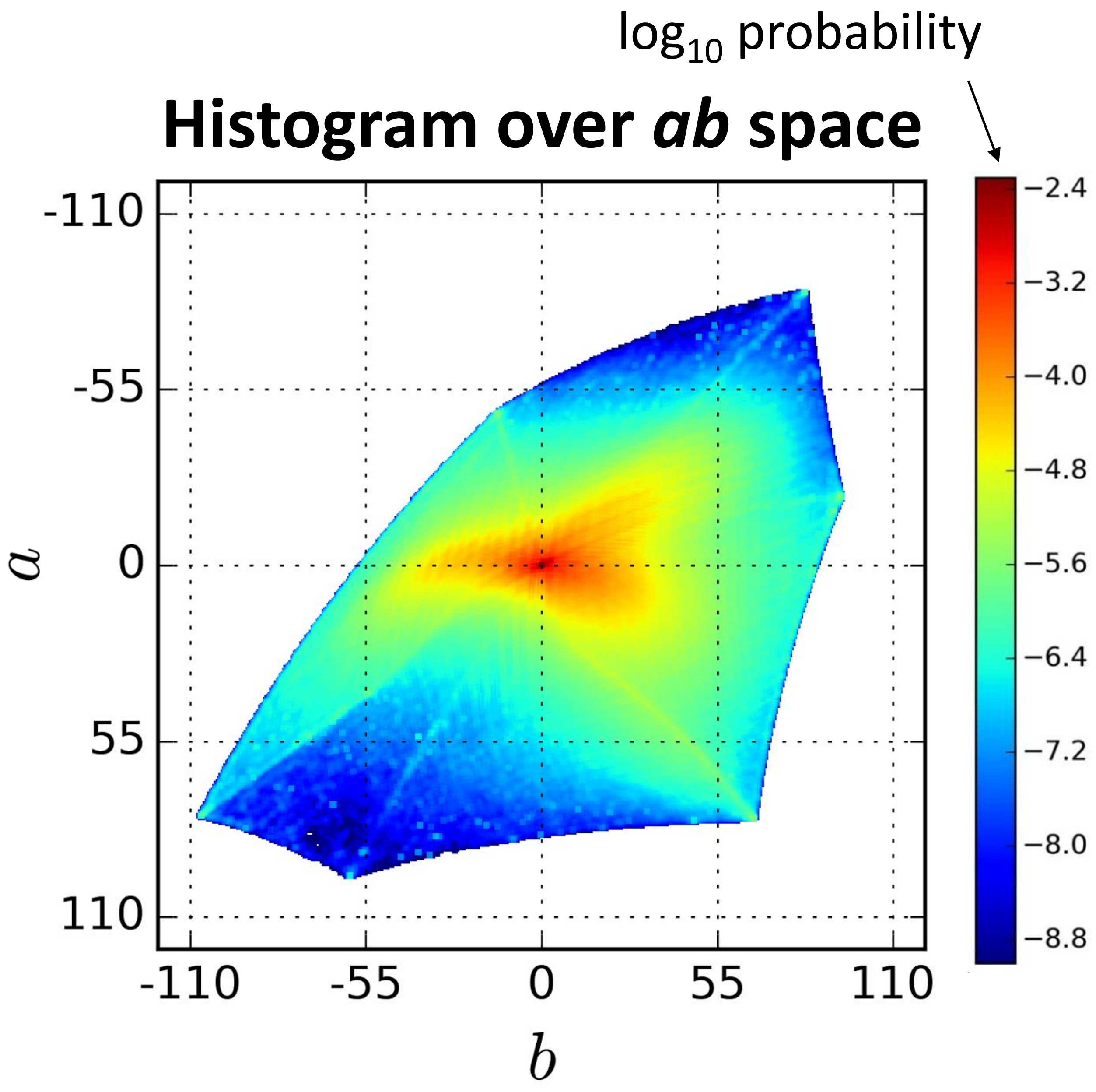
Better Loss Function

- 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 **multinomial classification**

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



Better Loss Function

- Regression with L2 loss inadequate

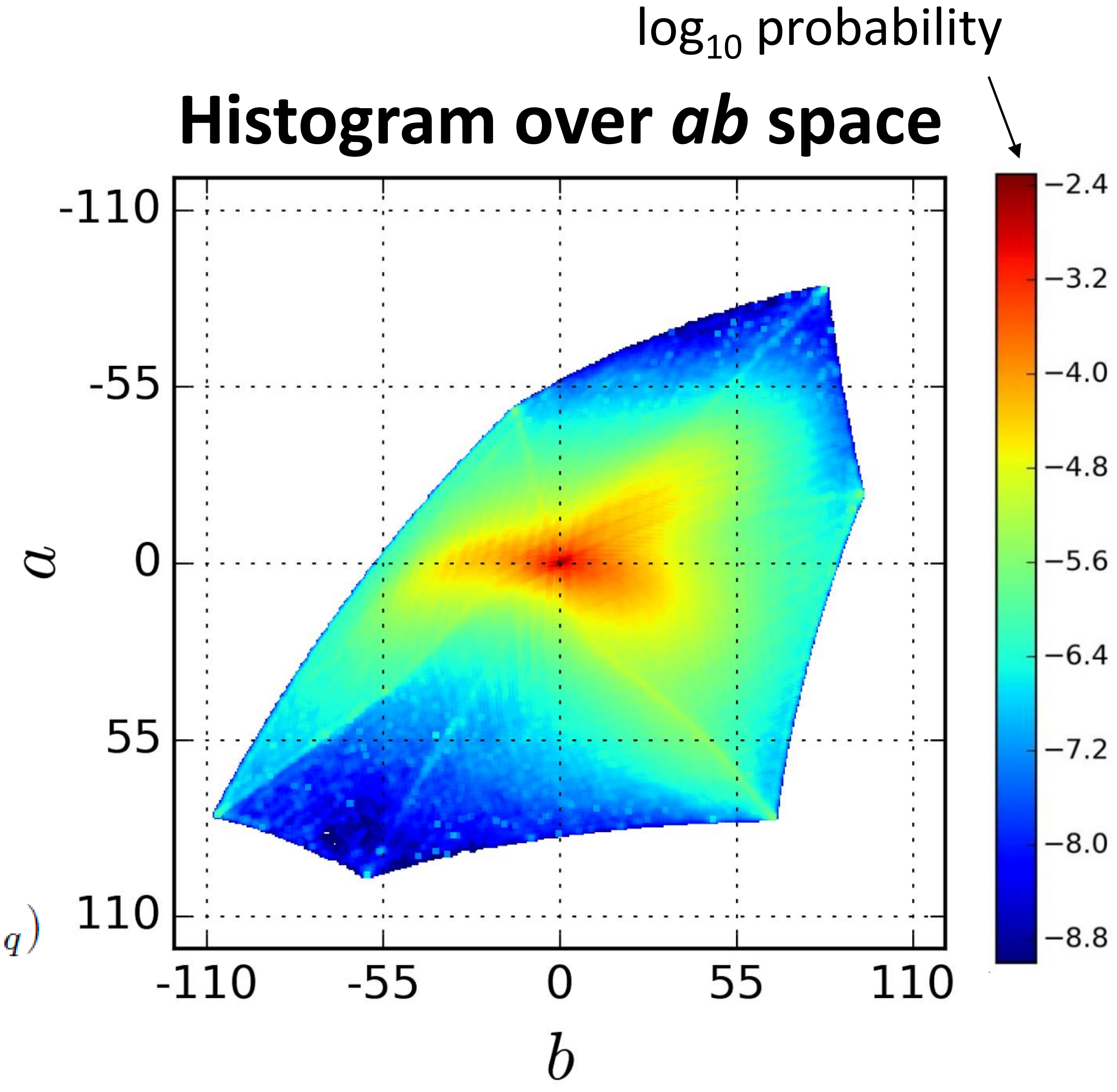
$$\mathrm{L}_2(\widehat{\mathbf{Y}}, \mathbf{Y}) = \frac{1}{2} \sum_{h,w} \|\mathbf{Y}_{h,w} - \widehat{\mathbf{Y}}_{h,w}\|_2^2$$

- Use **multinomial classification**

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

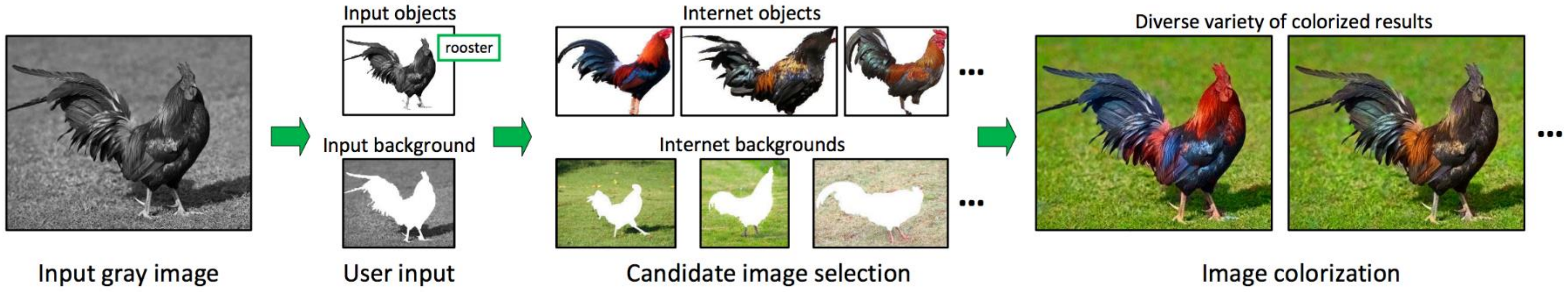
- **Class rebalancing** to encourage learning of *rare* colors

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



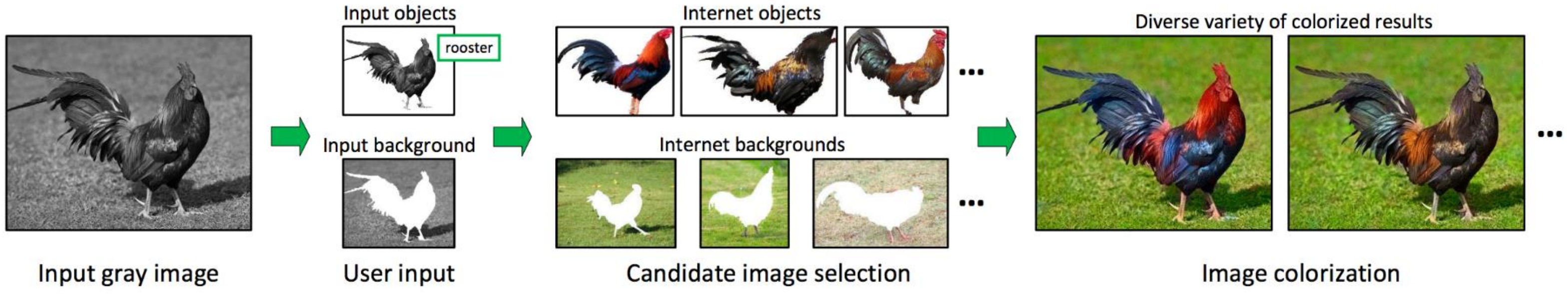
Non-parametric

Hertzmann et al. In SIGGRAPH, 2001.
Welsh et al. In TOG, 2002.
Irony et al. In Eurographics, 2005.
Liu et al. In TOG, 2008.
Chia et al. In ACM 2011.
Gupta et al. In ACM, 2012.



Non-parametric

Hertzmann et al. In SIGGRAPH, 2001.
Welsh et al. In TOG, 2002.
Irony et al. In Eurographics, 2005.
Liu et al. In TOG, 2008.
Chia et al. In ACM 2011.
Gupta et al. In ACM, 2012.

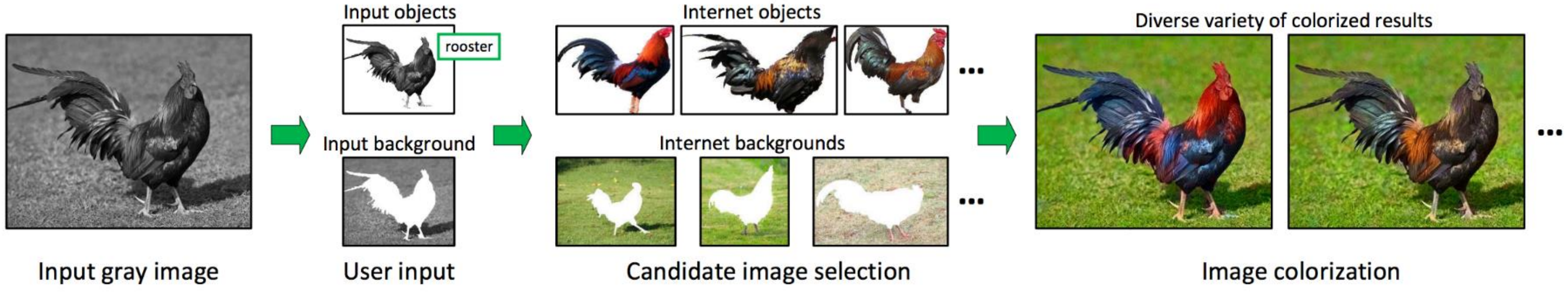


Parametric

L2 Regression

Non-parametric

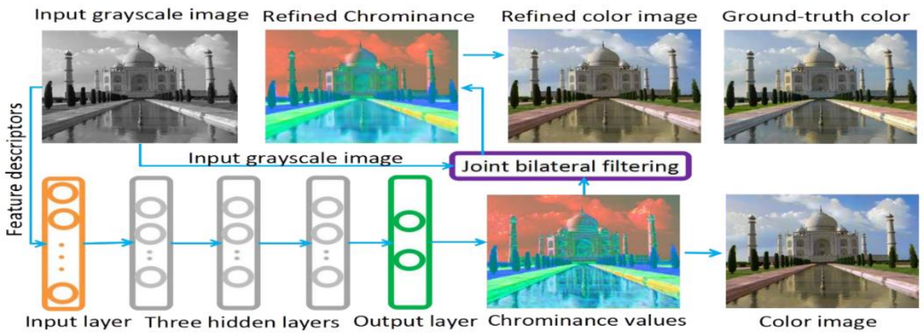
Hertzmann et al. In SIGGRAPH, 2001.
Welsh et al. In TOG, 2002.
Irony et al. In Eurographics, 2005.
Liu et al. In TOG, 2008.
Chia et al. In ACM 2011.
Gupta et al. In ACM, 2012.



Parametric

L2 Regression

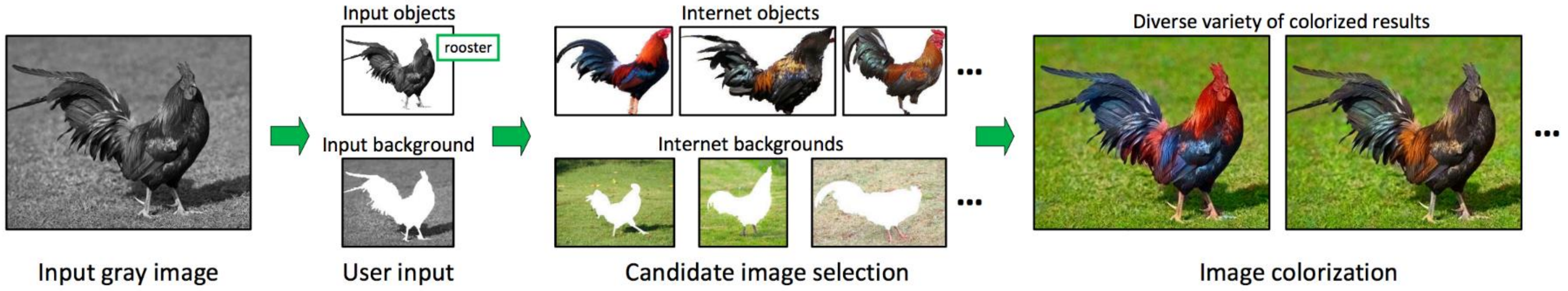
Hand-engineered Features



Deshpande et al. Cheng et al. In ICCV 2015.

Non-parametric

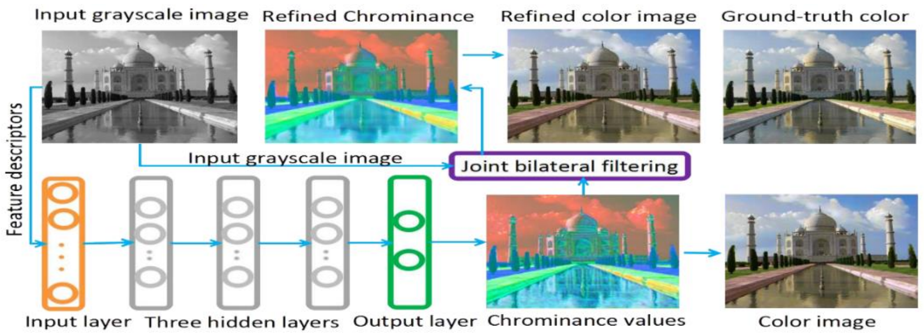
Hertzmann et al. In SIGGRAPH, 2001.
Welsh et al. In TOG, 2002.
Irony et al. In Eurographics, 2005.
Liu et al. In TOG, 2008.
Chia et al. In ACM 2011.
Gupta et al. In ACM, 2012.



Parametric

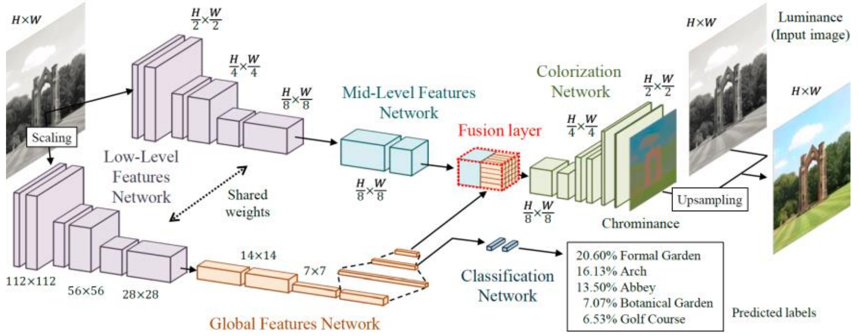
Hand-engineered Features

L2 Regression



Deshpande et al. Cheng et al. In ICCV 2015.

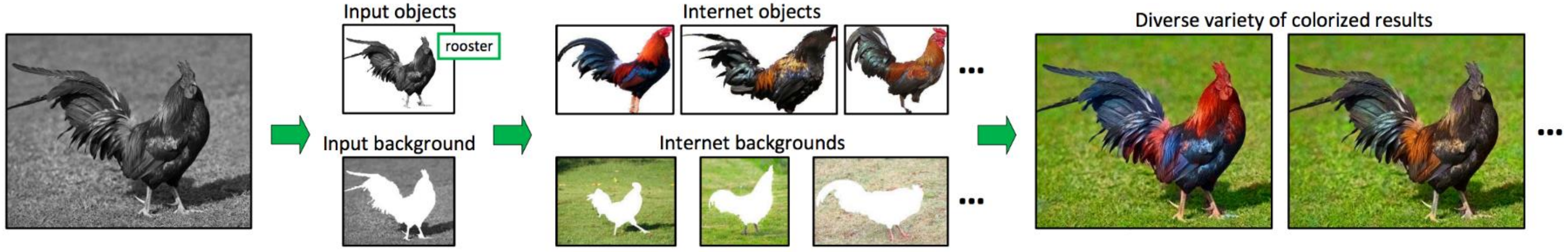
Deep Networks



Dahl. Jan 2016. Iizuka et al. In SIGGRAPH, 2016.

Non-parametric

Hertzmann et al. In SIGGRAPH, 2001.
Welsh et al. In TOG, 2002.
Irony et al. In Eurographics, 2005.
Liu et al. In TOG, 2008.
Chia et al. In ACM 2011.
Gupta et al. In ACM, 2012.

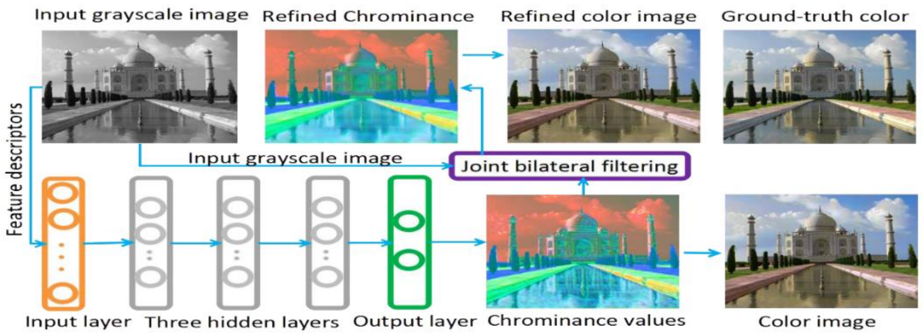


Parametric

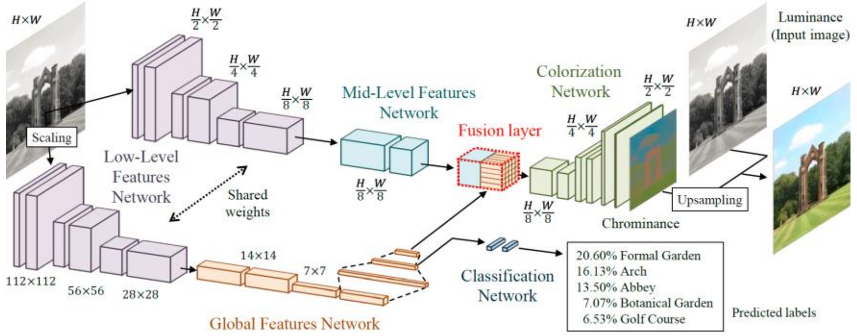
Hand-engineered Features

Deep Networks

L2 Regression

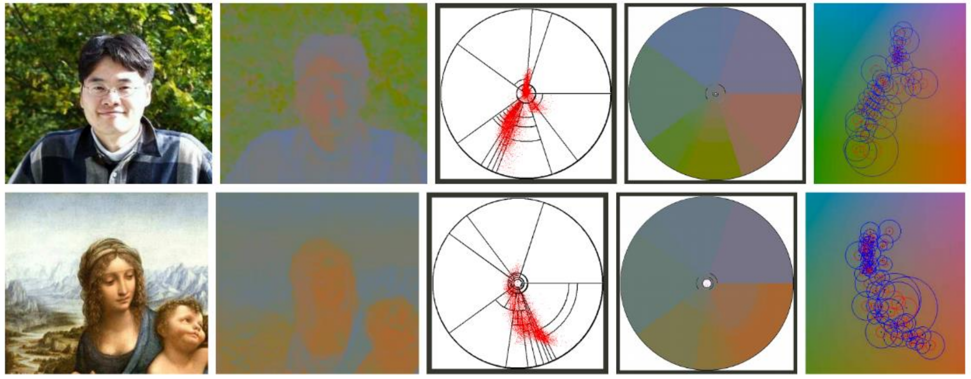


Deshpande et al. Cheng et al. In ICCV 2015.



Dahl. Jan 2016. Iizuka et al. In SIGGRAPH, 2016.

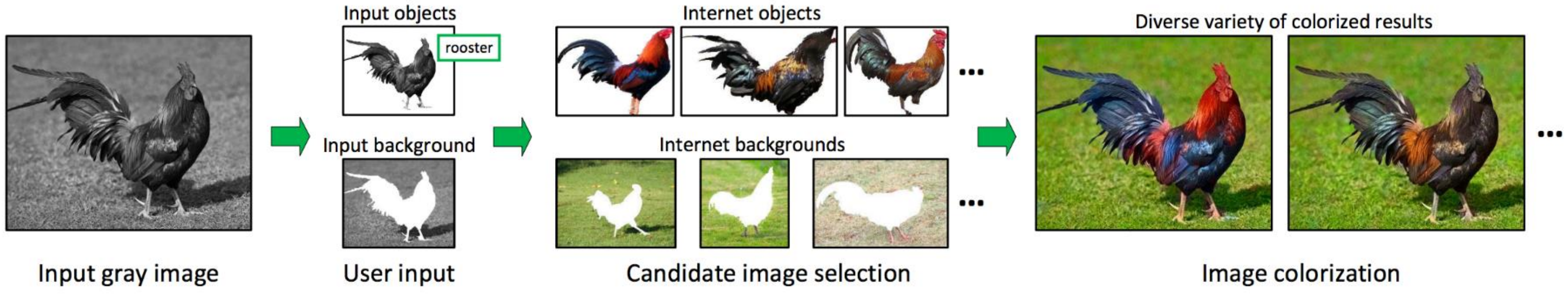
Classification



Charpiat et al. In ECCV 2008.

Non-parametric

Hertzmann et al. In SIGGRAPH, 2001.
Welsh et al. In TOG, 2002.
Irony et al. In Eurographics, 2005.
Liu et al. In TOG, 2008.
Chia et al. In ACM 2011.
Gupta et al. In ACM, 2012.

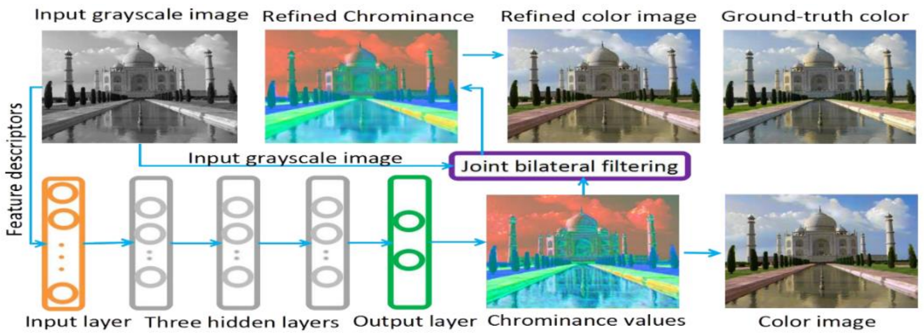


Parametric

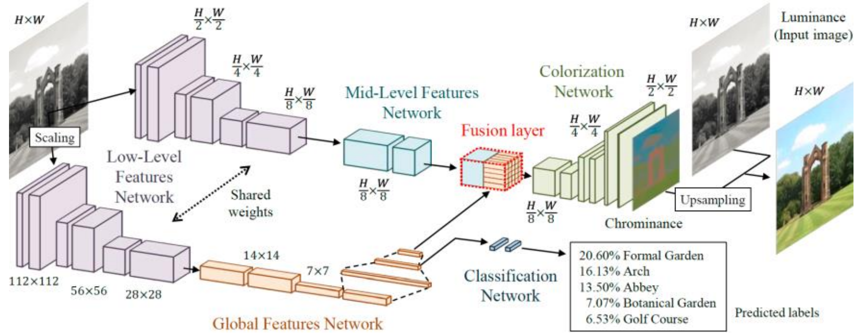
Hand-engineered Features

Deep Networks

L2 Regression

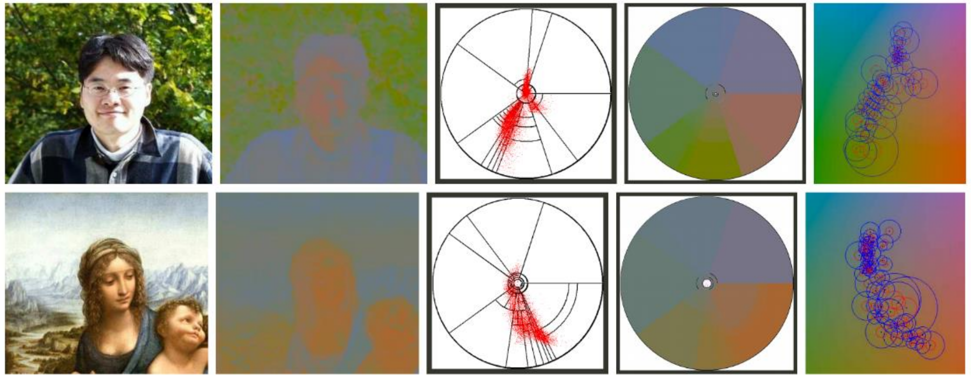


Deshpande et al. Cheng et al. In ICCV 2015.

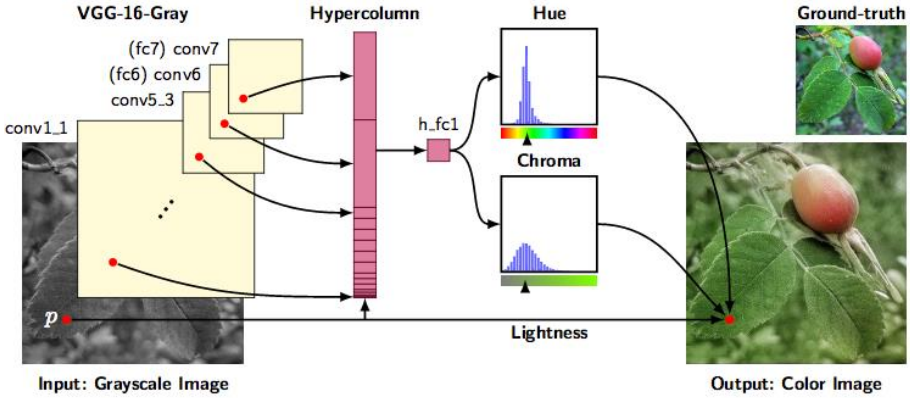


Dahl. Jan 2016. Iizuka et al. In SIGGRAPH, 2016.

Classification



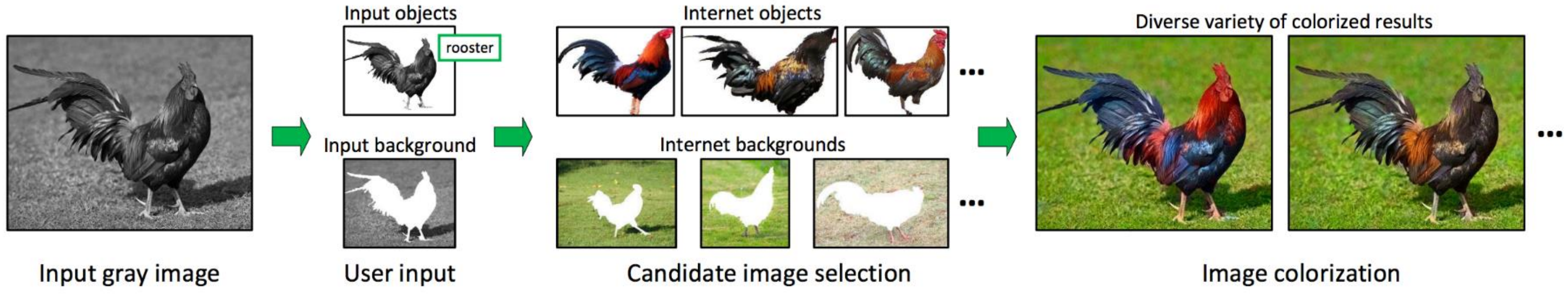
Charpiat et al. In ECCV 2008.



Larsson et al. In ECCV 2016. [Concurrent]

Non-parametric

Hertzmann et al. In SIGGRAPH, 2001.
Welsh et al. In TOG, 2002.
Irony et al. In Eurographics, 2005.
Liu et al. In TOG, 2008.
Chia et al. In ACM 2011.
Gupta et al. In ACM, 2012.

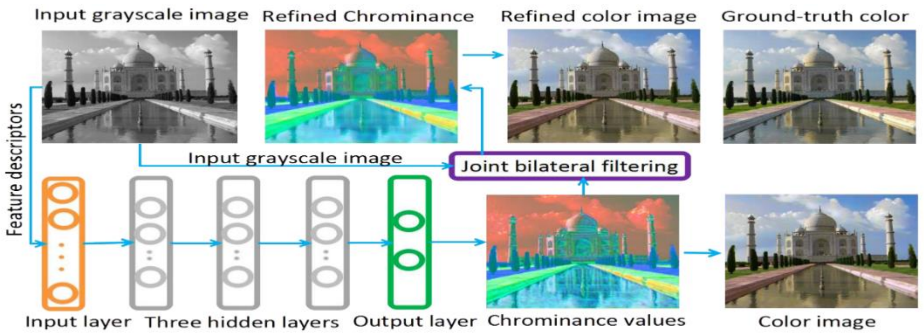


Parametric

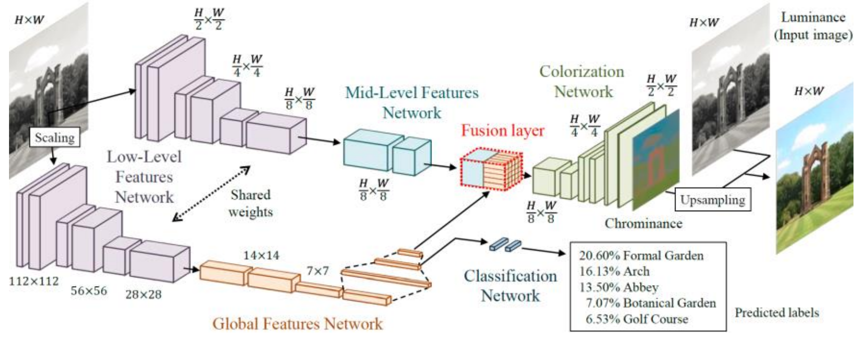
Hand-engineered Features

Deep Networks

L2 Regression

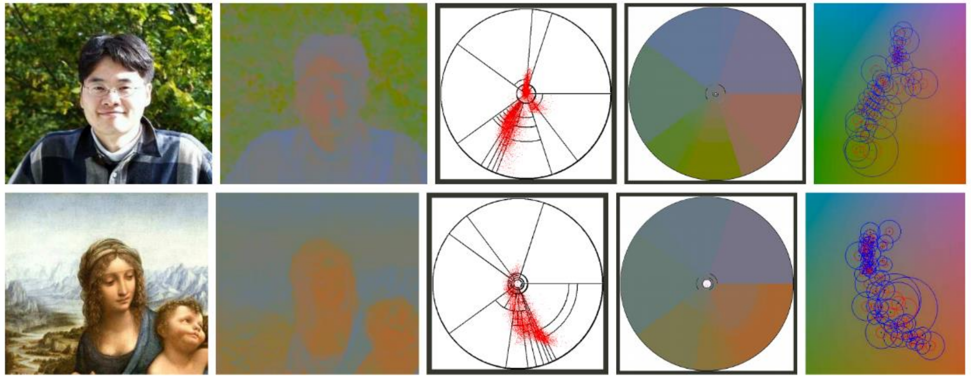


Deshpande et al. Cheng et al. In ICCV 2015.

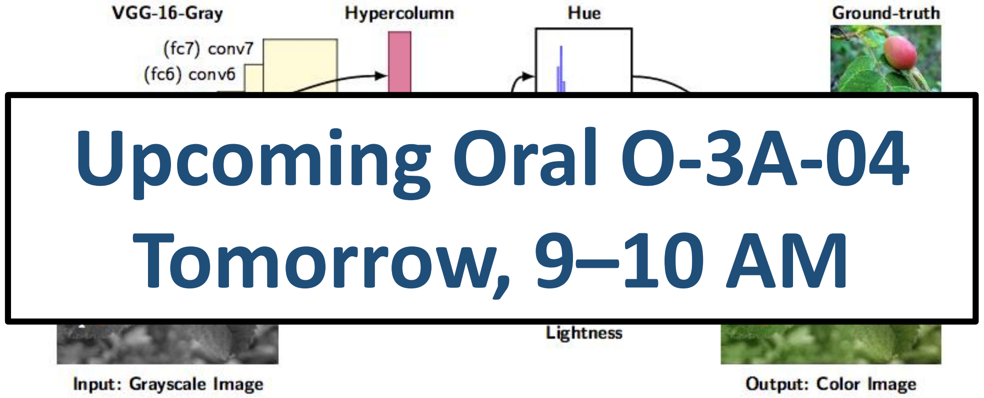


Dahl. Jan 2016. Iizuka et al. In SIGGRAPH, 2016.

Classification



Charpiat et al. In ECCV 2008.



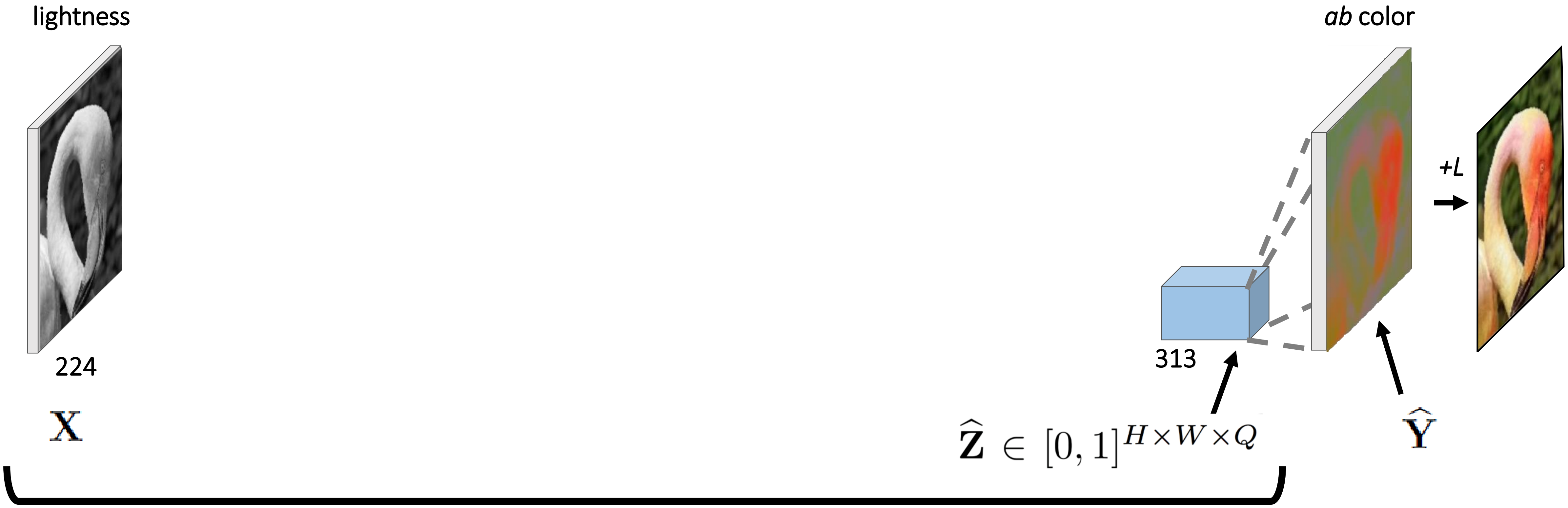
Upcoming Oral O-3A-04
Tomorrow, 9–10 AM

Larsson et al. In ECCV 2016. [Concurrent]

Network Architecture

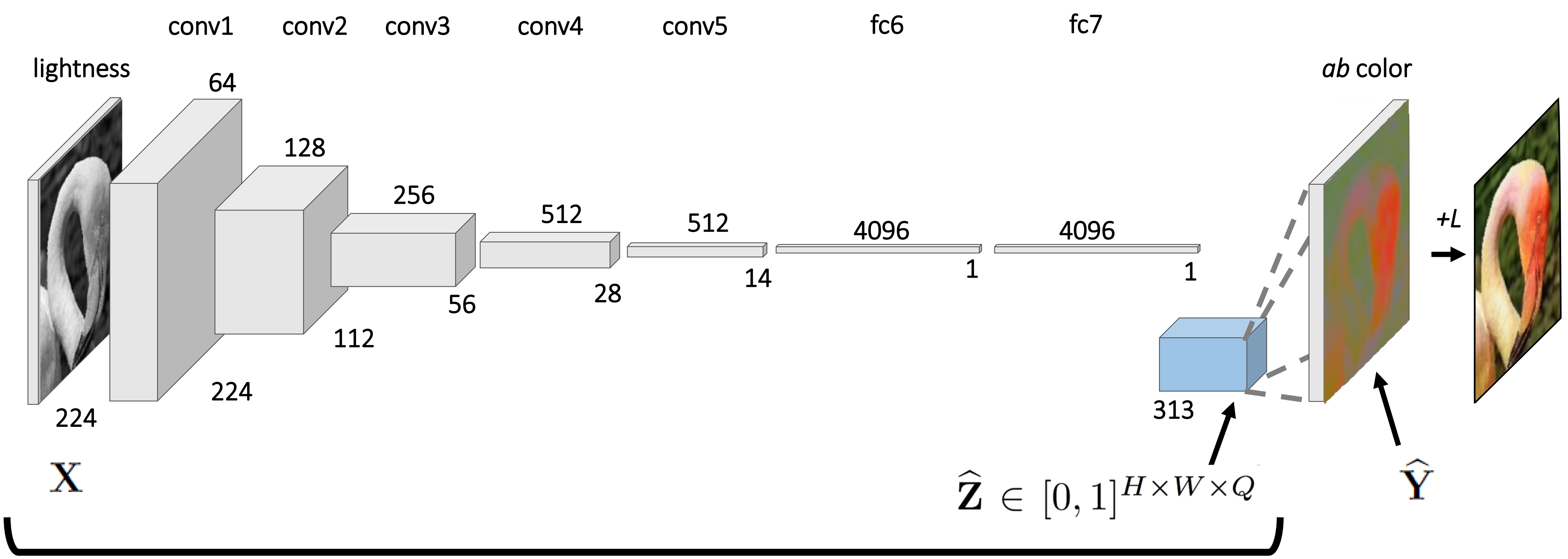
$$\widehat{\mathbf{Y}} = \mathcal{F}(\mathbf{X})$$

Network Architecture



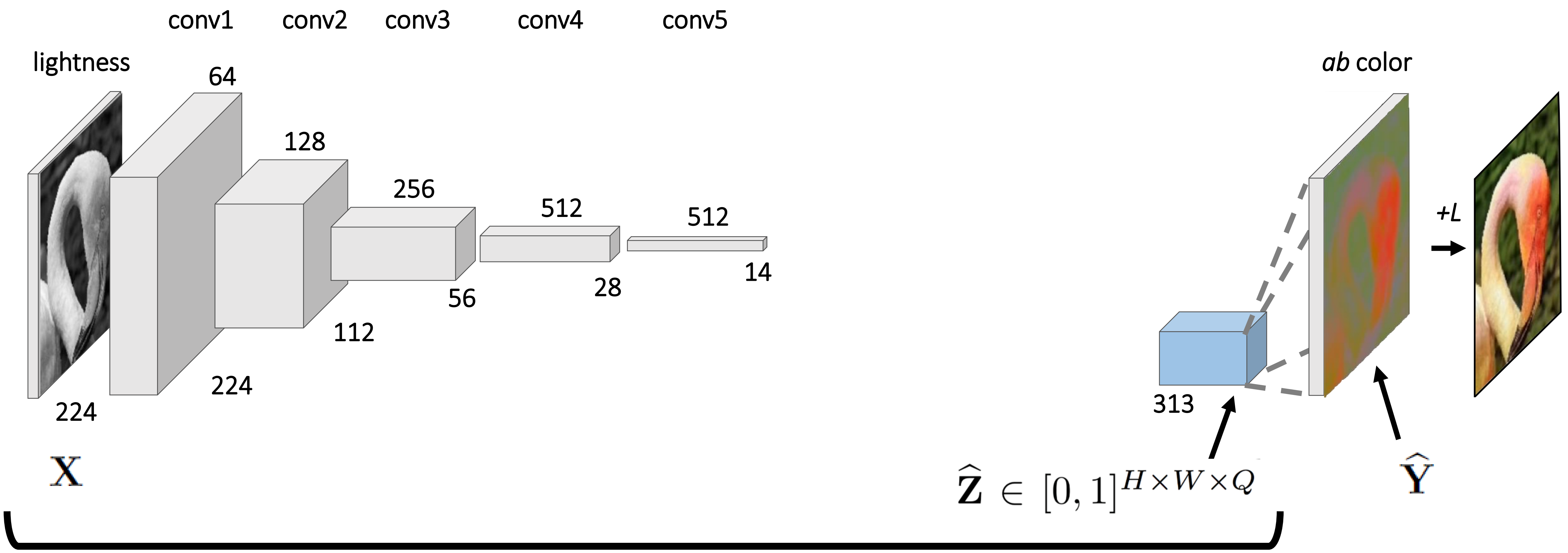
$$\widehat{\mathbf{Z}} = \mathcal{G}(\mathbf{X})$$

Network Architecture



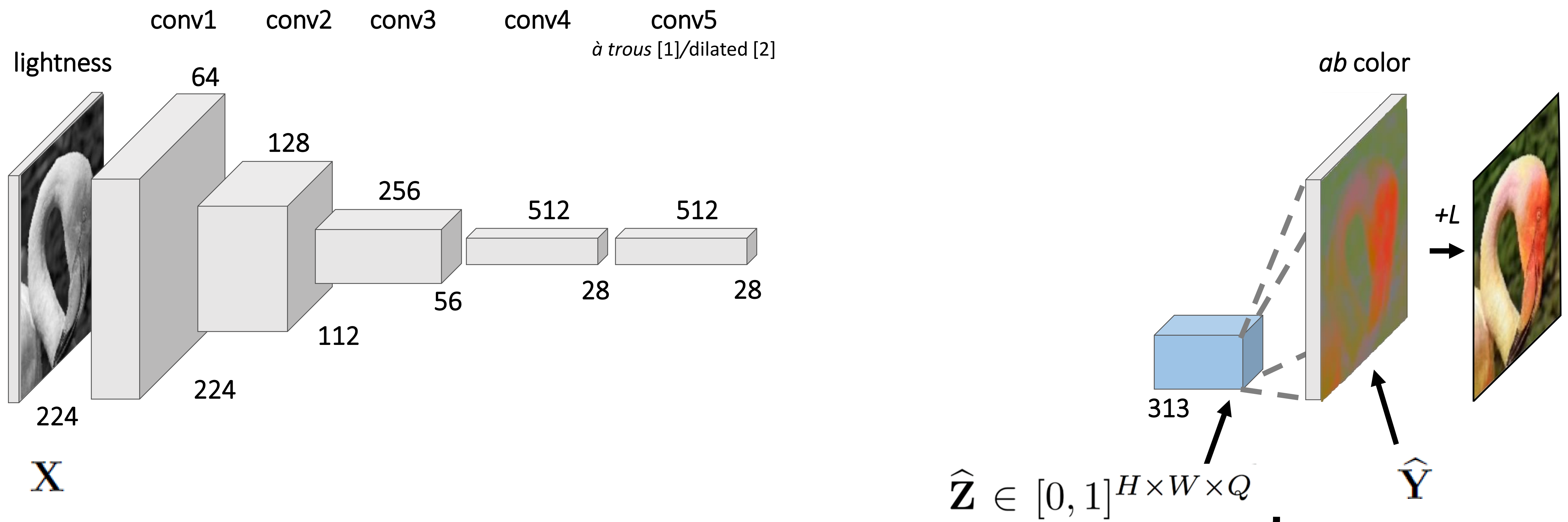
$$\widehat{\mathbf{Z}} = \mathcal{G}(\mathbf{X})$$

Network Architecture



$$\widehat{\mathbf{Z}} = \mathcal{G}(\mathbf{X})$$

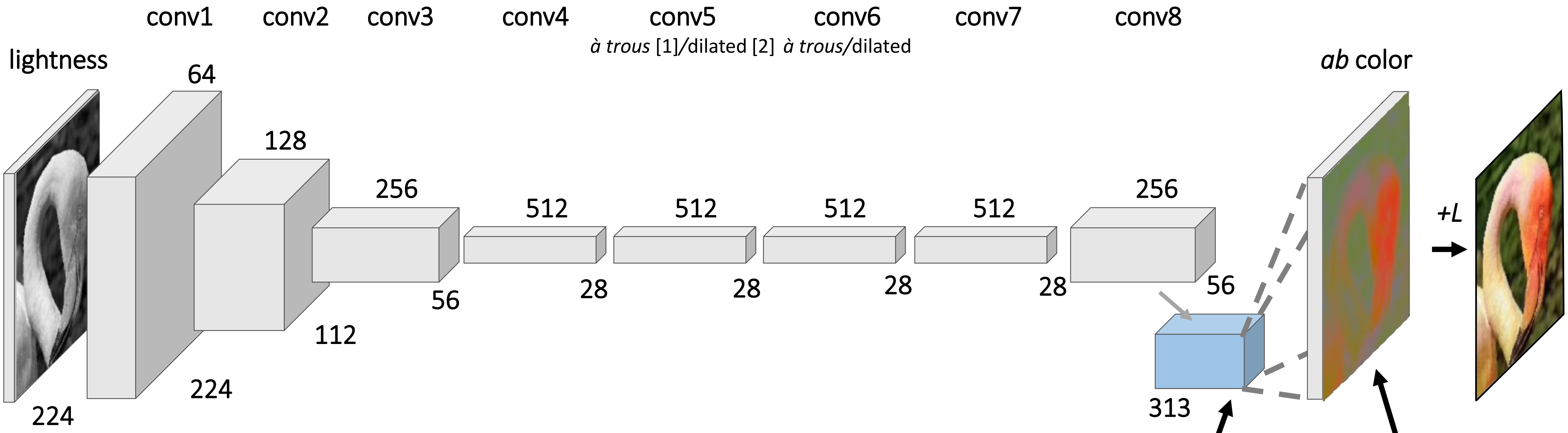
Network Architecture



$$\widehat{\mathbf{Z}} = \mathcal{G}(\mathbf{X})$$

[1] Chen *et al.* In arXiv, 2016.
[2] Yu and Koltun. In ICLR, 2016

Network Architecture



$\mathbf{X}$

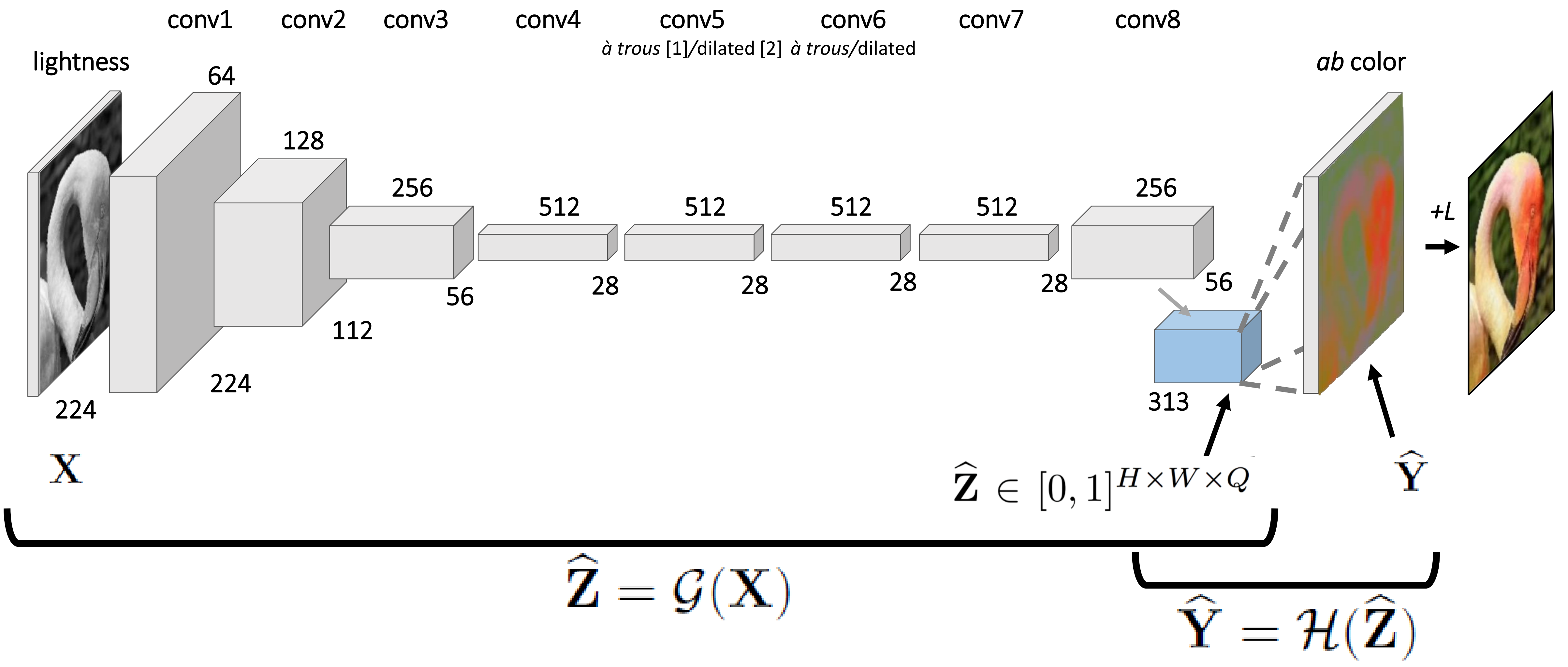
$\widehat{\mathbf{Z}} \in [0,1]^{H \times W \times Q}$

$\widehat{\mathbf{Y}}$

$\widehat{\mathbf{Z}} = \mathcal{G}(\mathbf{X})$

[1] Chen *et al.* In arXiv, 2016.
[2] Yu and Koltun. In ICLR, 2016

Network Architecture



[1] Chen *et al.* In arXiv, 2016.
[2] Yu and Koltun. In ICLR, 2016

Input

Input

L2 Regression

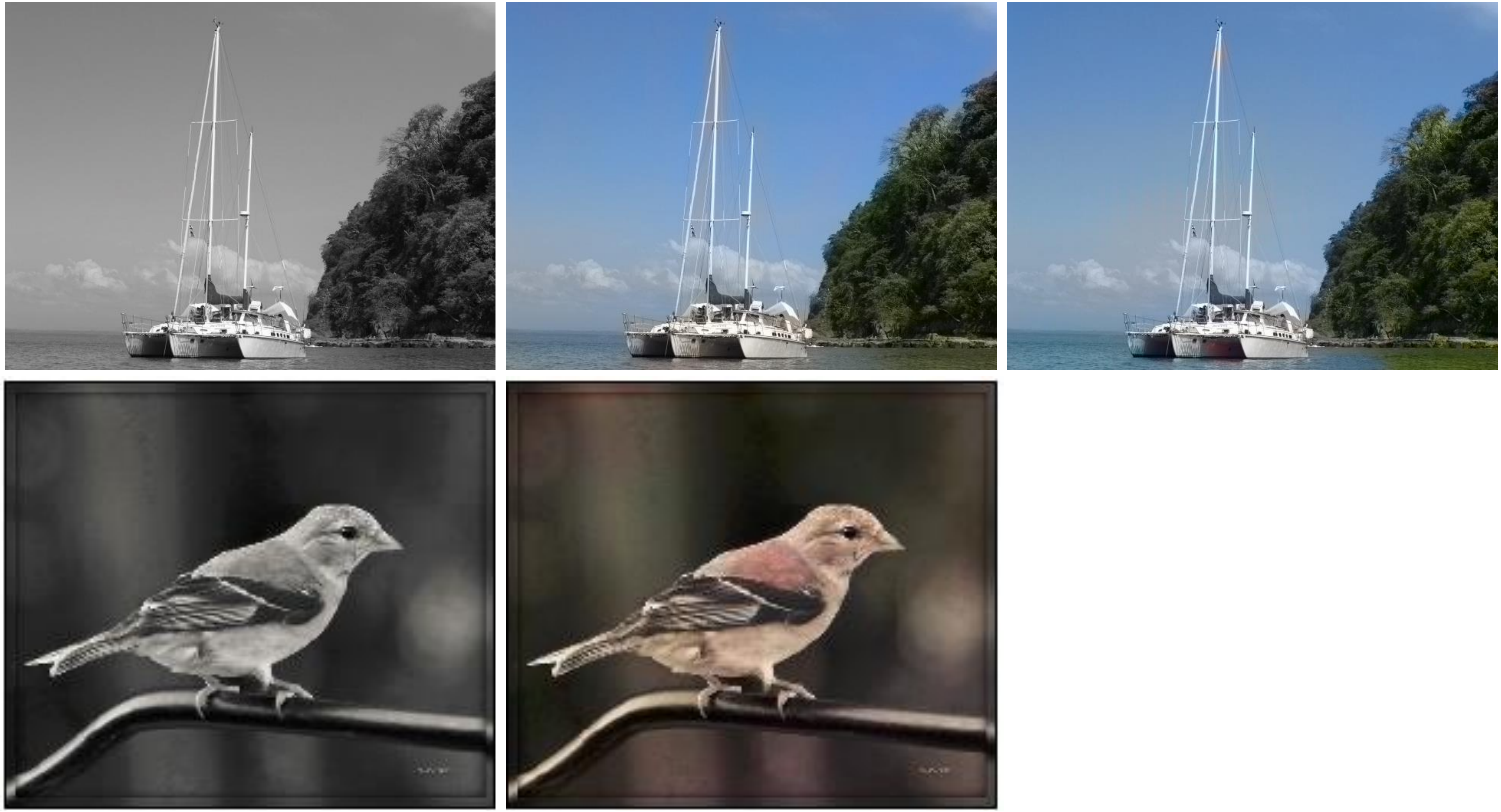
Input L2 Regression Class w/ Rebalancing

Input

L2 Regression

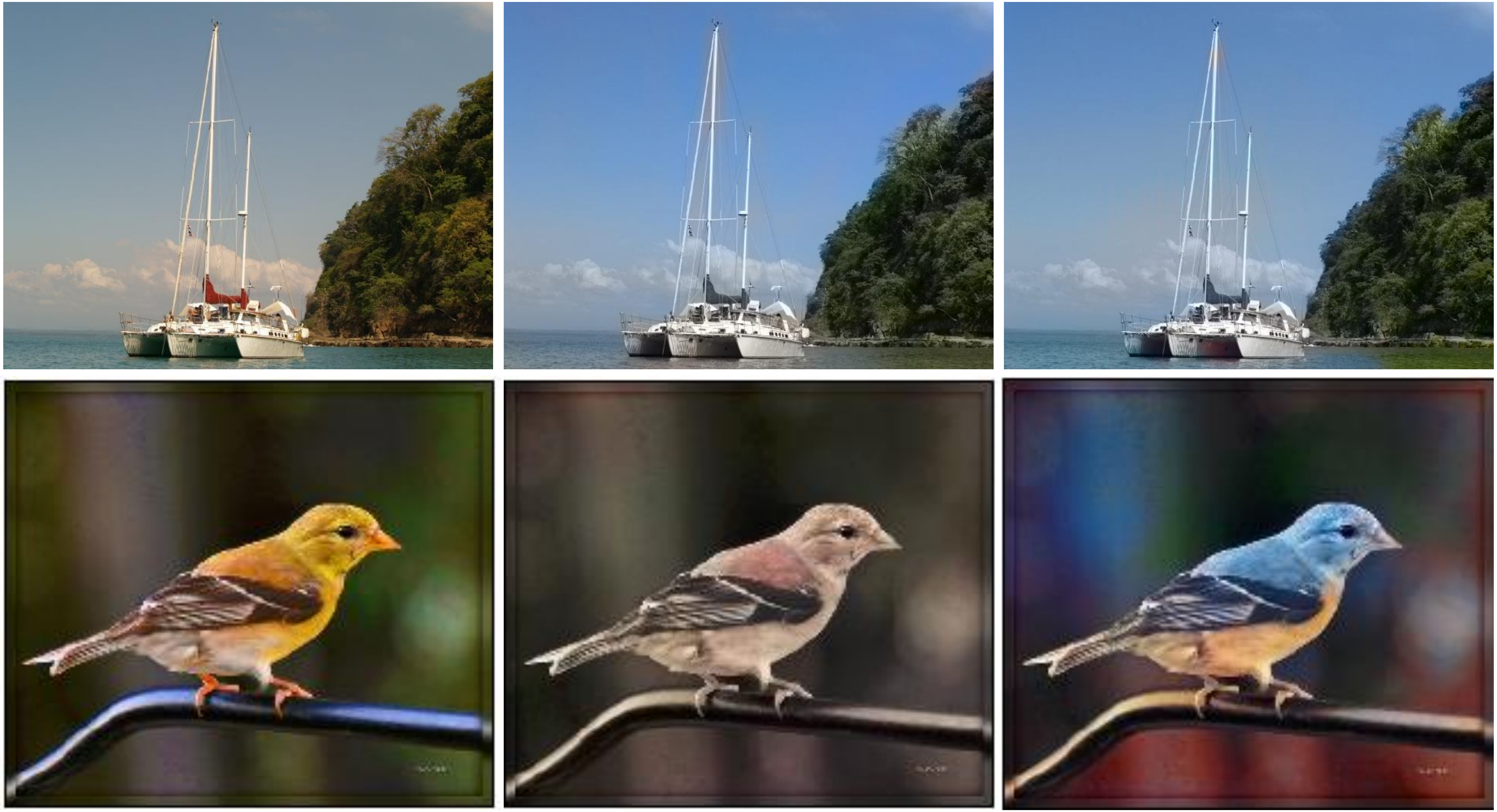
Class w/ Rebalancing

Input | L2 Regression | Class w/ Rebalancing



Input | L2 Regression | Class w/ Rebalancing

Ground Truth | L2 Regression | Class w/ Rebalancing



Failure Cases

Biases

Evaluation

	Visual Quality
Quantitative	Per-pixel accuracy

Evaluation

	Visual Quality
Quantitative	Per-pixel accuracy Perceptual realism Semantic interpretability
Qualitative	Low-level stimuli Legacy grayscale photos

Evaluation

	Visual Quality	**Representation Learning**
Quantitative	Per-pixel accuracy Perceptual realism Semantic interpretability	Task generalization ImageNet classification Task & dataset generalization PASCAL classification, detection, segmentation
Qualitative	Low-level stimuli Legacy grayscale photos	Hidden unit activations

Evaluation

	Visual Quality	Representation Learning
Quantitative	Per-pixel accuracy Perceptual realism Semantic interpretability	Task generalization ImageNet classification Task & dataset generalization PASCAL classification, detection, segmentation
Qualitative	Low-level stimuli Legacy grayscale photos	Hidden unit activations

Evaluation

	Visual Quality	Representation Learning
Quantitative	Per-pixel accuracy **Perceptual realism** Semantic interpretability	Task generalization ImageNet classification Task & dataset generalization PASCAL classification, detection, segmentation
Qualitative	Low-level stimuli Legacy grayscale photos	Hidden unit activations

Perceptual Realism / Amazon Mechanical Turk Test

clap if “fake”

clap if “fake”

Fake, 0% fooled

clap if “fake”

clap if “fake”

Fake, 55% fooled

clap if “fake”

clap if “fake”

Fake, 58% fooled

from Reddit /u/SherySantucci

Recolorized by Reddit ColorizeBot

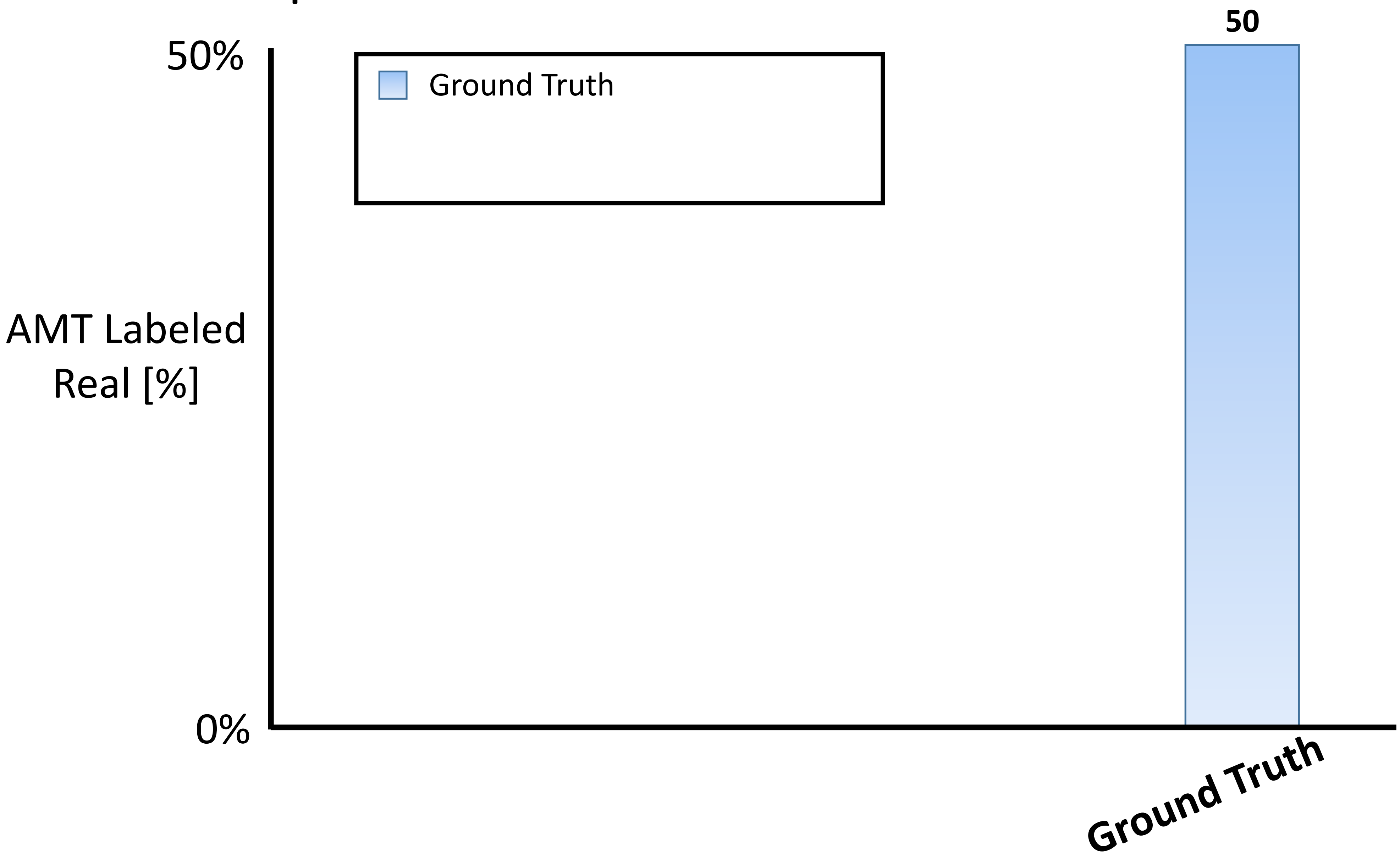
Photo taken by Reddit /u/Timteroo, Mural from street artist Eduardo Kobra

Recolorized by Reddit ColorizeBot

Perceptual Realism Test

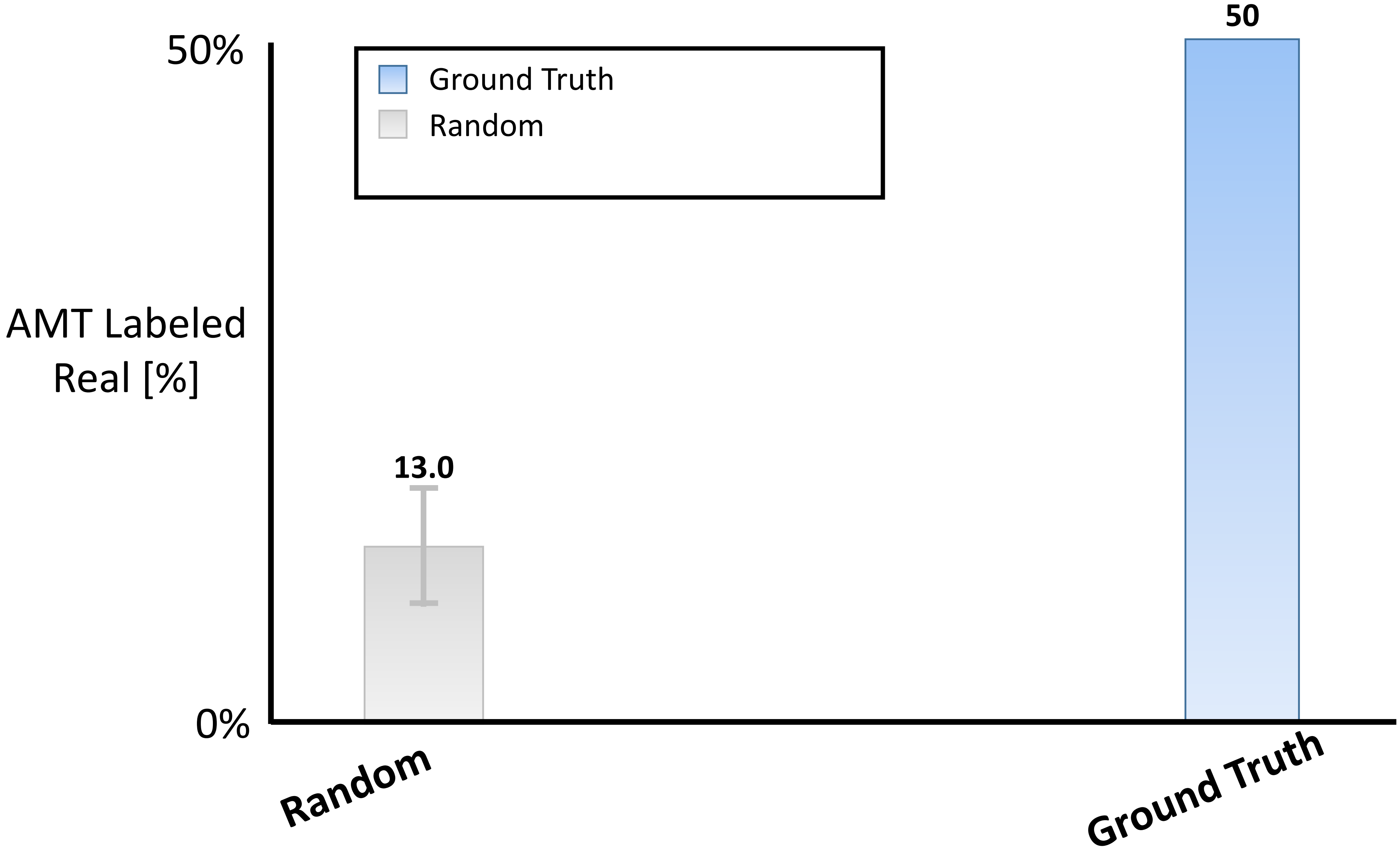
1600 images tested per algorithm

Perceptual Realism Test



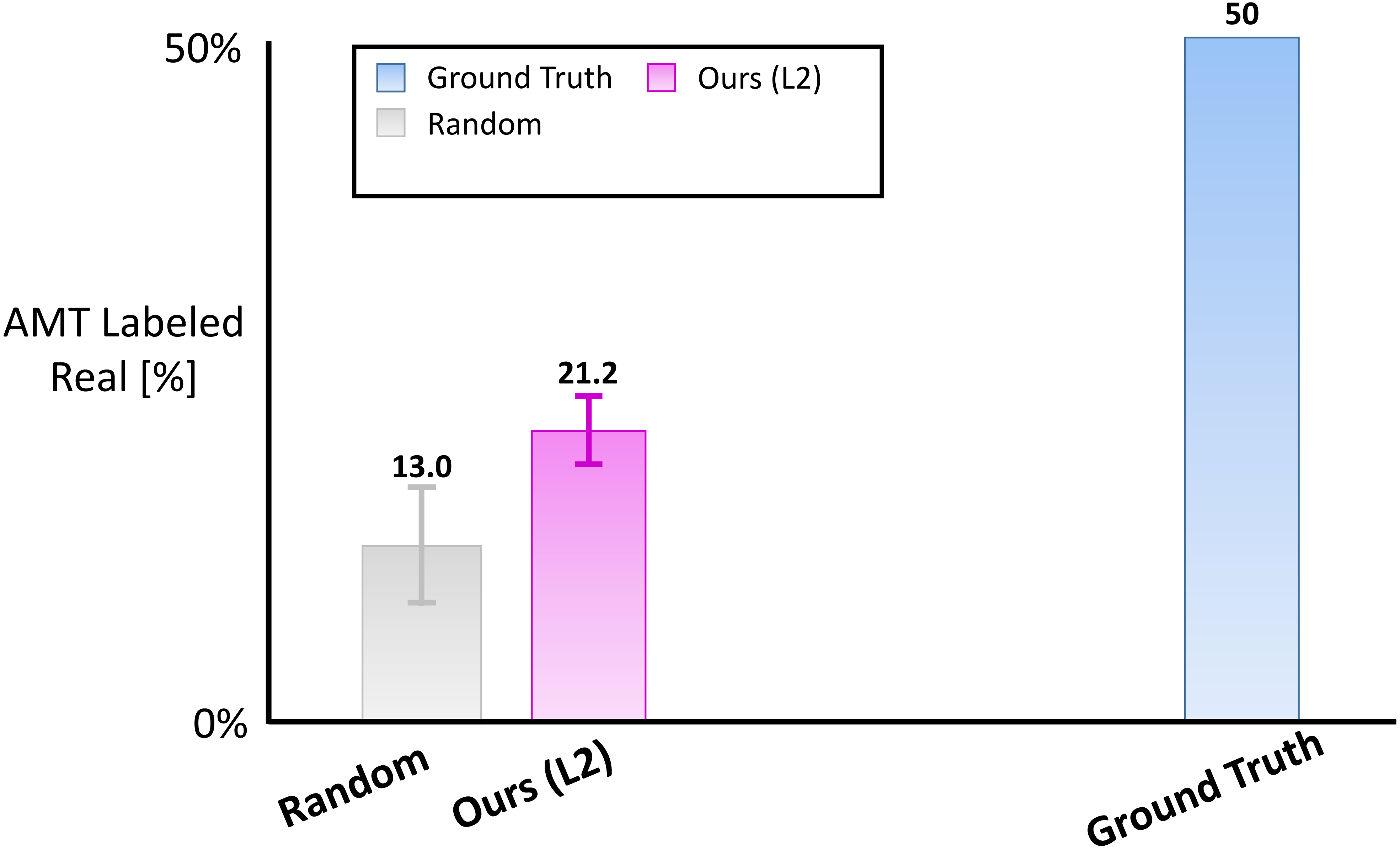
1600 images tested per algorithm

Perceptual Realism Test



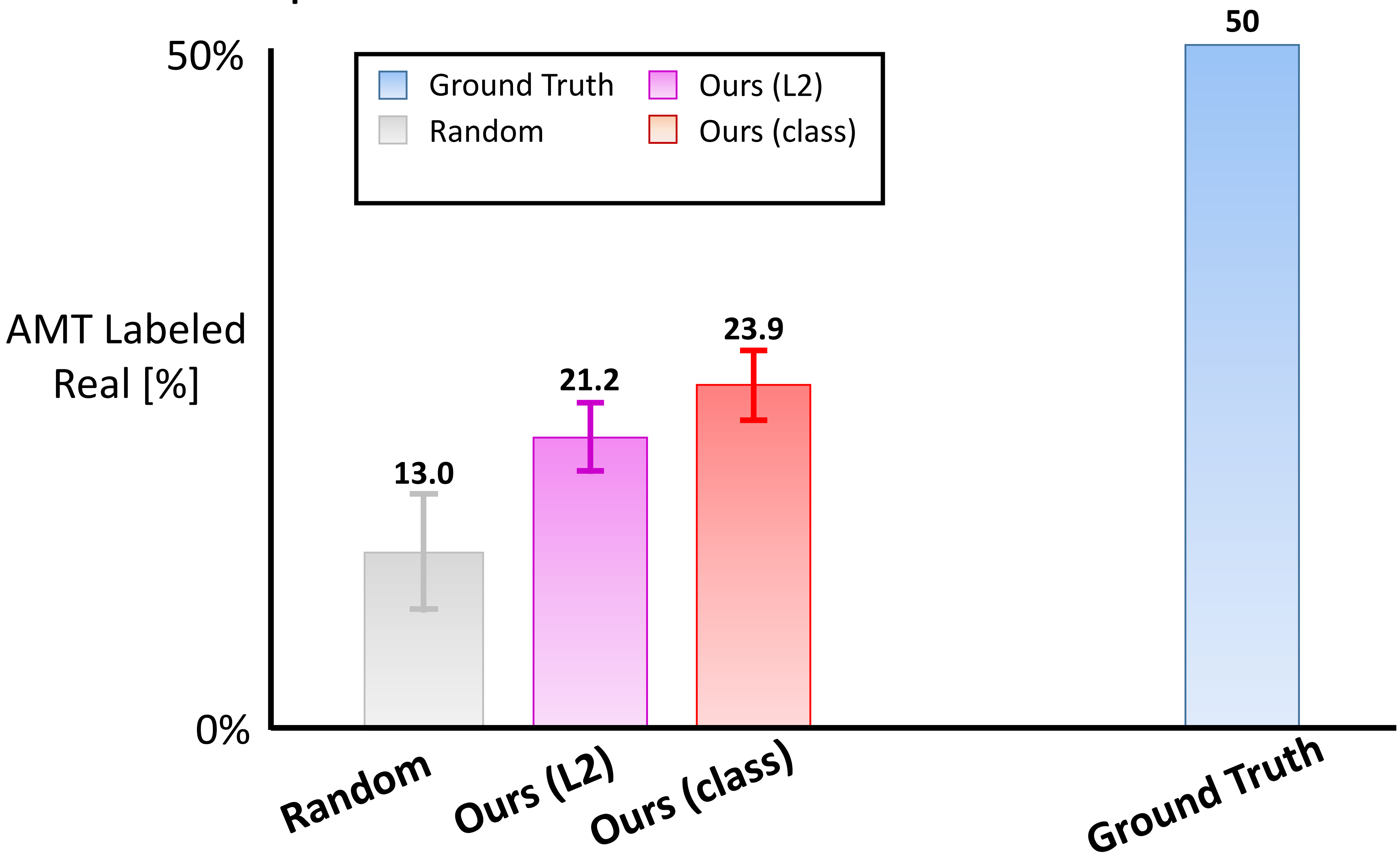
1600 images tested per algorithm

Perceptual Realism Test



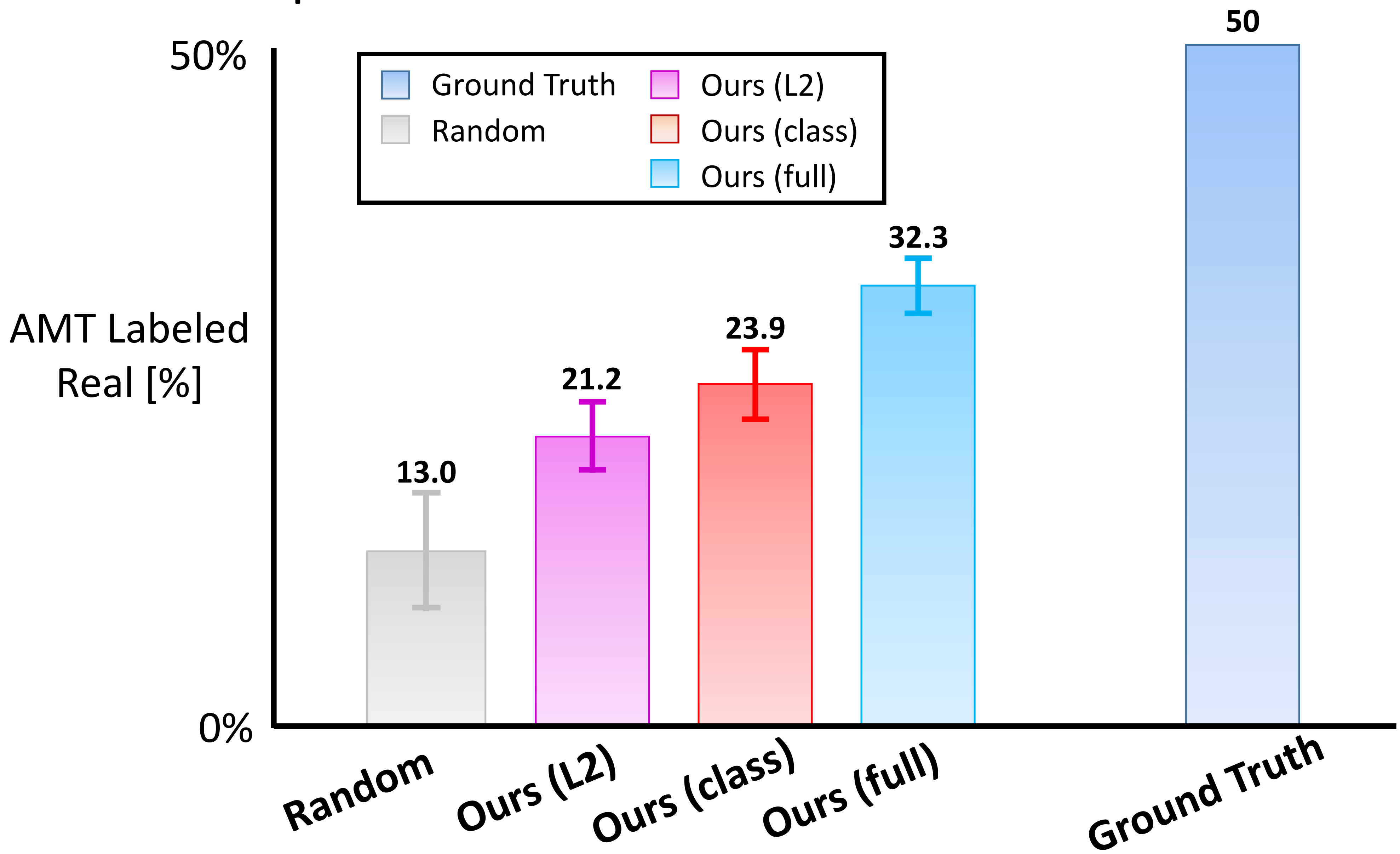
1600 images tested per algorithm

Perceptual Realism Test



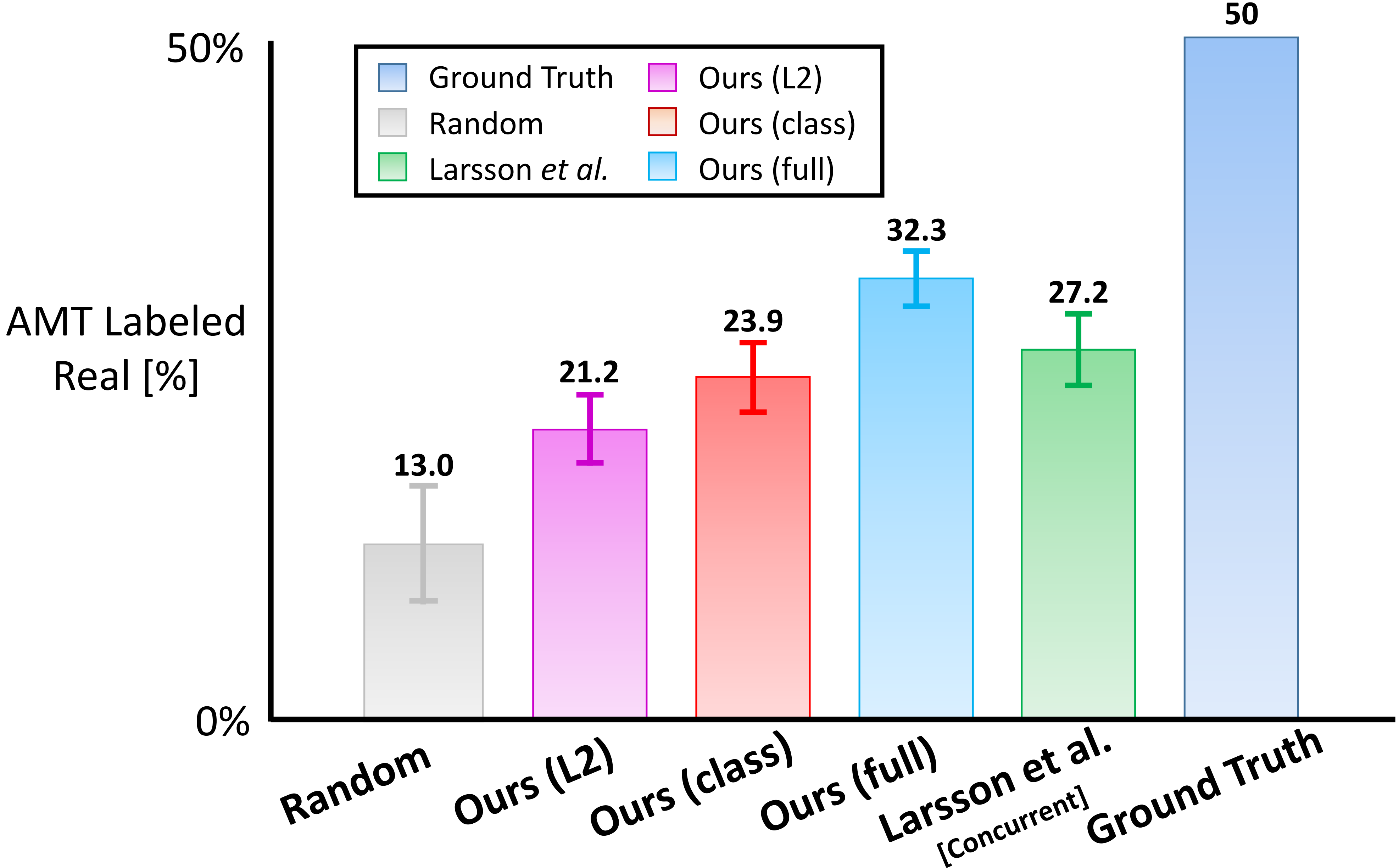
1600 images tested per algorithm

Perceptual Realism Test



1600 images tested per algorithm

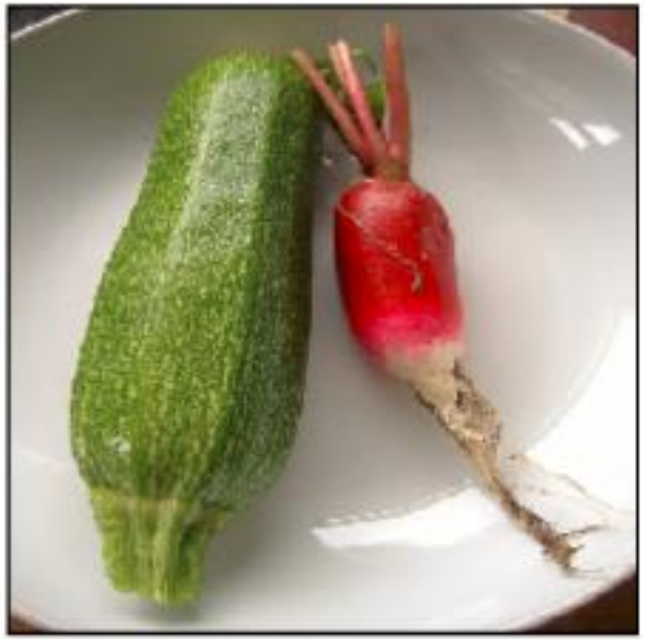
Perceptual Realism Test



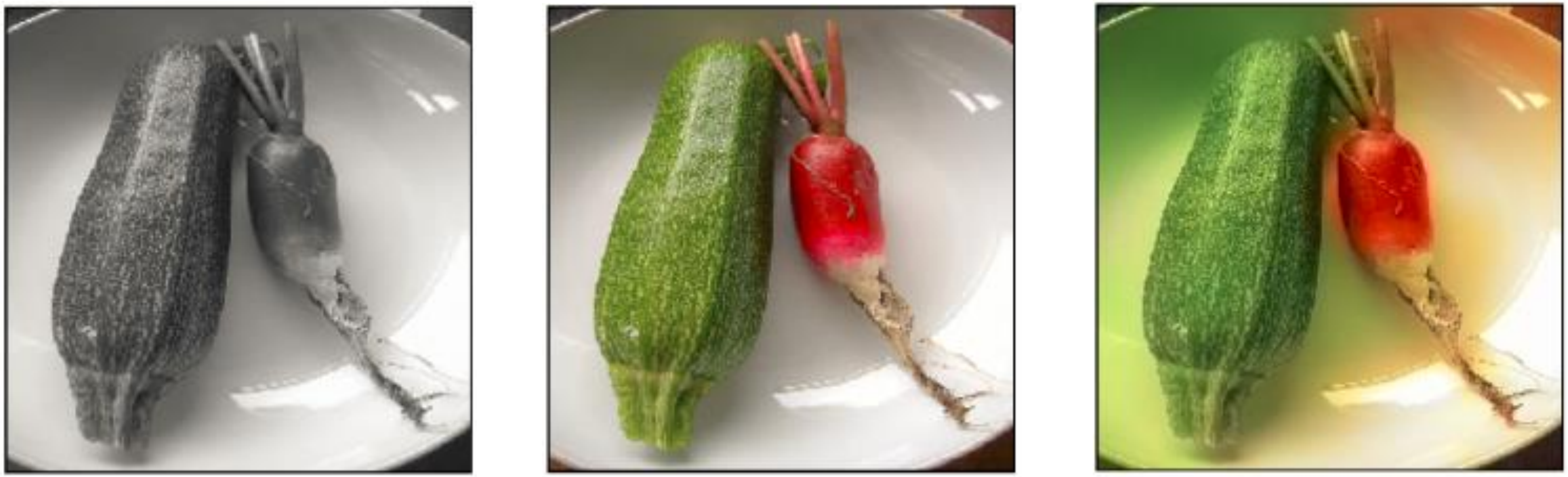
1600 images tested per algorithm

Input

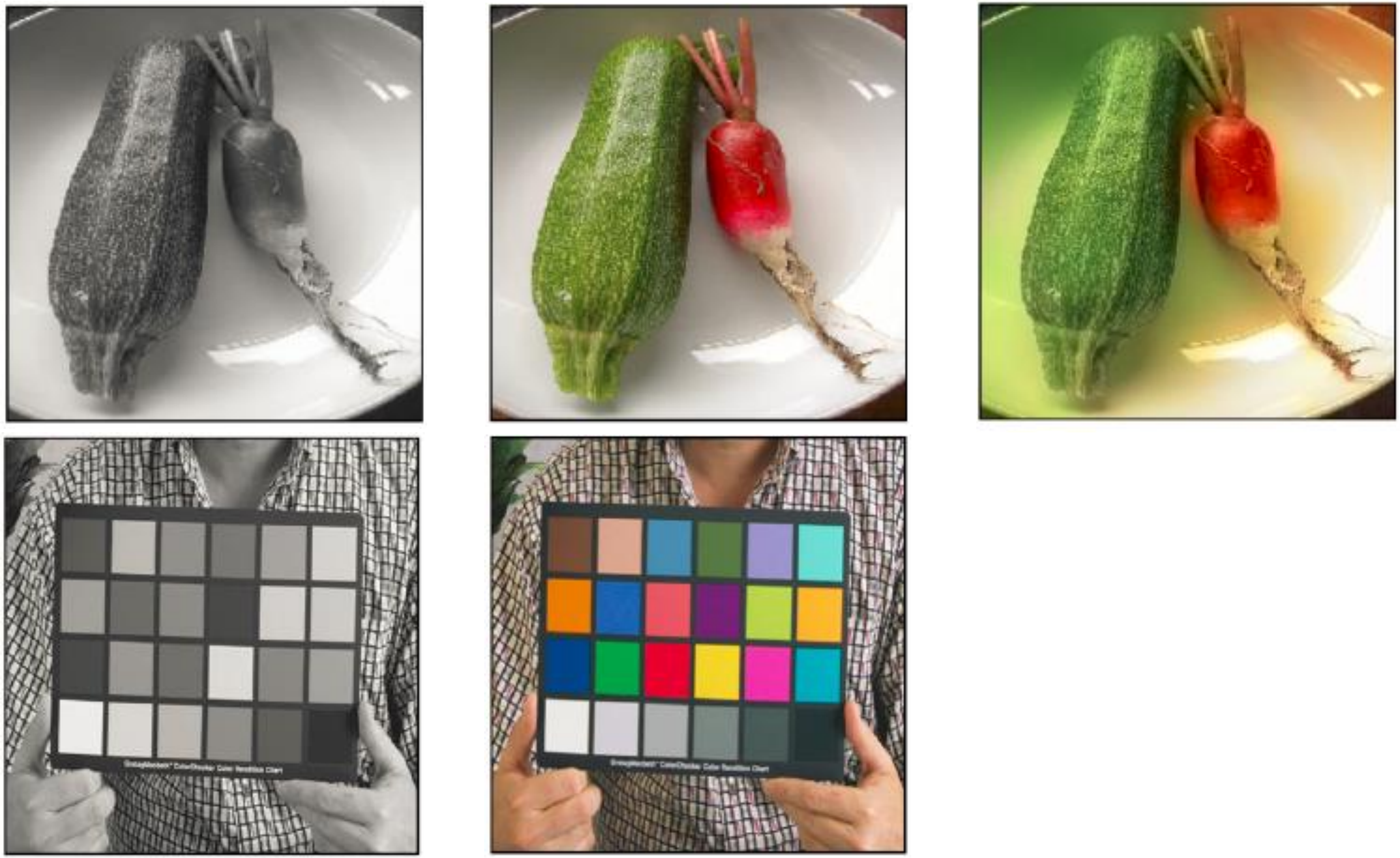
Ground Truth

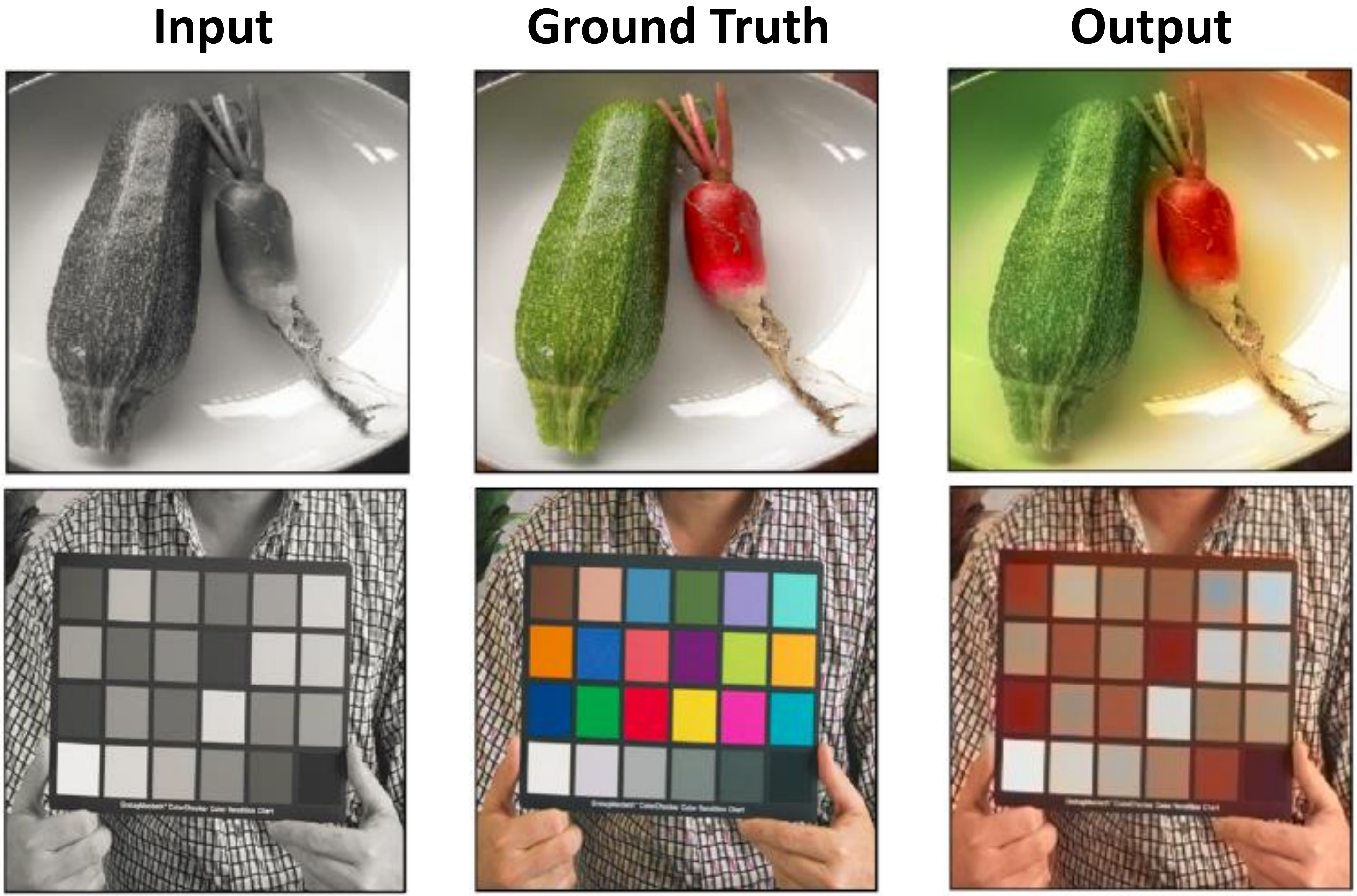


Input **Ground Truth** **Output**



Input **Ground Truth** **Output**




Input
Ground Truth
Output

Predicting Labels from Data

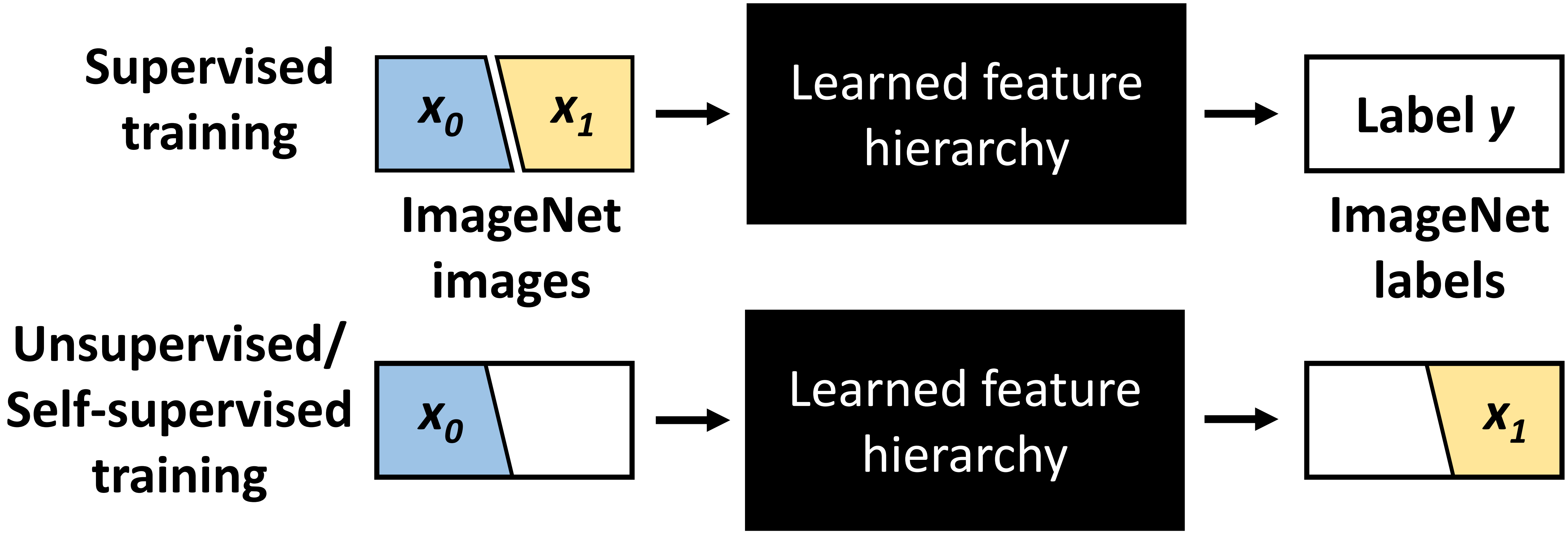
Supervised training

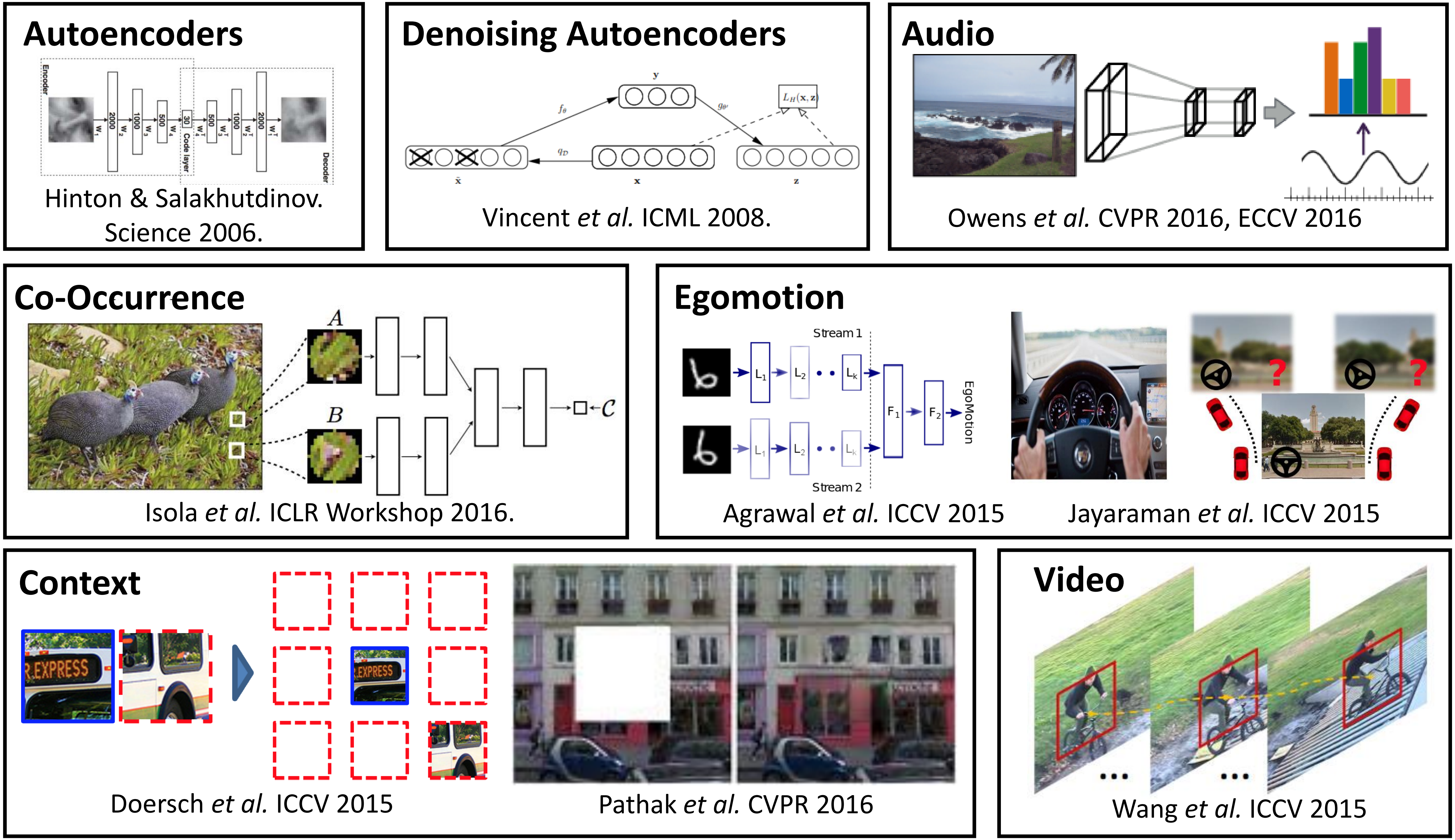
Data *x* → Learned feature hierarchy → **Label *y***

ImageNet images

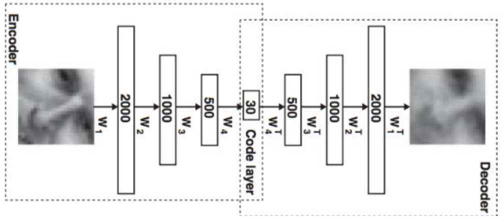
ImageNet labels

Predicting Data from Data



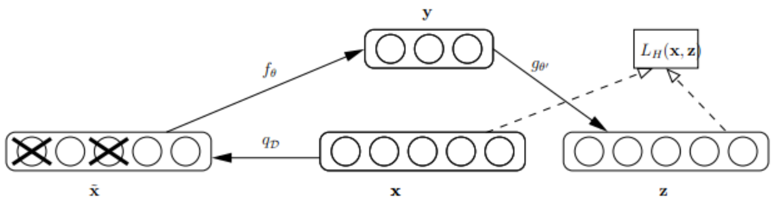

Autoencoders
Encoder
2000
1000
500
30
Code layer
Decoder
Hinton & Salakhutdinov.
Science 2006.
Denoising Autoencoders
y
x
z
Vincent et al. ICML 2008.
Audio
Owens et al. CVPR 2016, ECCV 2016
Co-Occurrence
A
B
C
Isola et al. ICLR Workshop 2016.
Egomotion
Stream 1
Stream 2
EgoMotion
Agrawal et al. ICCV 2015
Jayaraman et al. ICCV 2015
Context
EXPRESS
Doersch et al. ICCV 2015
Pathak et al. CVPR 2016
Video
Wang et al. ICCV 2015

Autoencoders



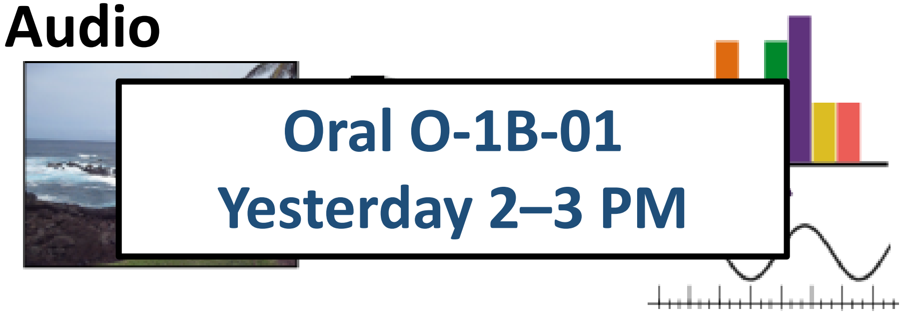
Hinton & Salakhutdinov. Science 2006.

Denoising Autoencoders



Vincent *et al.* ICML 2008.

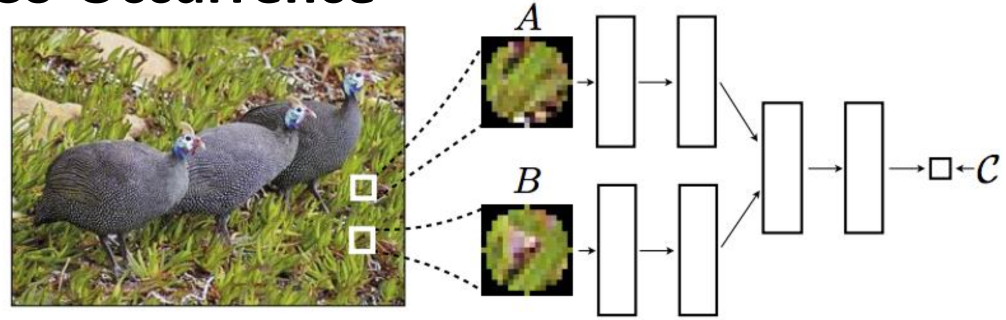
Audio



Oral O-1B-01
Yesterday 2–3 PM

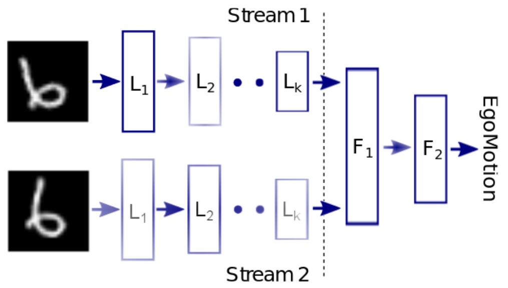
Owens *et al.* CVPR 2016, ECCV 2016

Co-Occurrence



Isola *et al.* ICLR Workshop 2016.

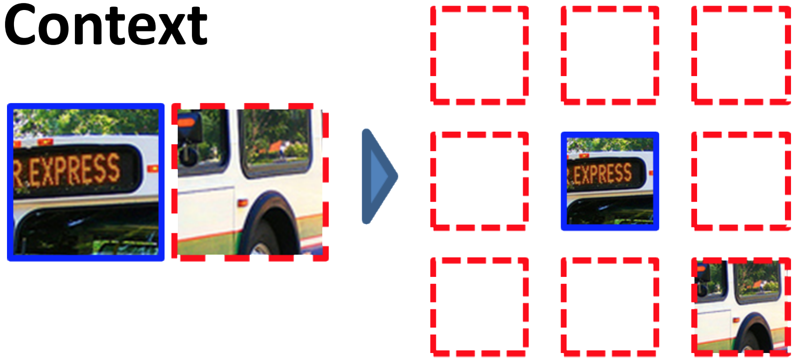
Egomotion



Agrawal *et al.* ICCV 2015

Jayaraman *et al.* ICCV 2015

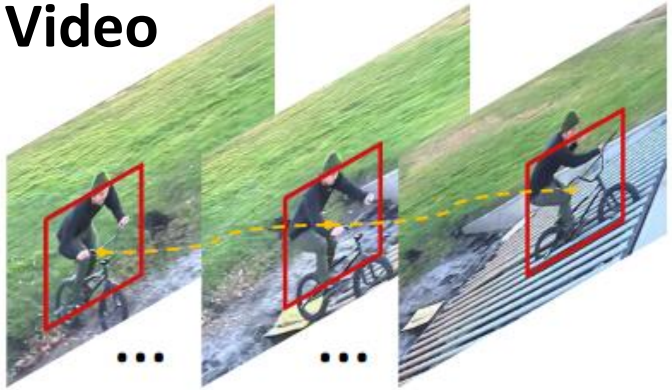
Context



Doersch *et al.* ICCV 2015

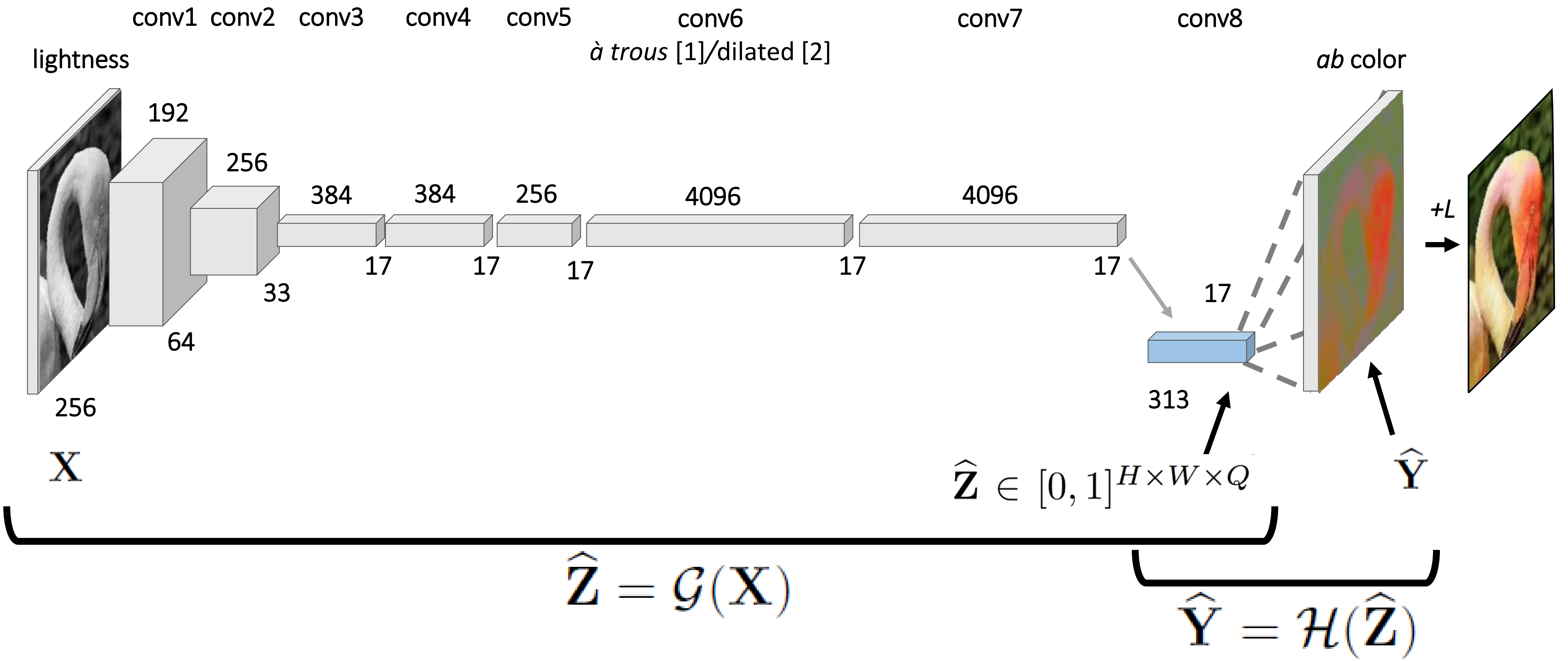
Pathak *et al.* CVPR 2016

Video



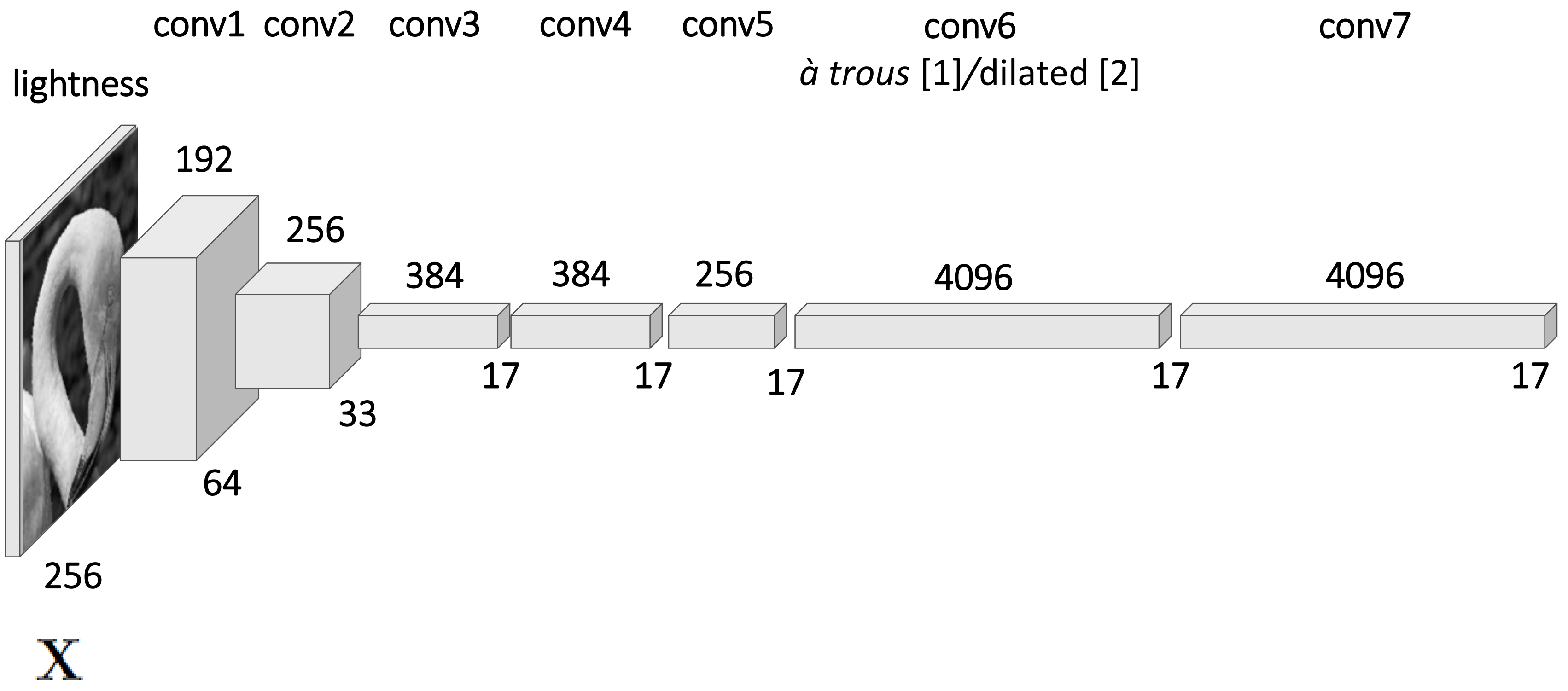
Wang *et al.* ICCV 2015

Cross-Channel Encoder



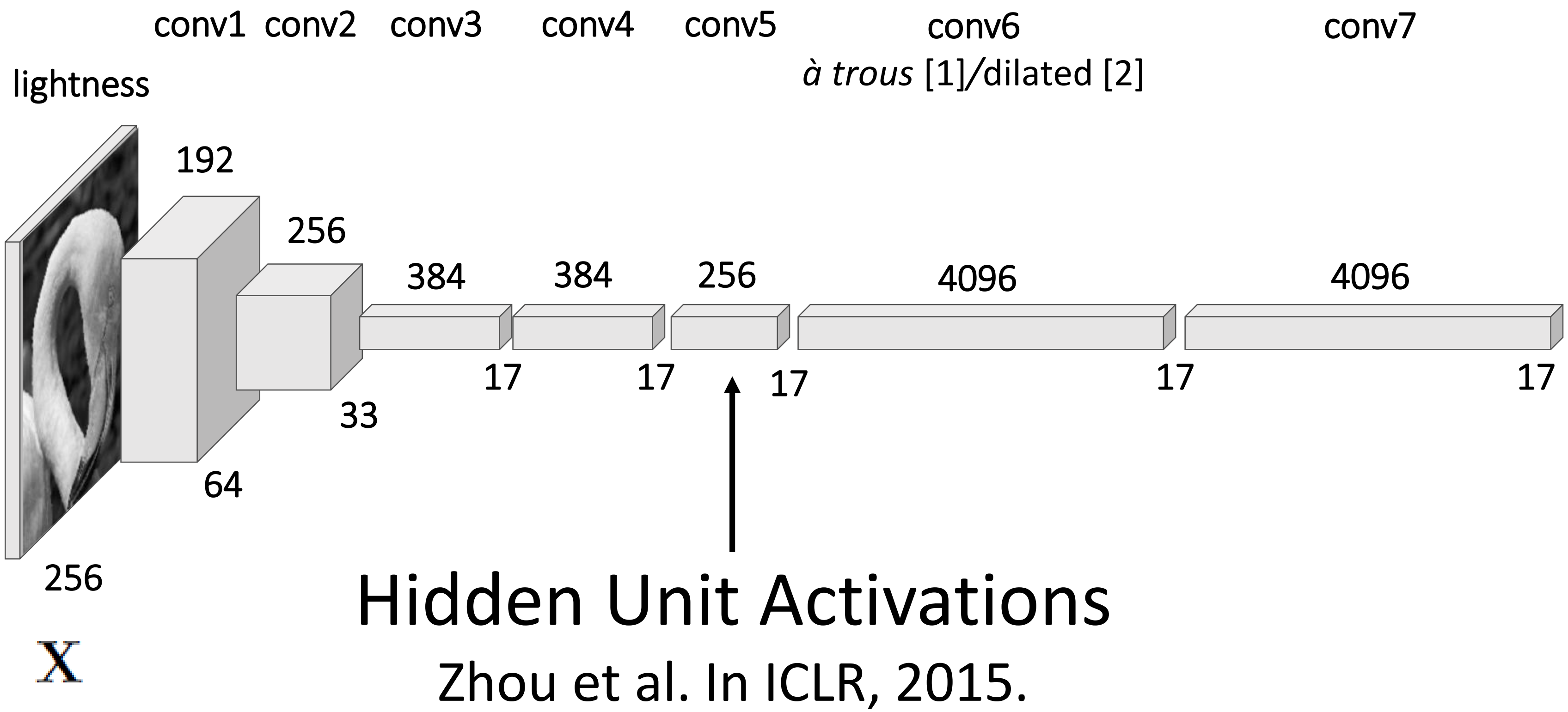
[1] Chen *et al.* In arXiv, 2016.
[2] Yu and Koltun. In ICLR, 2016

Cross-Channel Encoder



[1] Chen *et al.* In arXiv, 2016.
[2] Yu and Koltun. In ICLR, 2016

Cross-Channel Encoder



[1] Chen *et al.* In arXiv, 2016.
[2] Yu and Koltun. In ICLR, 2016

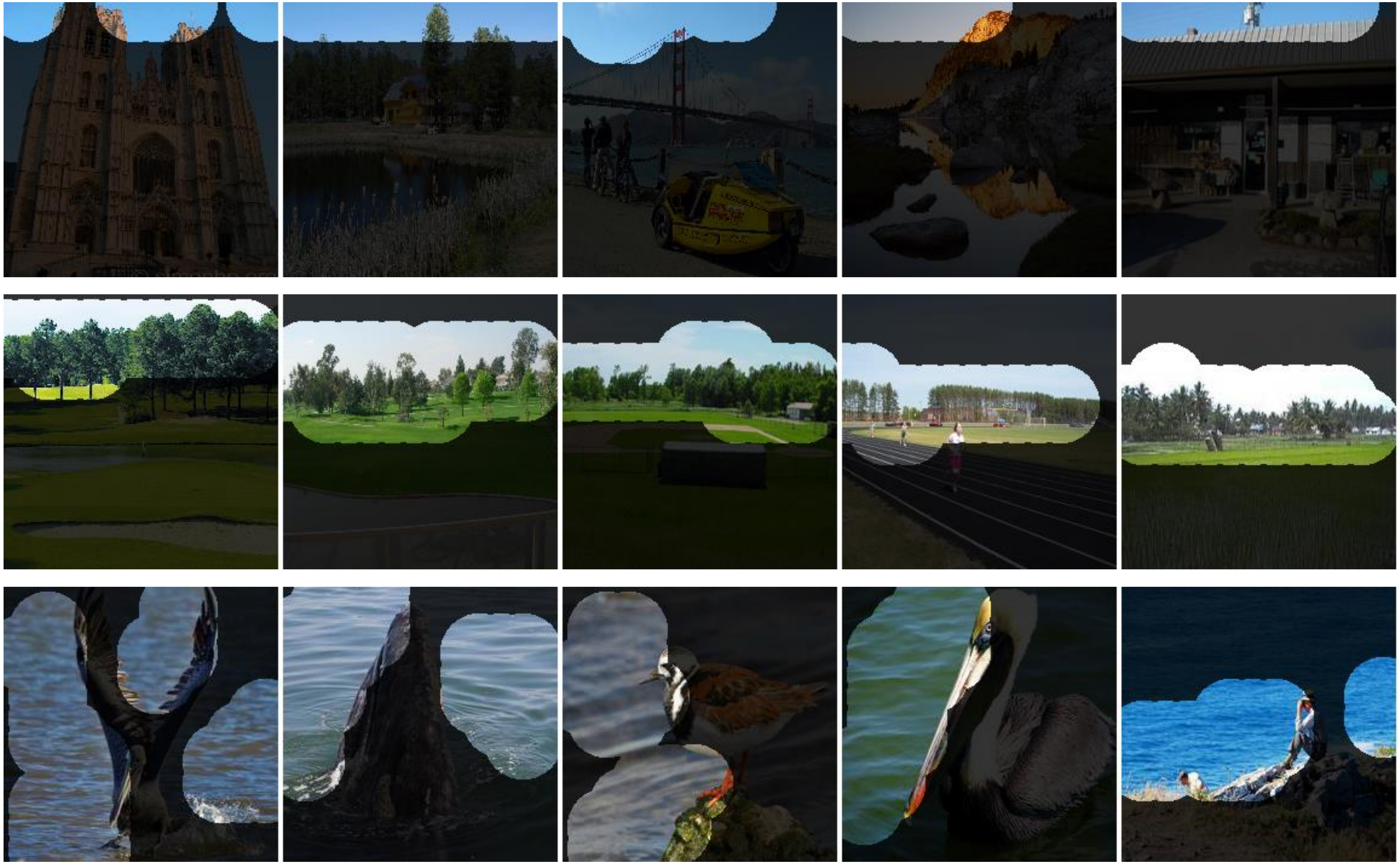
Hidden Unit (conv5) Activations

Hidden Unit (conv5) Activations

sky

trees

water



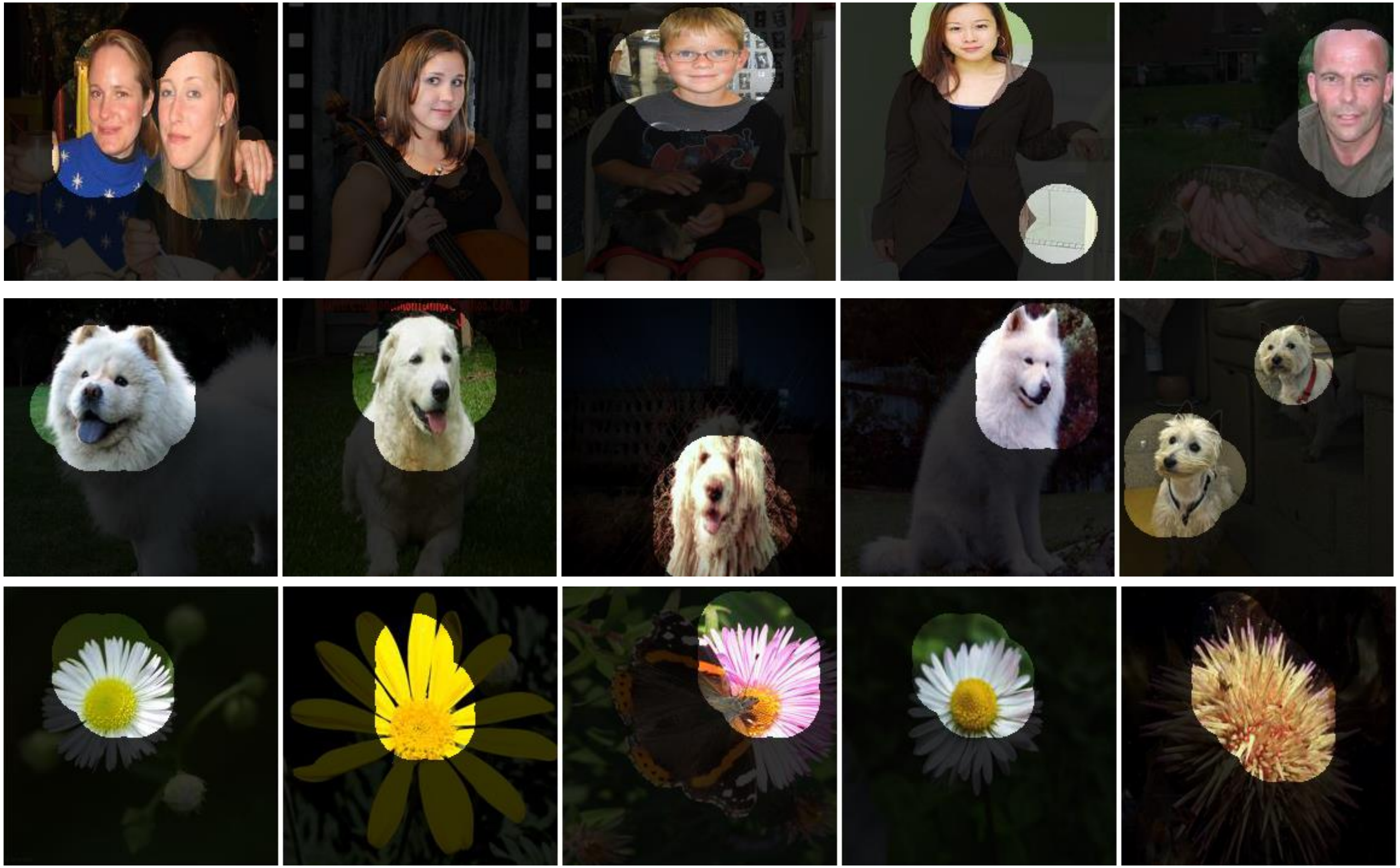
Hidden Unit (conv5) Activations

Hidden Unit (conv5) Activations

faces

dog faces

flowers



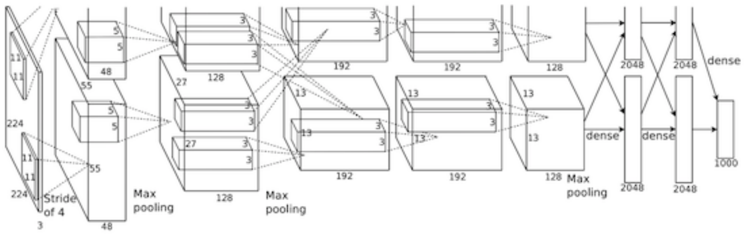
Dataset & Task Generalization on PASCAL VOC

Dataset & Task Generalization on PASCAL VOC

Does the feature representation
transfer to other datasets and tasks?

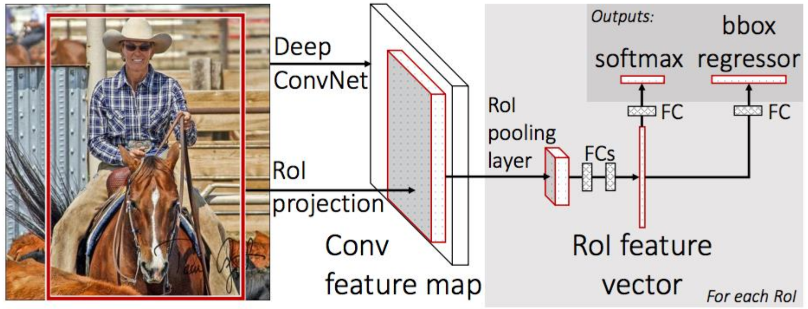
Dataset & Task Generalization on PASCAL VOC

Does the feature representation
transfer to other datasets and tasks?



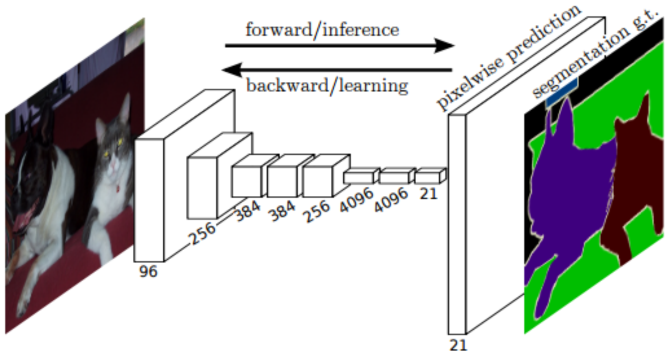
Classification

Krähenbühl et al. In ICLR, 2016.



Detection

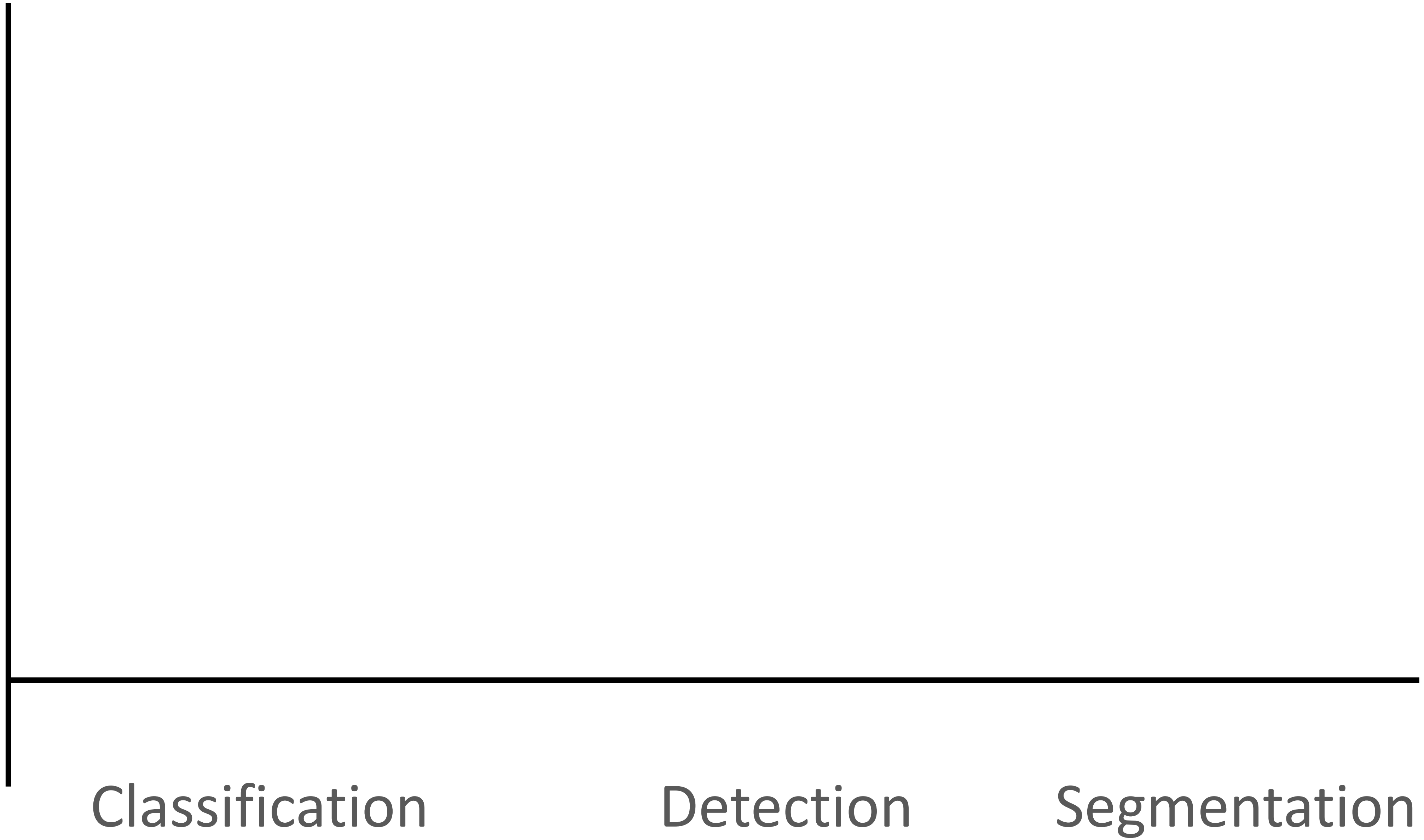
Fast R-CNN. Girshick. In ICCV, 2015.



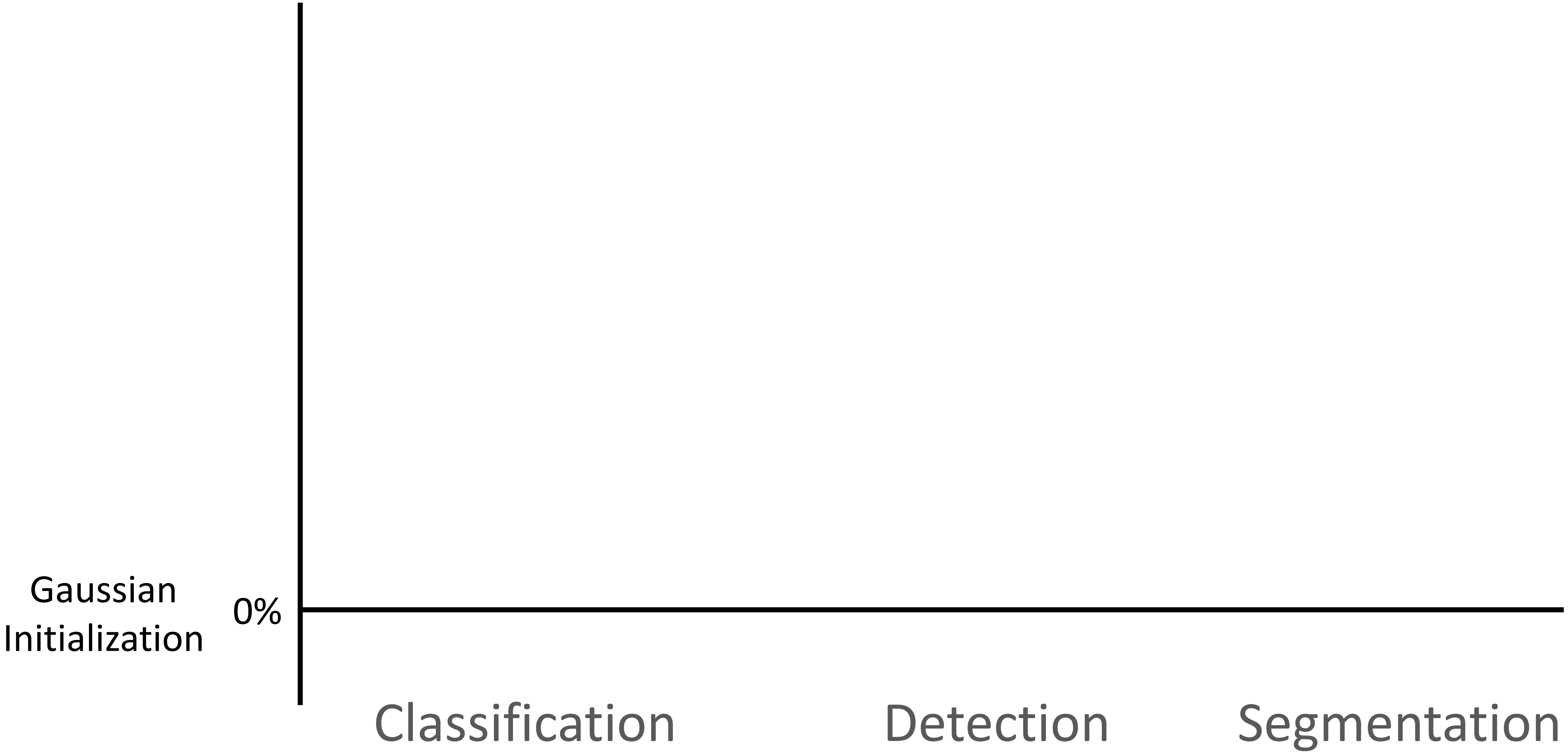
Segmentation

FCNs. Long et al. In CVPR, 2015.

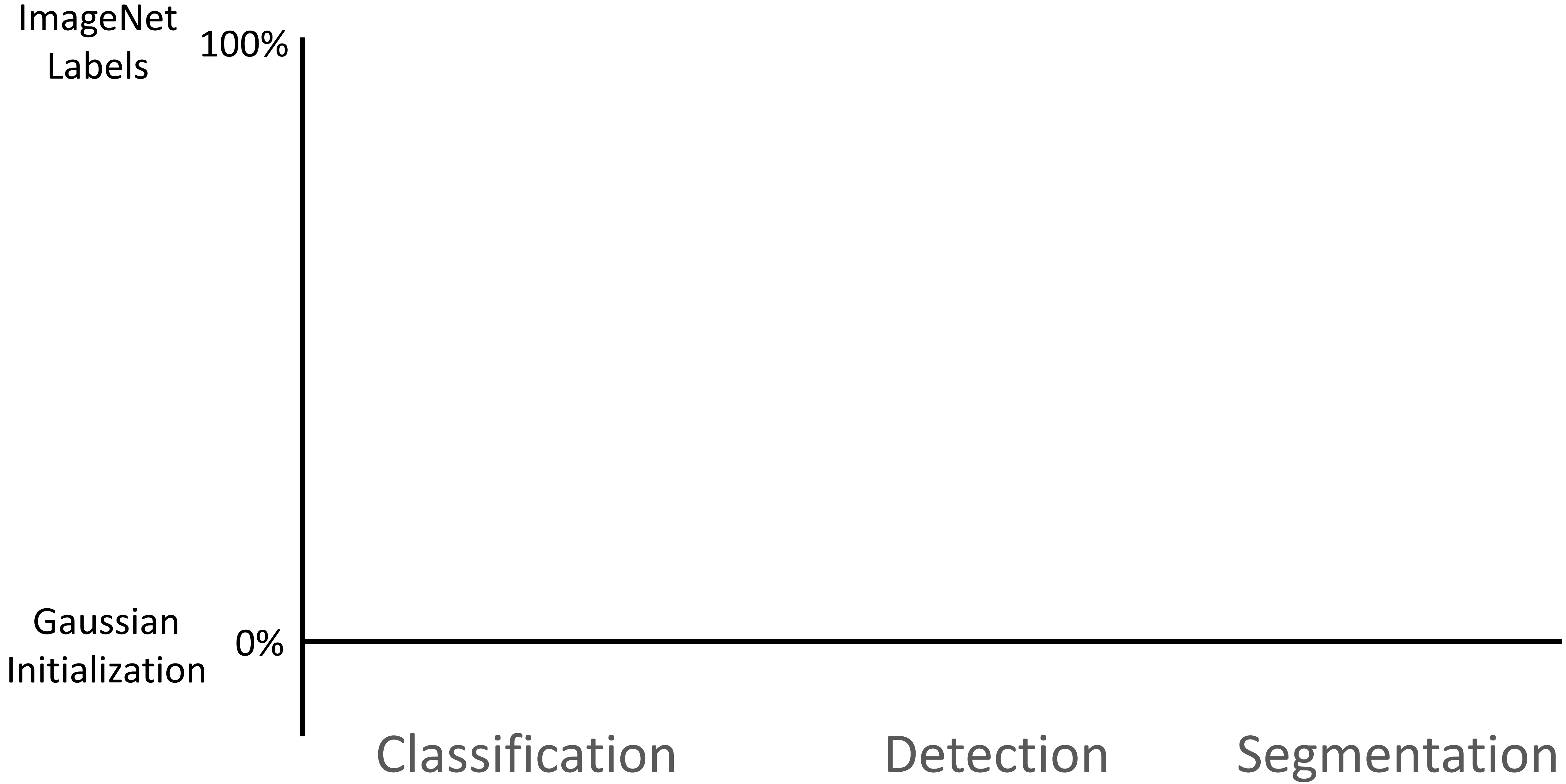
Dataset & Task Generalization on PASCAL VOC



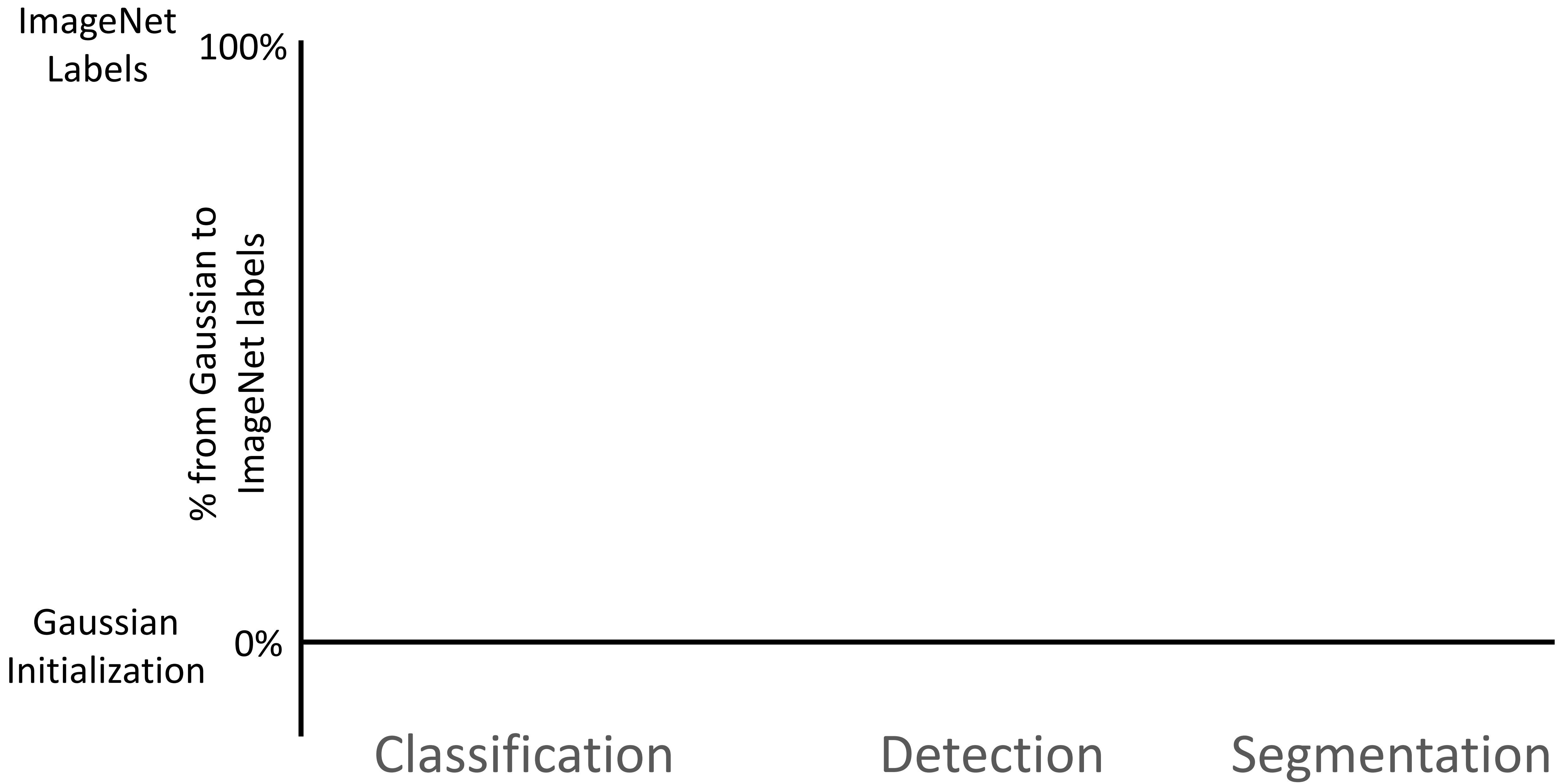
Dataset & Task Generalization on PASCAL VOC



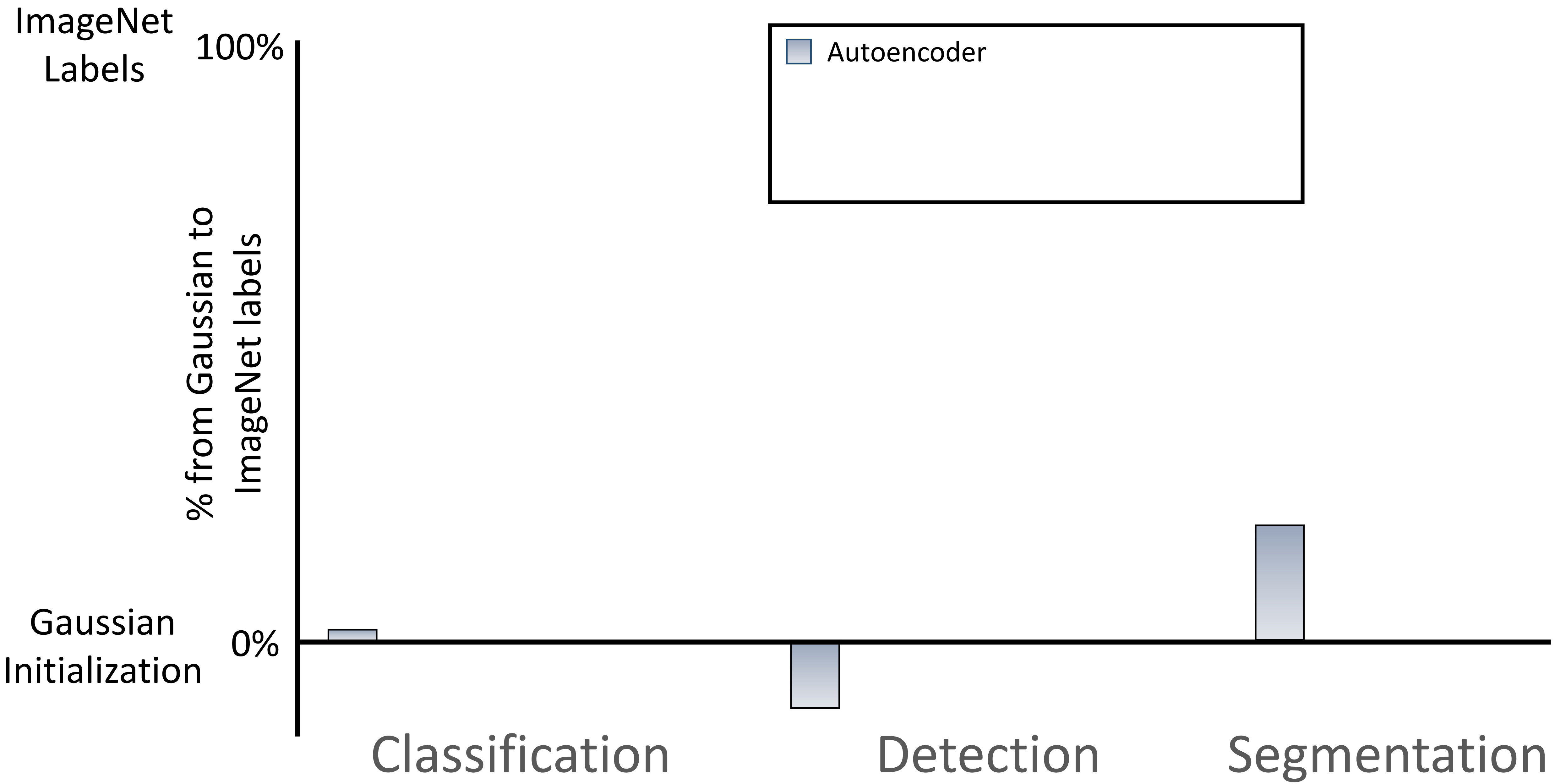
Dataset & Task Generalization on PASCAL VOC



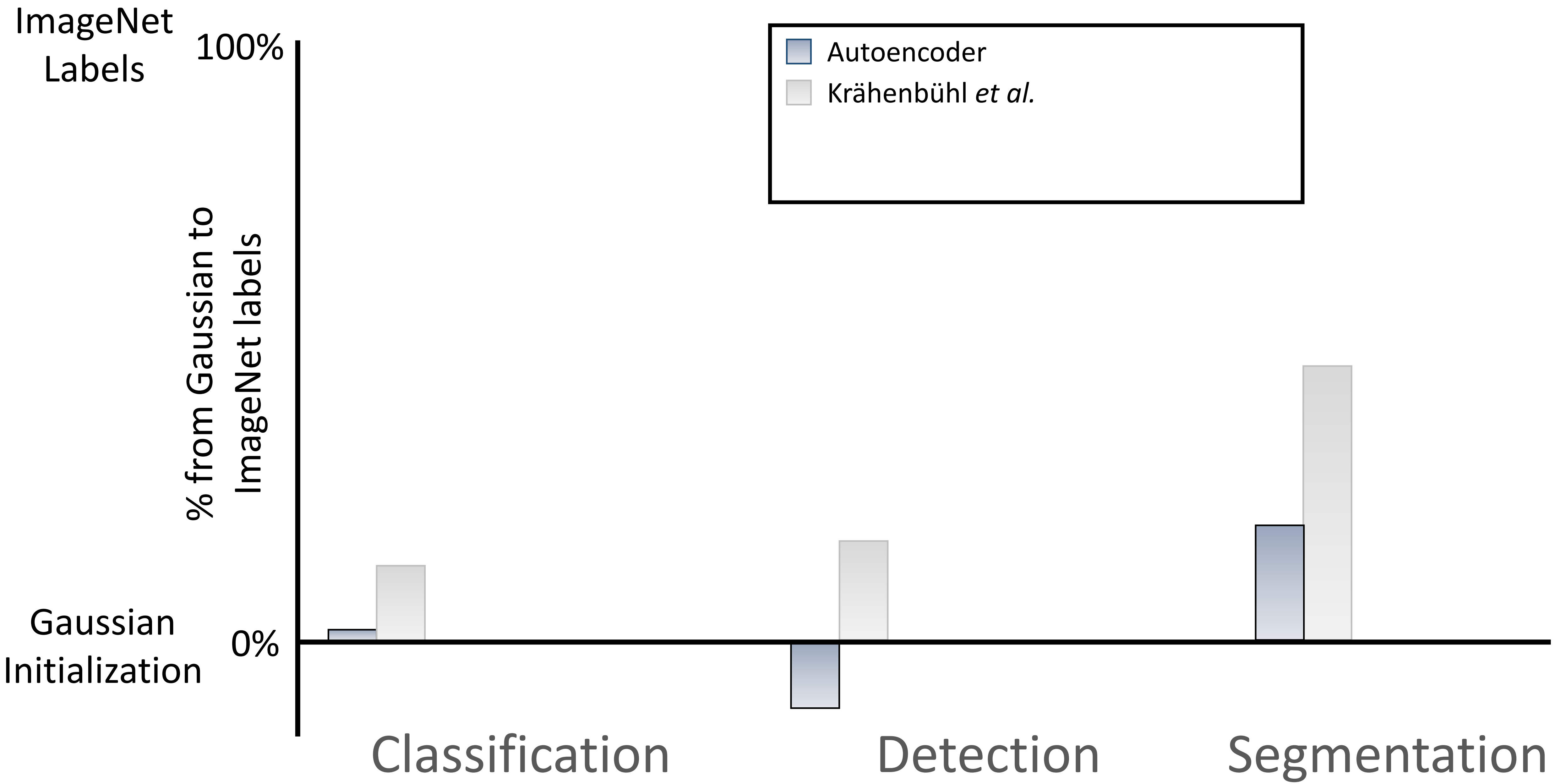
Dataset & Task Generalization on PASCAL VOC



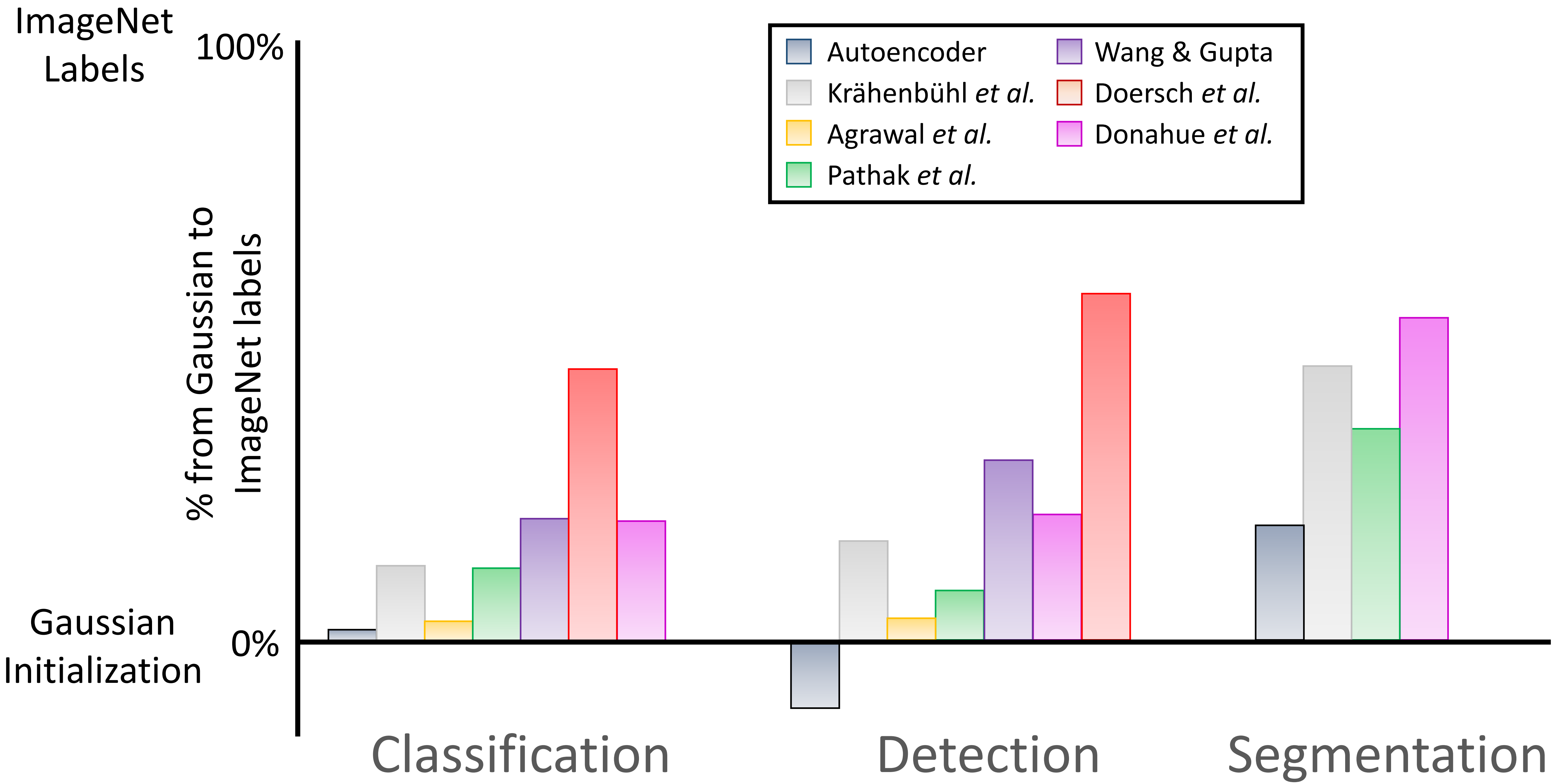
Dataset & Task Generalization on PASCAL VOC



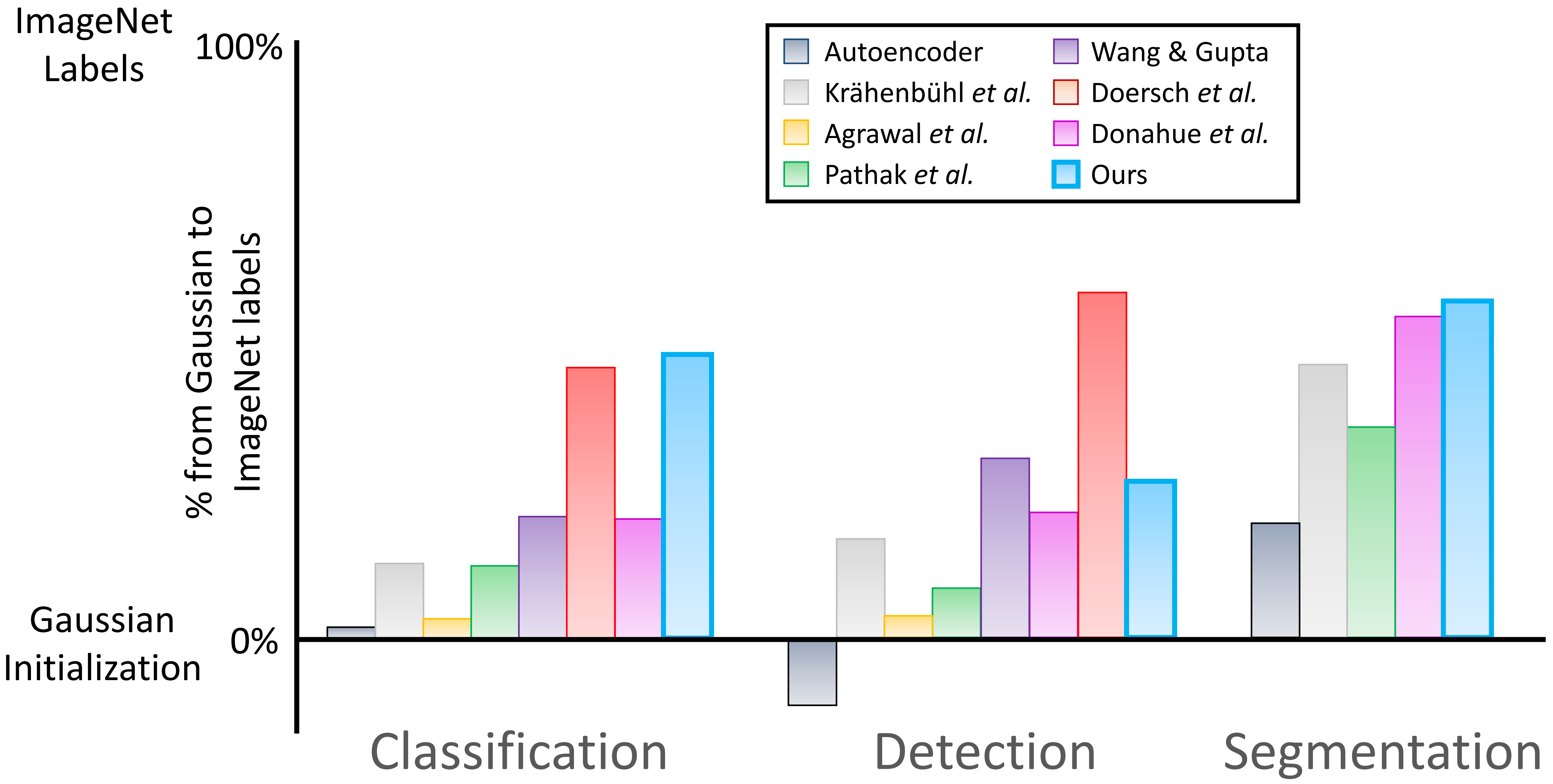
Dataset & Task Generalization on PASCAL VOC



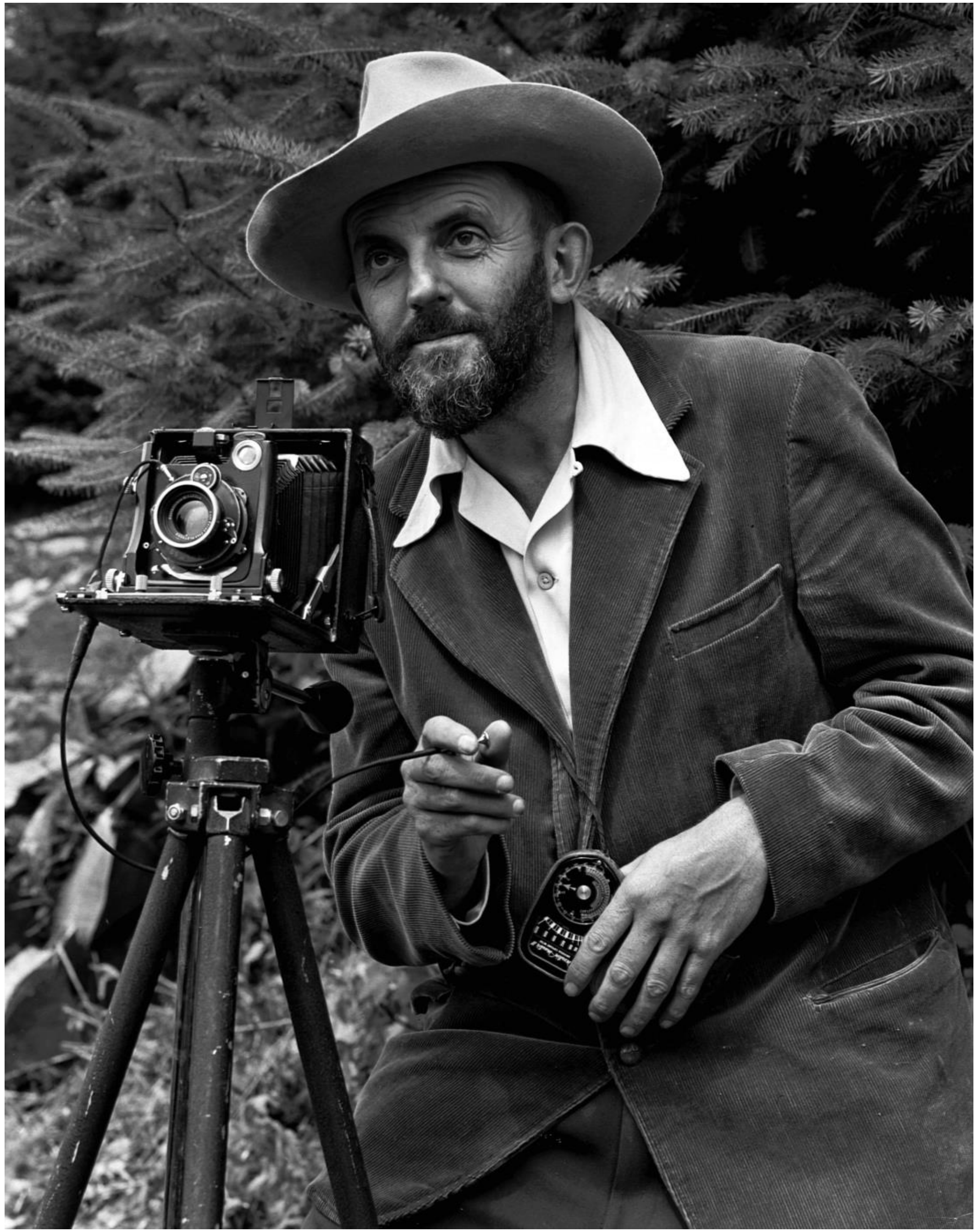
Dataset & Task Generalization on PASCAL VOC

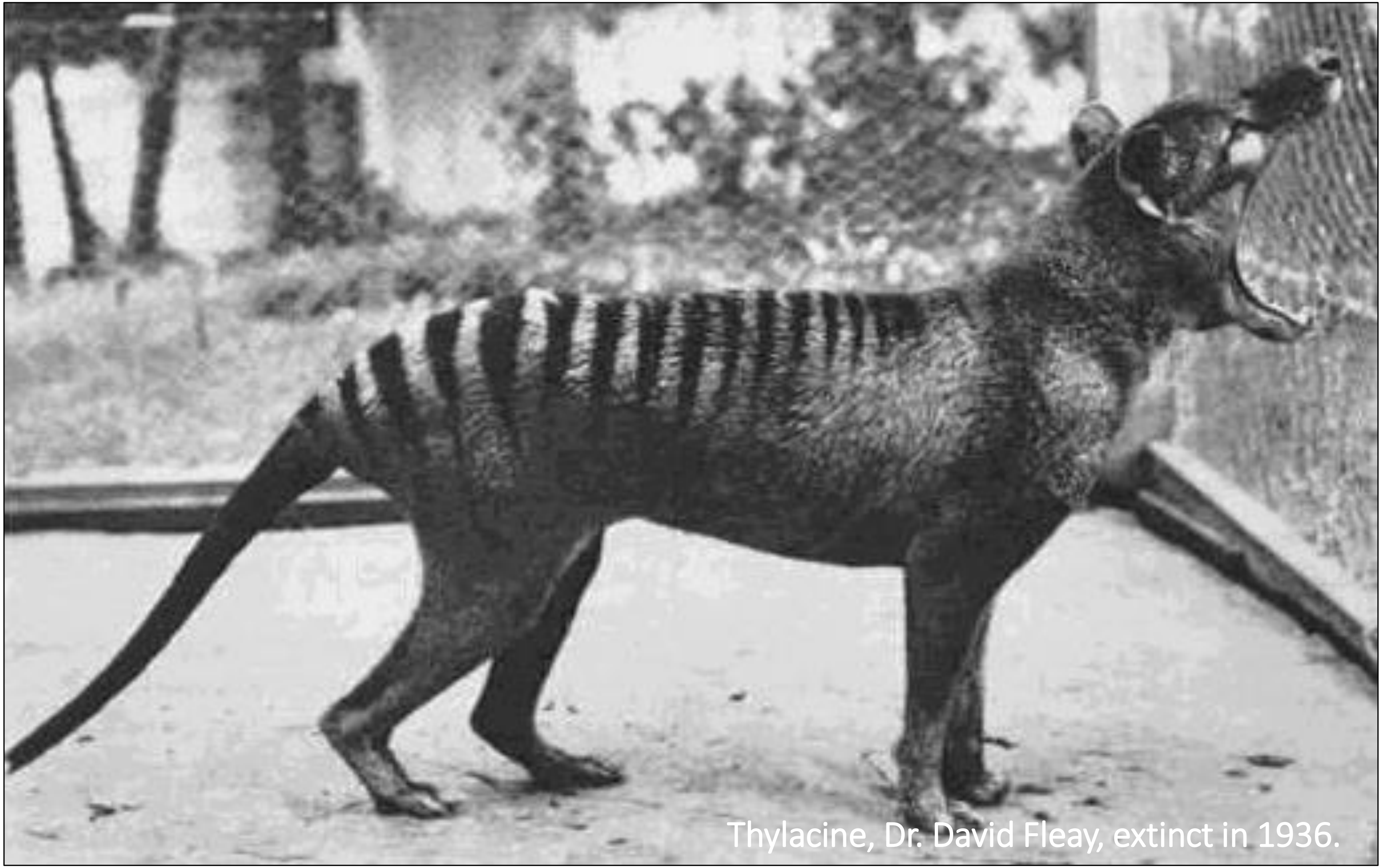


Dataset & Task Generalization on PASCAL VOC



Does the method work on ***legacy*** black and white photos?



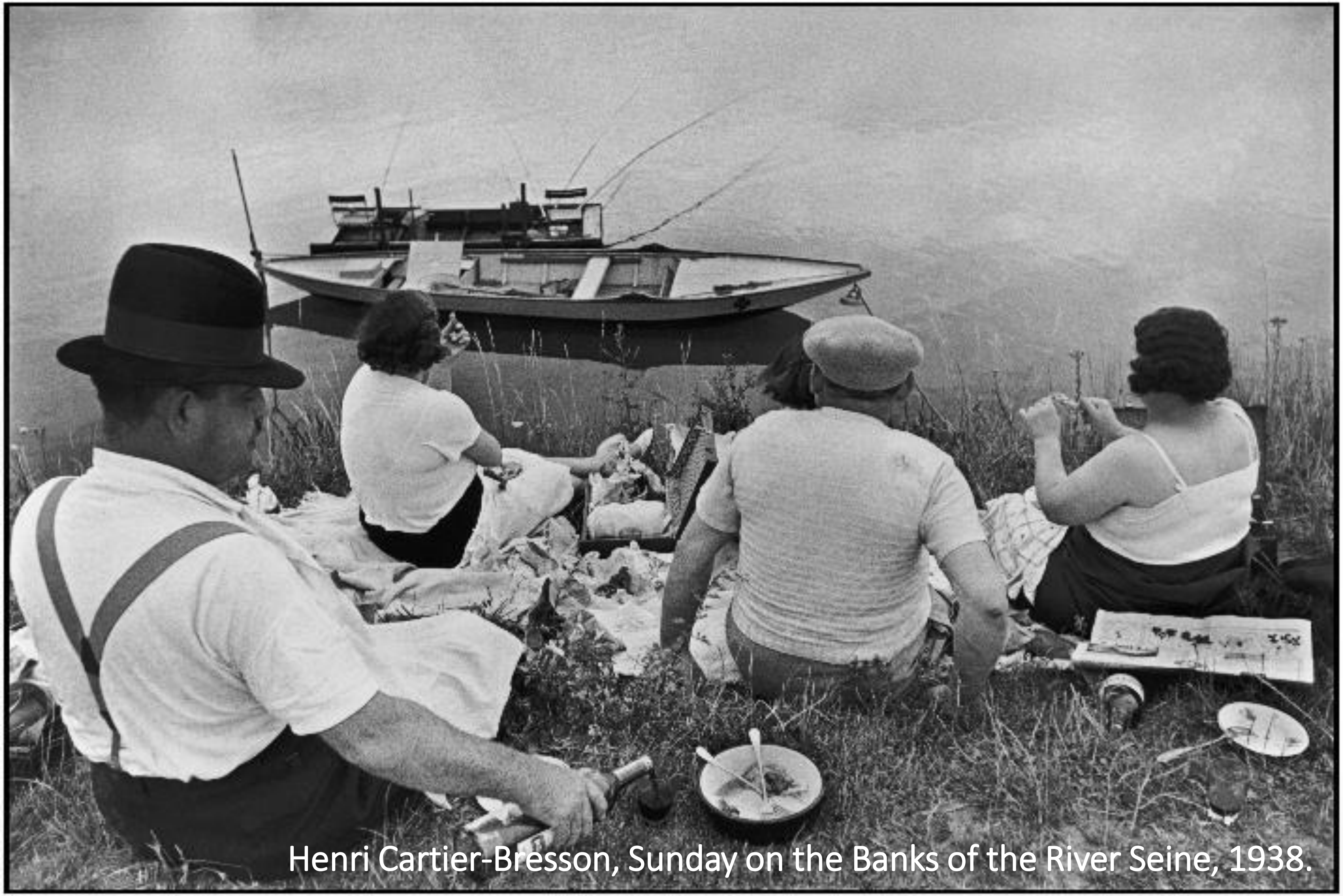


Thylacine, Dr. David Fleay, extinct in 1936.

Thylacine, Dr. David Fleay, extinct in 1936.

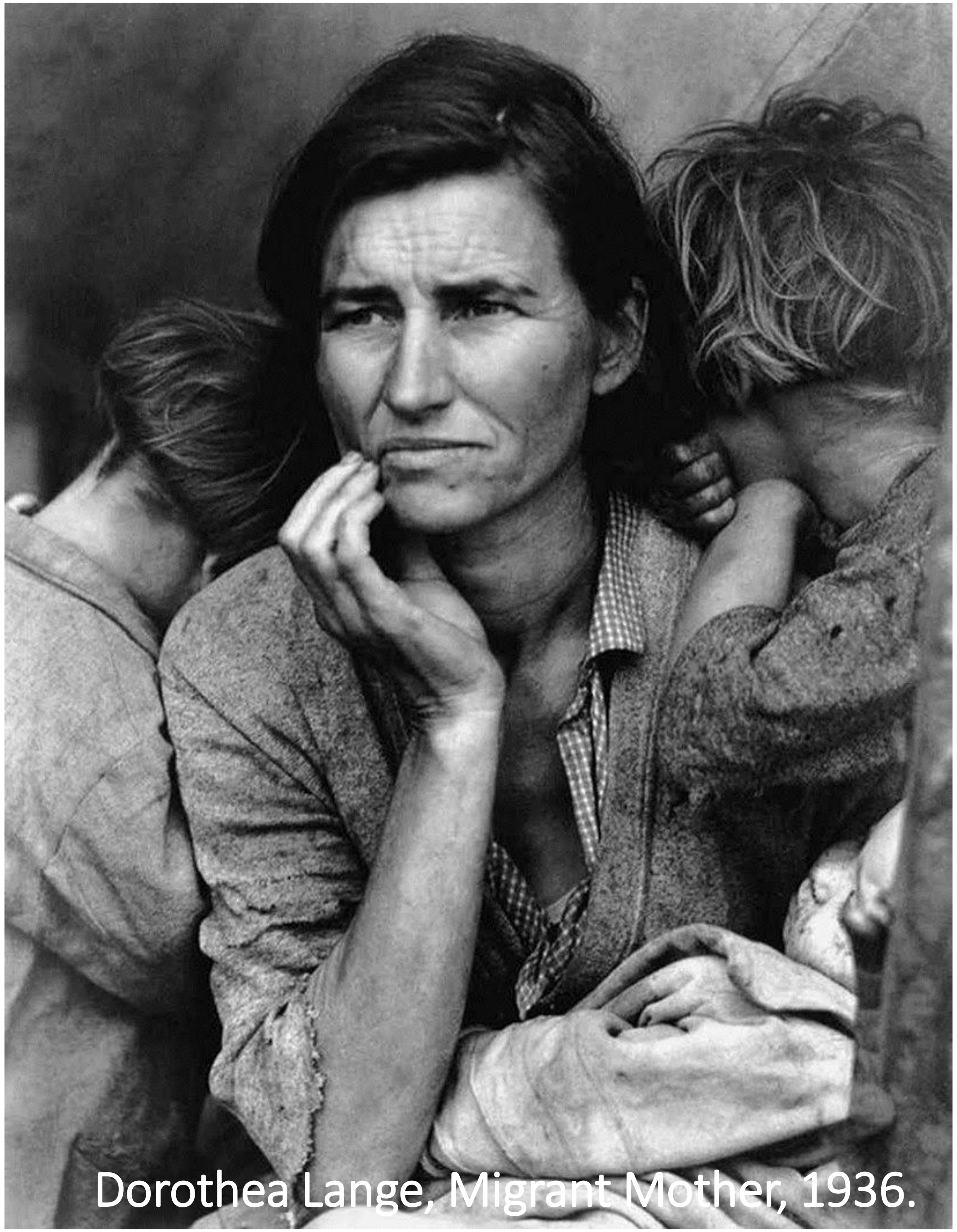
Amateur Family Photo, 1956.

Amateur Family Photo, 1956.

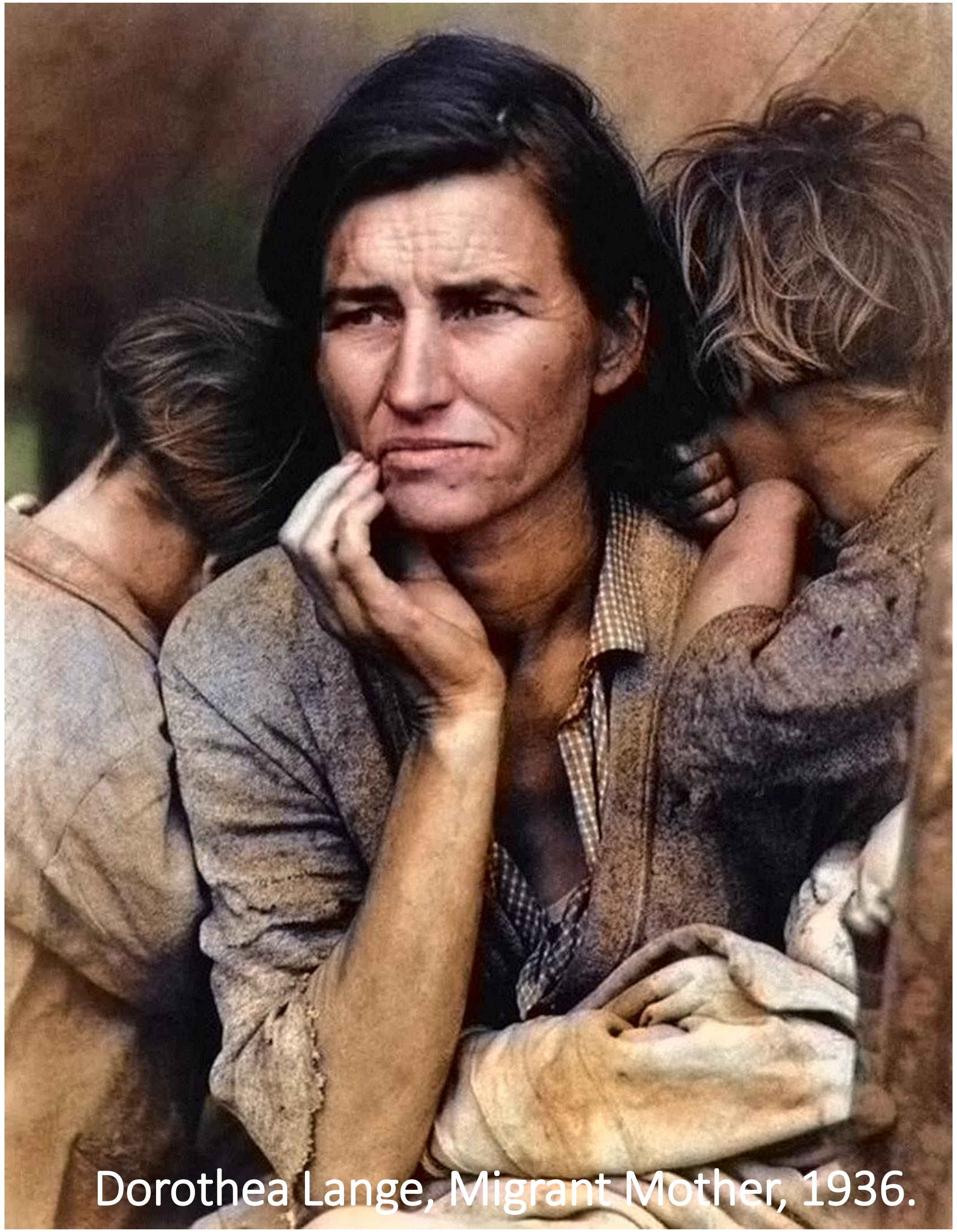


Henri Cartier-Bresson, Sunday on the Banks of the River Seine, 1938.


Henri Cartier-Bresson, Sunday on the Banks of the River Seine, 1938.



Dorothea Lange, Migrant Mother, 1936.



Dorothea Lange, Migrant Mother, 1936.

Additional Information

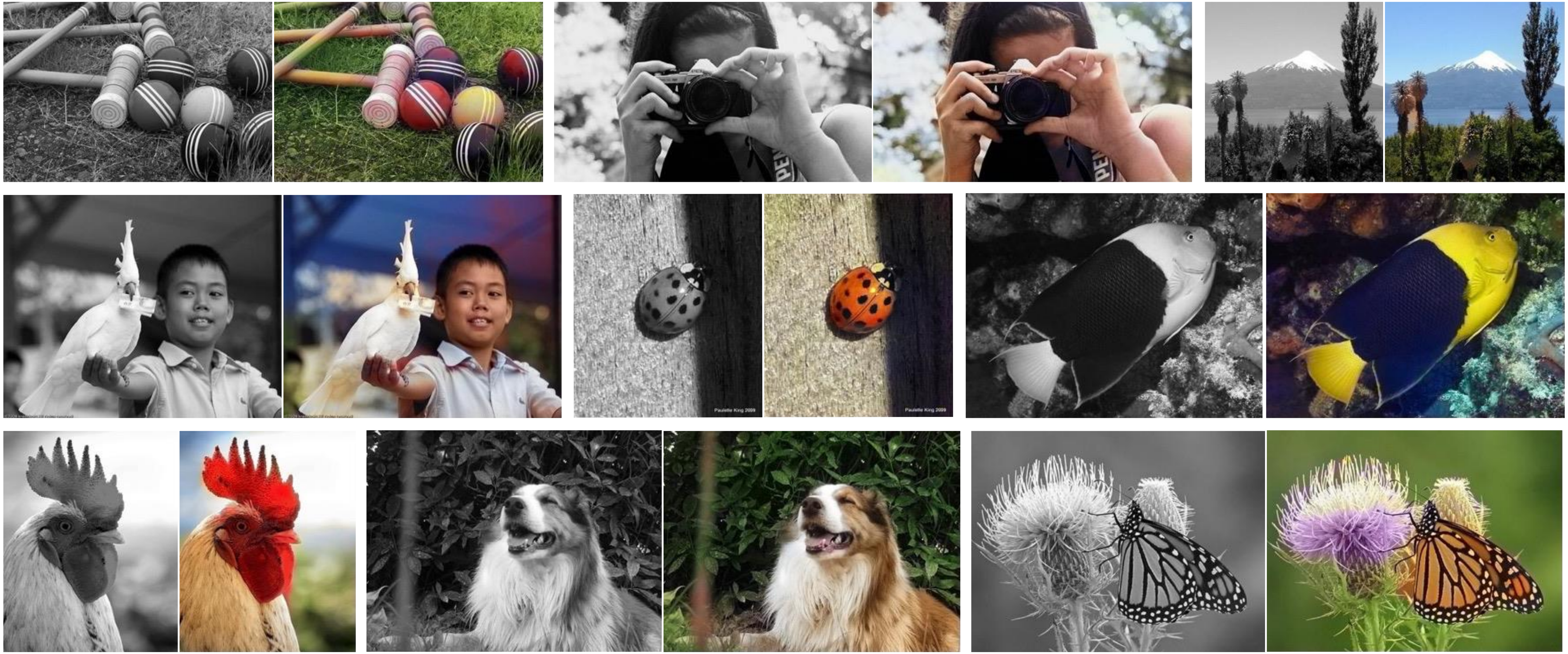
- Demo
  - http://demos.algorithmia.com/colorize-photos/
- Reddit ColorizeBot
  - Type “colorizebot” under any image post
- Code
  - https://github.com/richzhang/colorization
- Website – full paper, user examples, visualizations
  - http://richzhang.github.io/colorization

Lukas Graham – 7 Years

Submitted by Ron Zohar

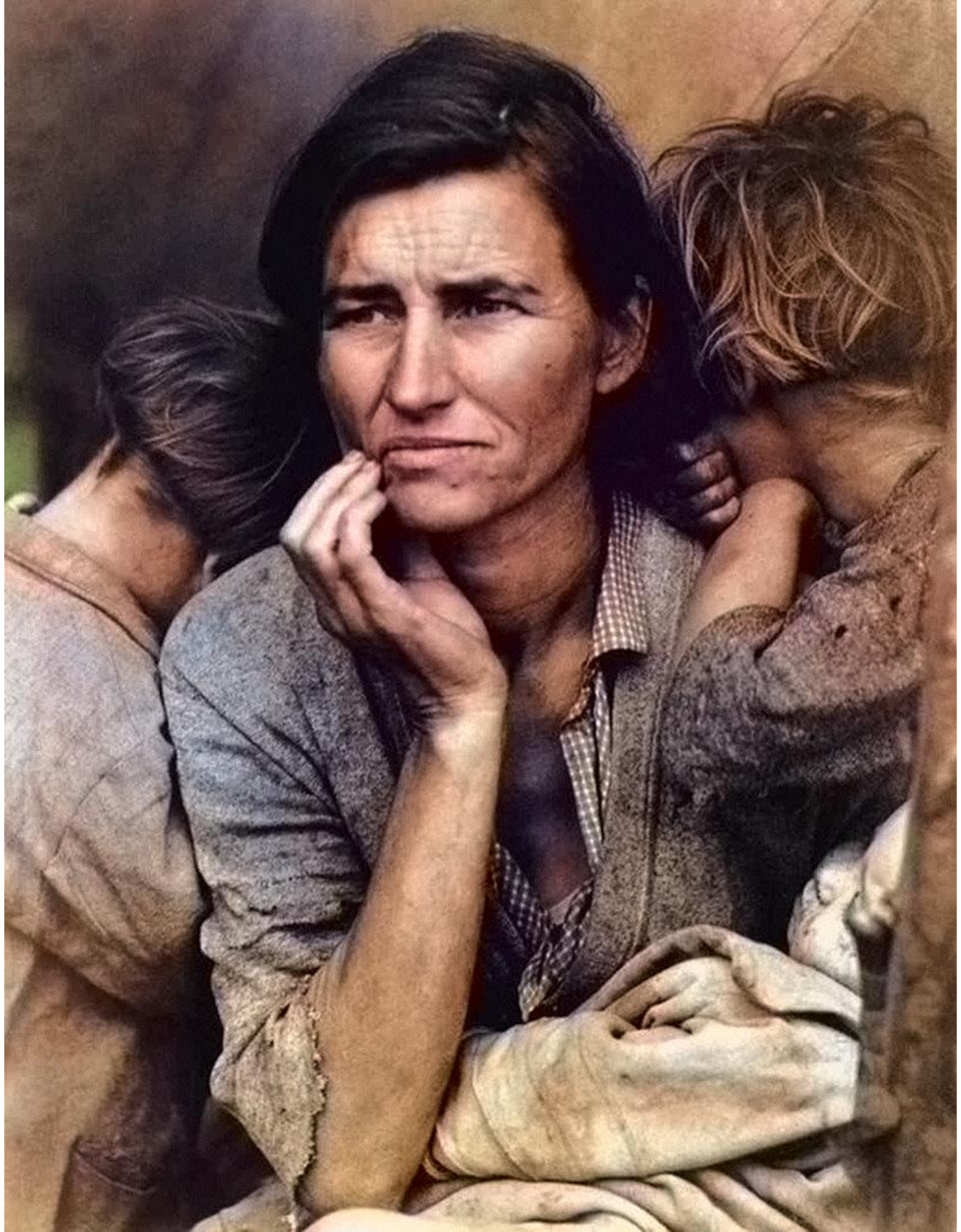
For the full paper, code, and live demo:

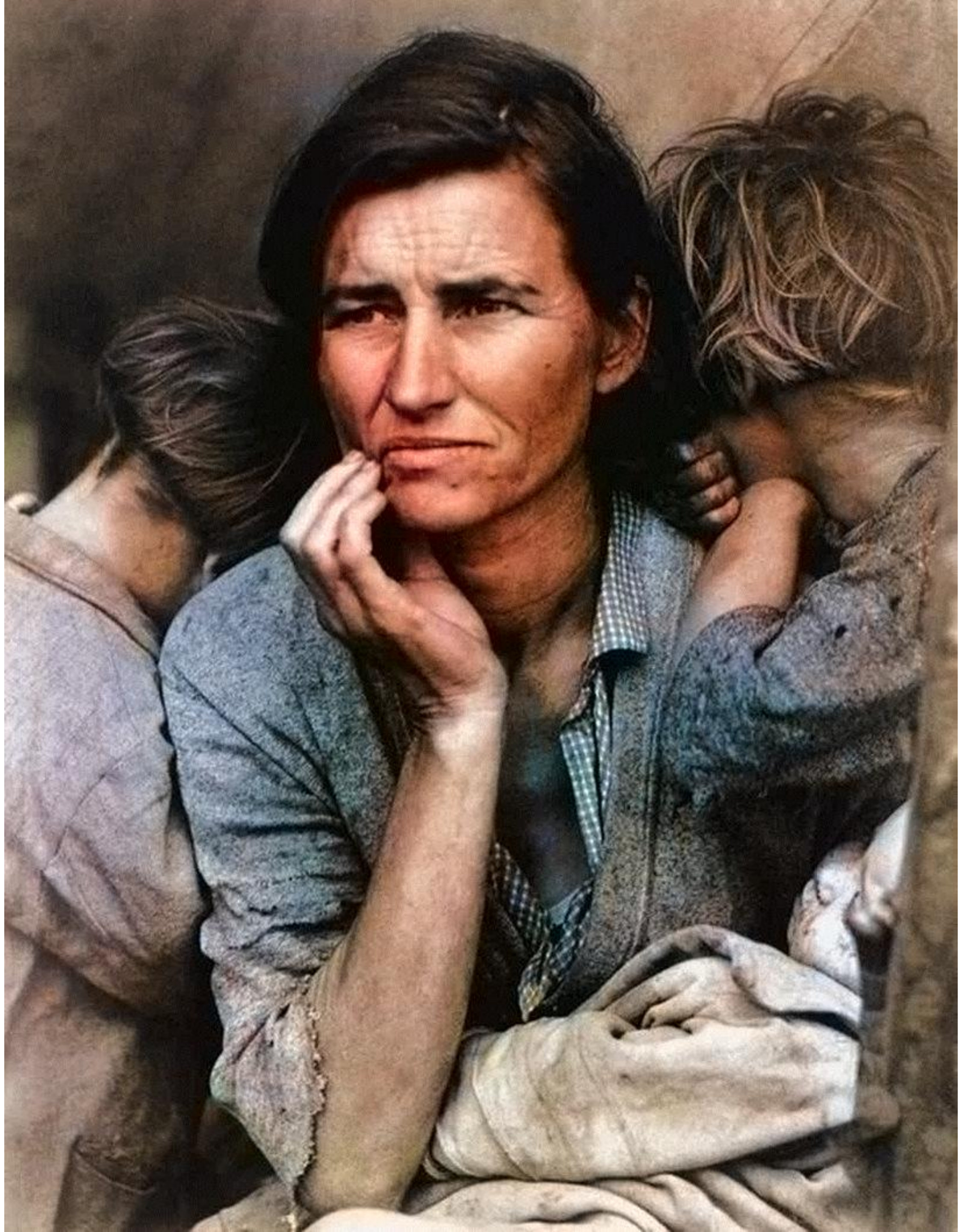
richzhang.github.io/colorization

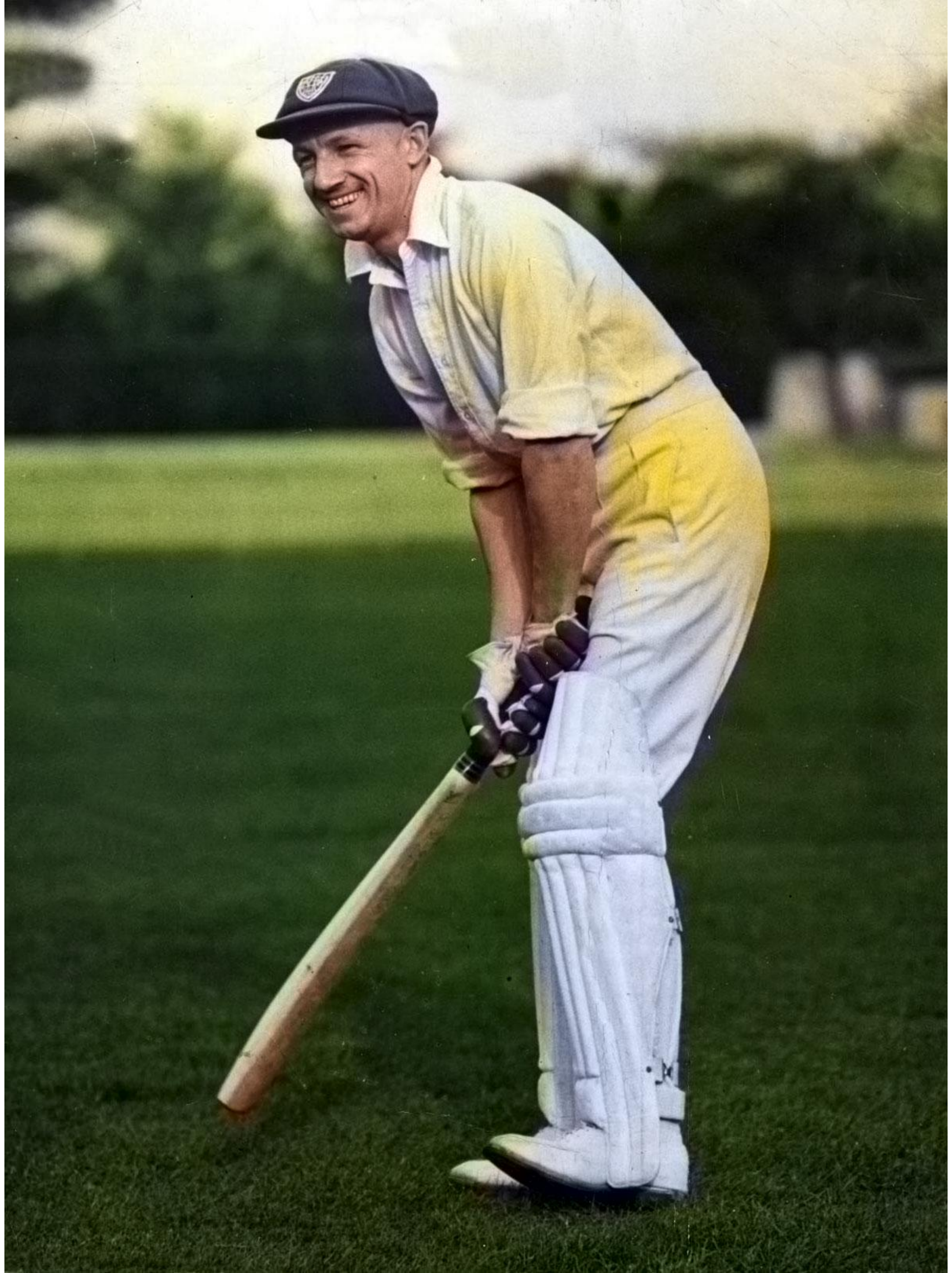


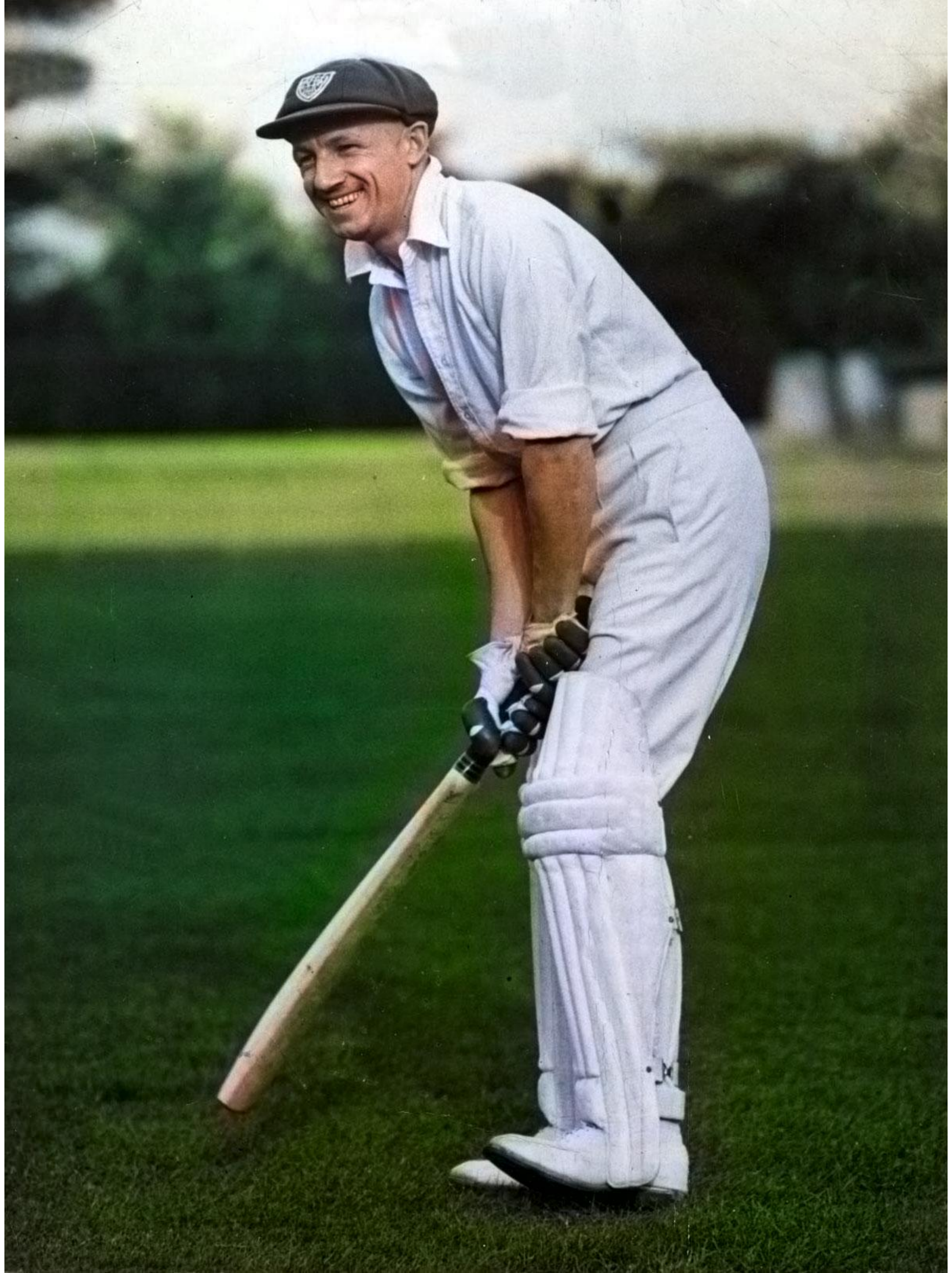
For the full paper, additional examples and our model:
richzhang.github.io/colorization

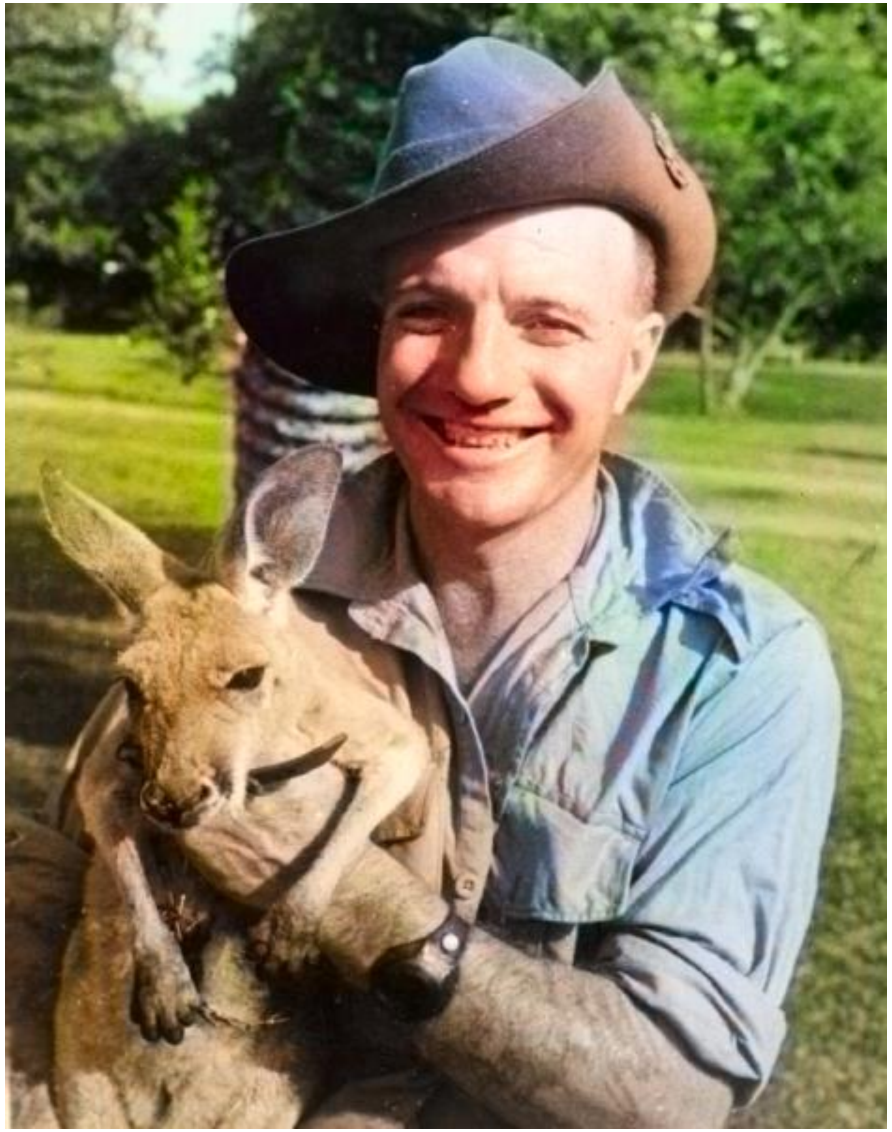
Backup

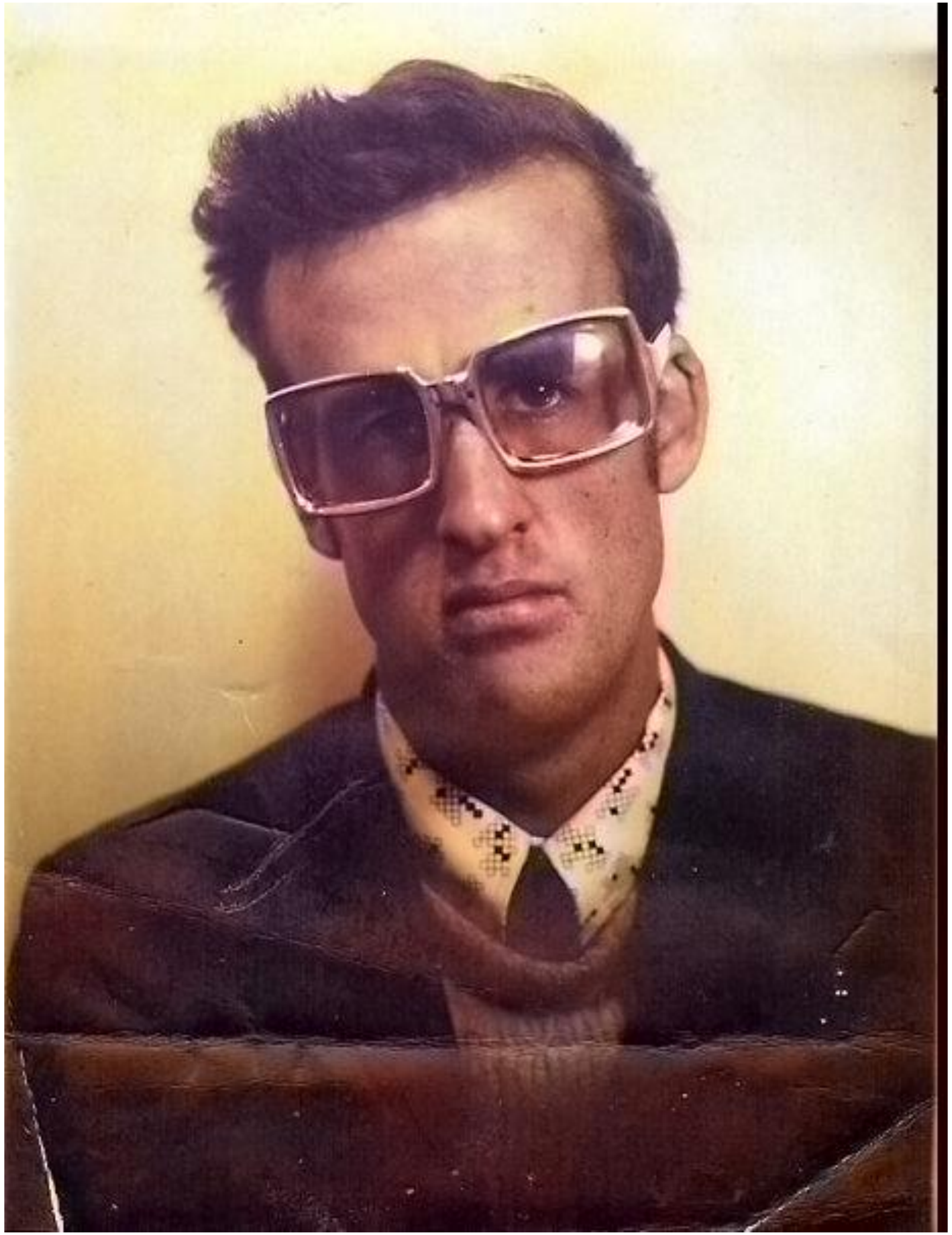


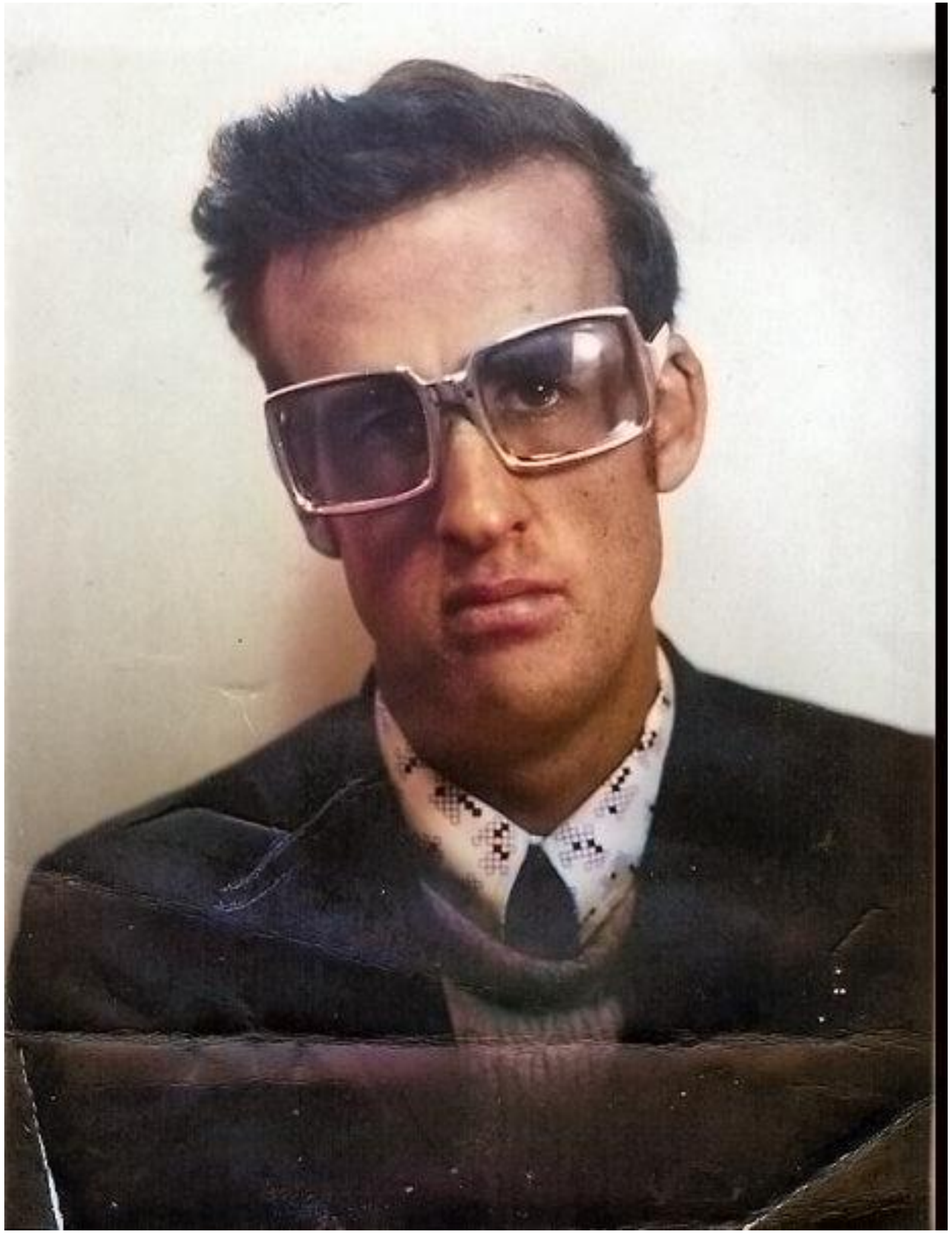


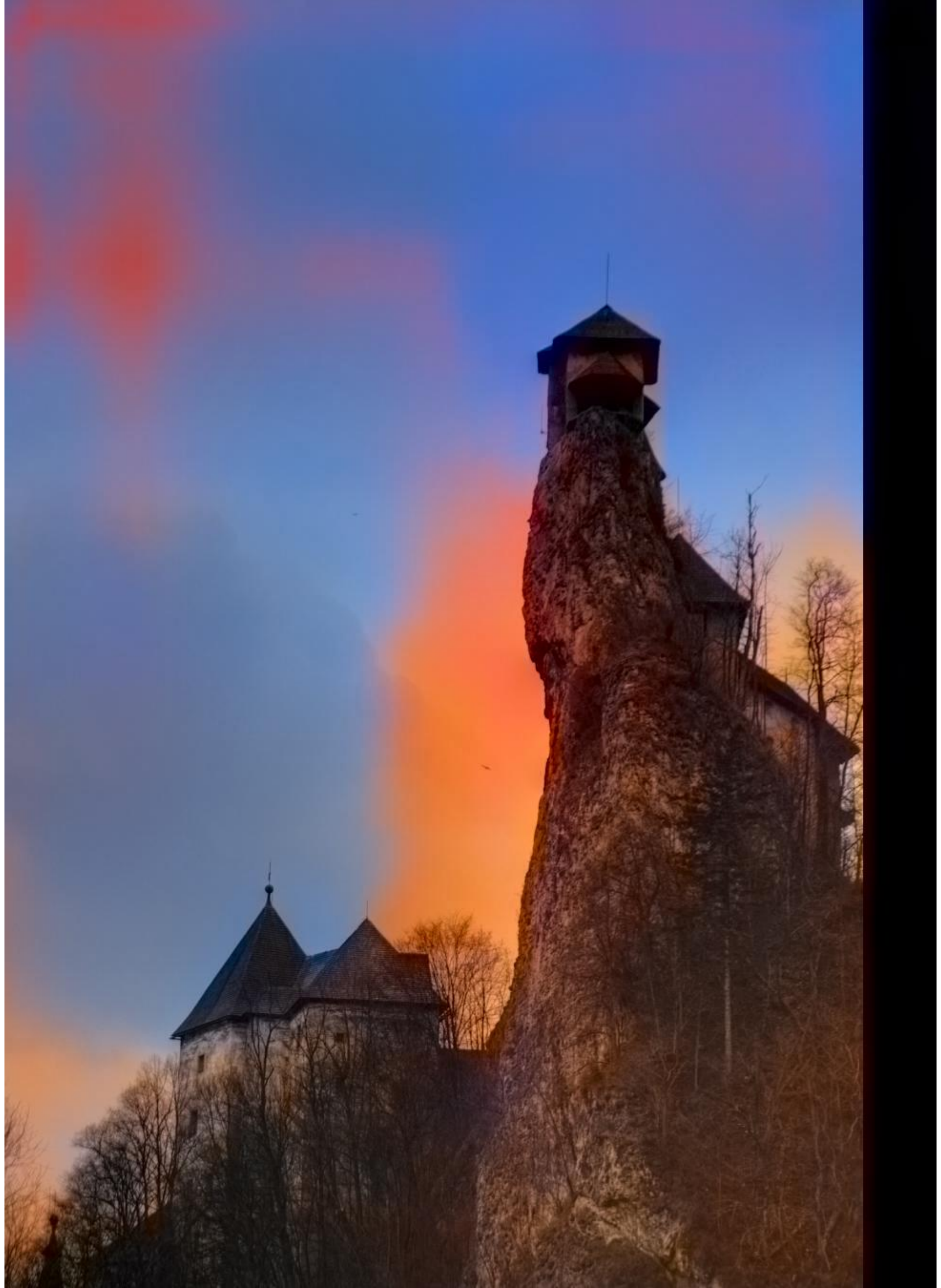


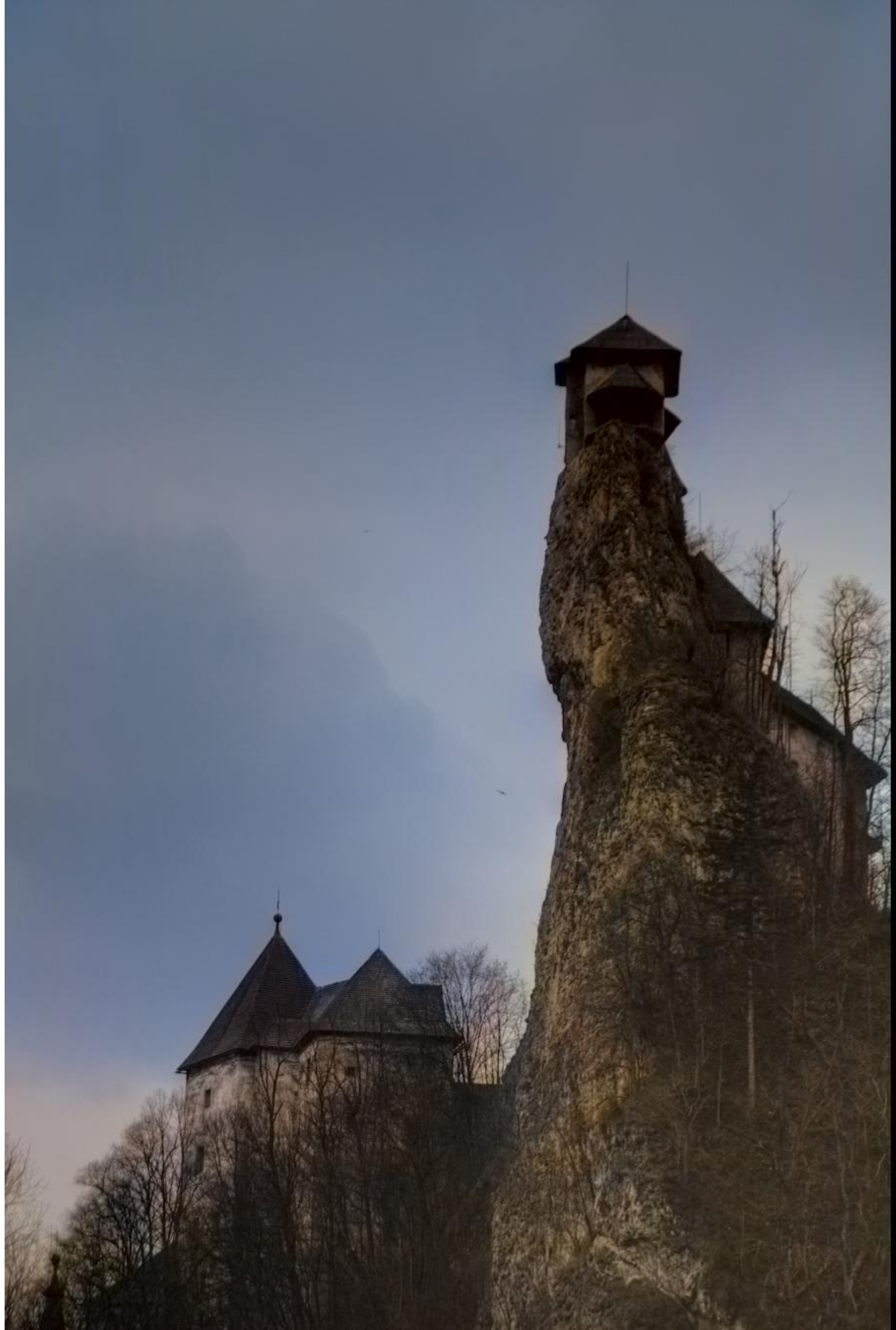






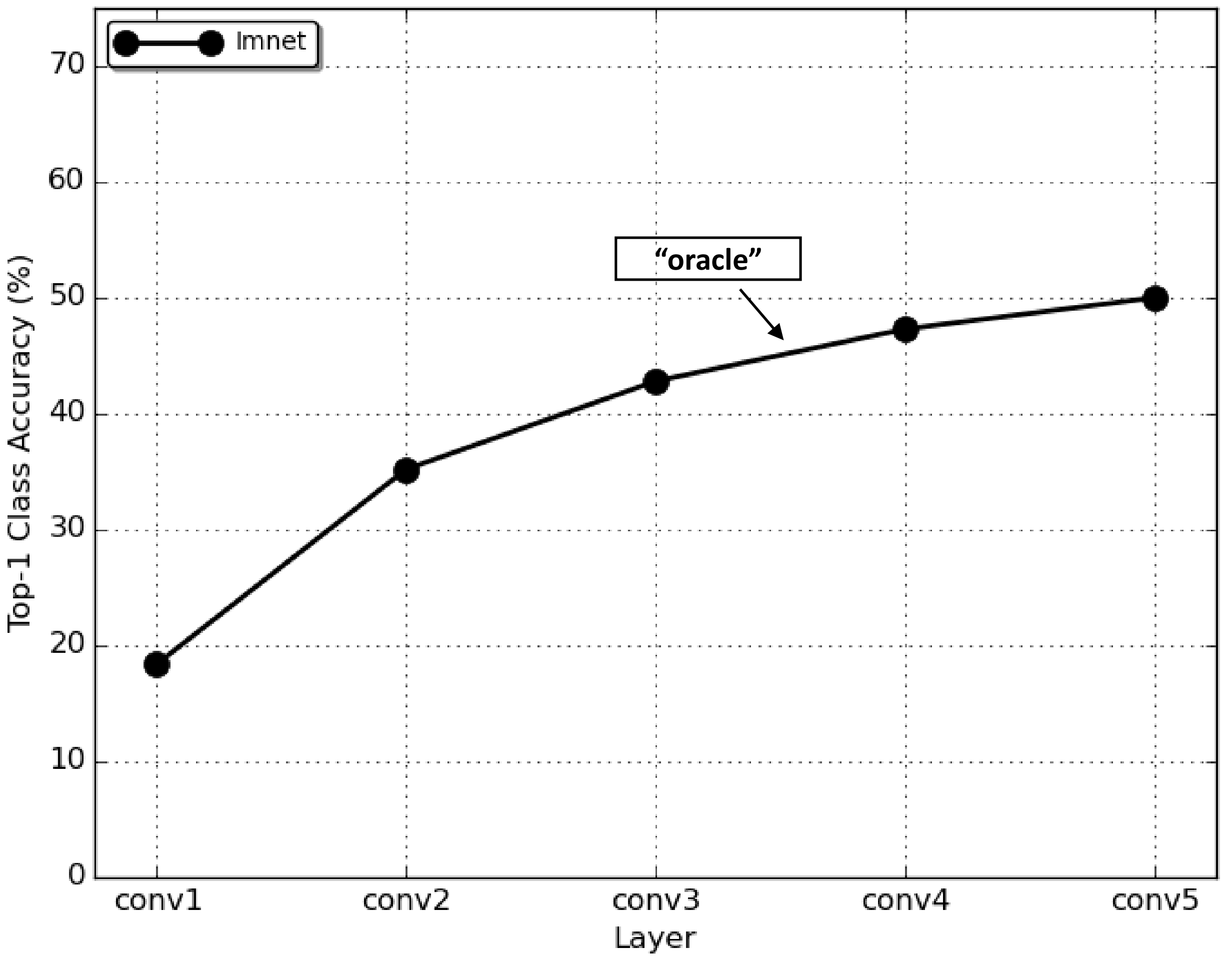






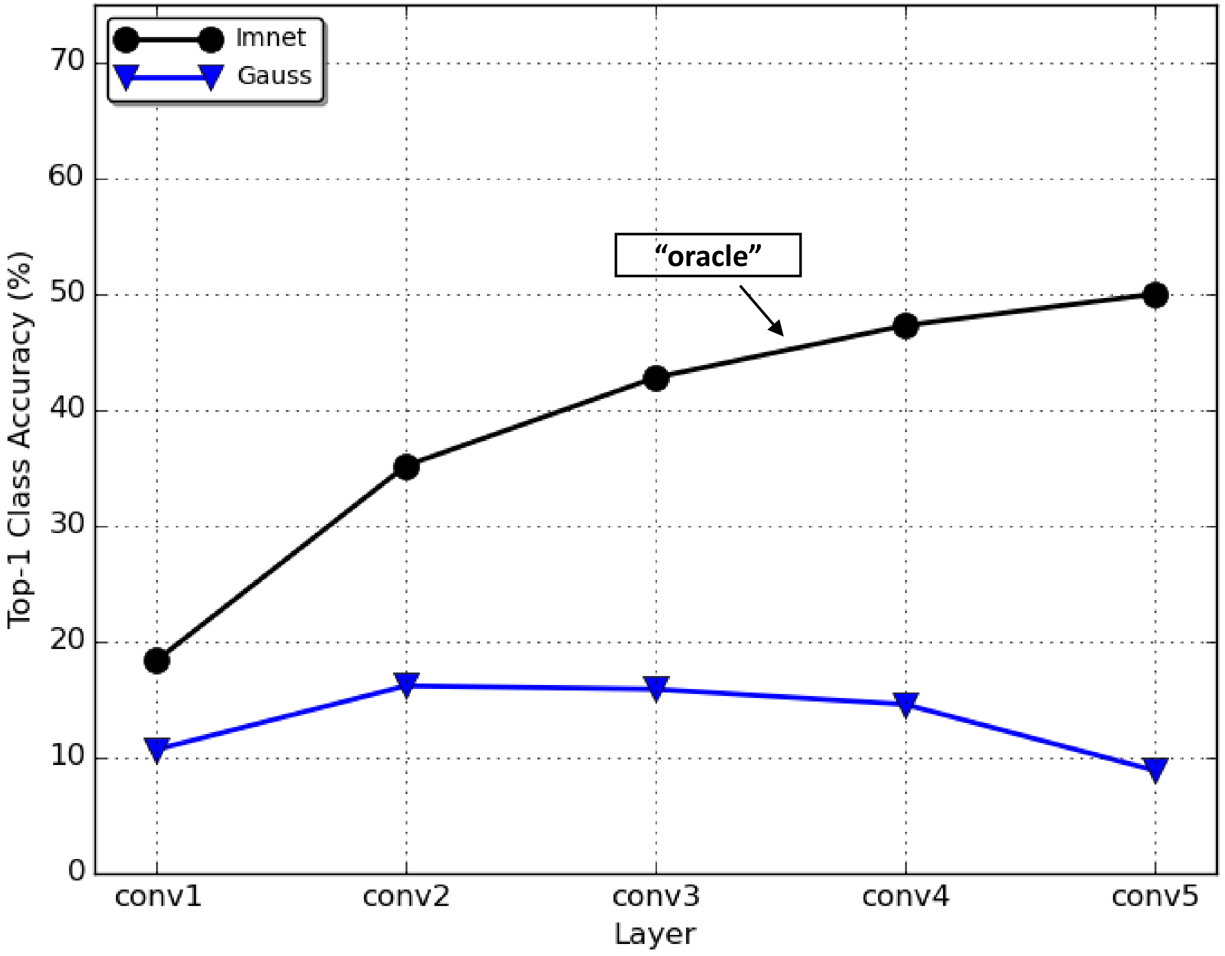
Task Generalization - ILSVRC linear classification

- Directly training on labels provides "oracle"



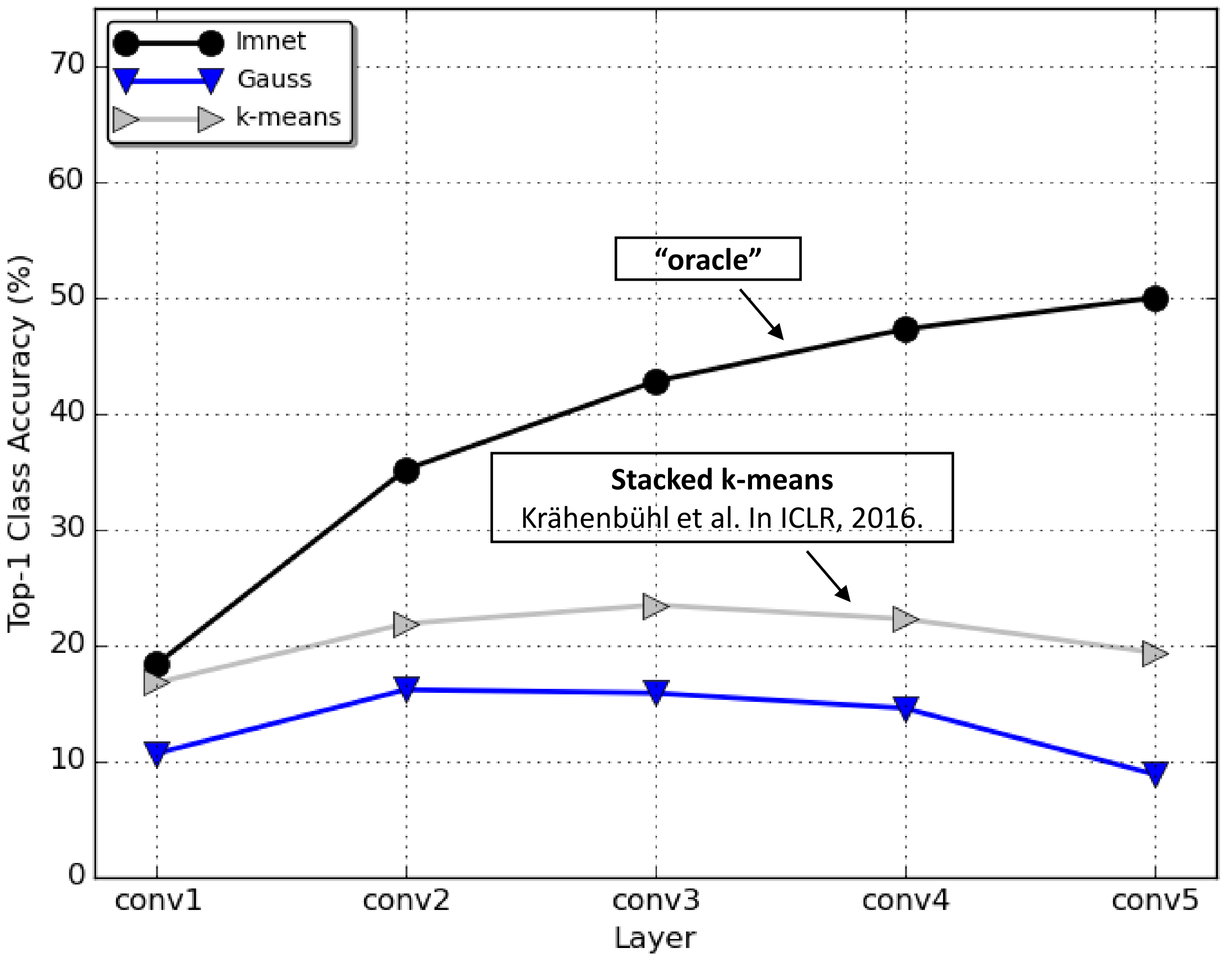
Task Generalization - ILSVRC linear classification

- Directly training on labels provides "oracle"



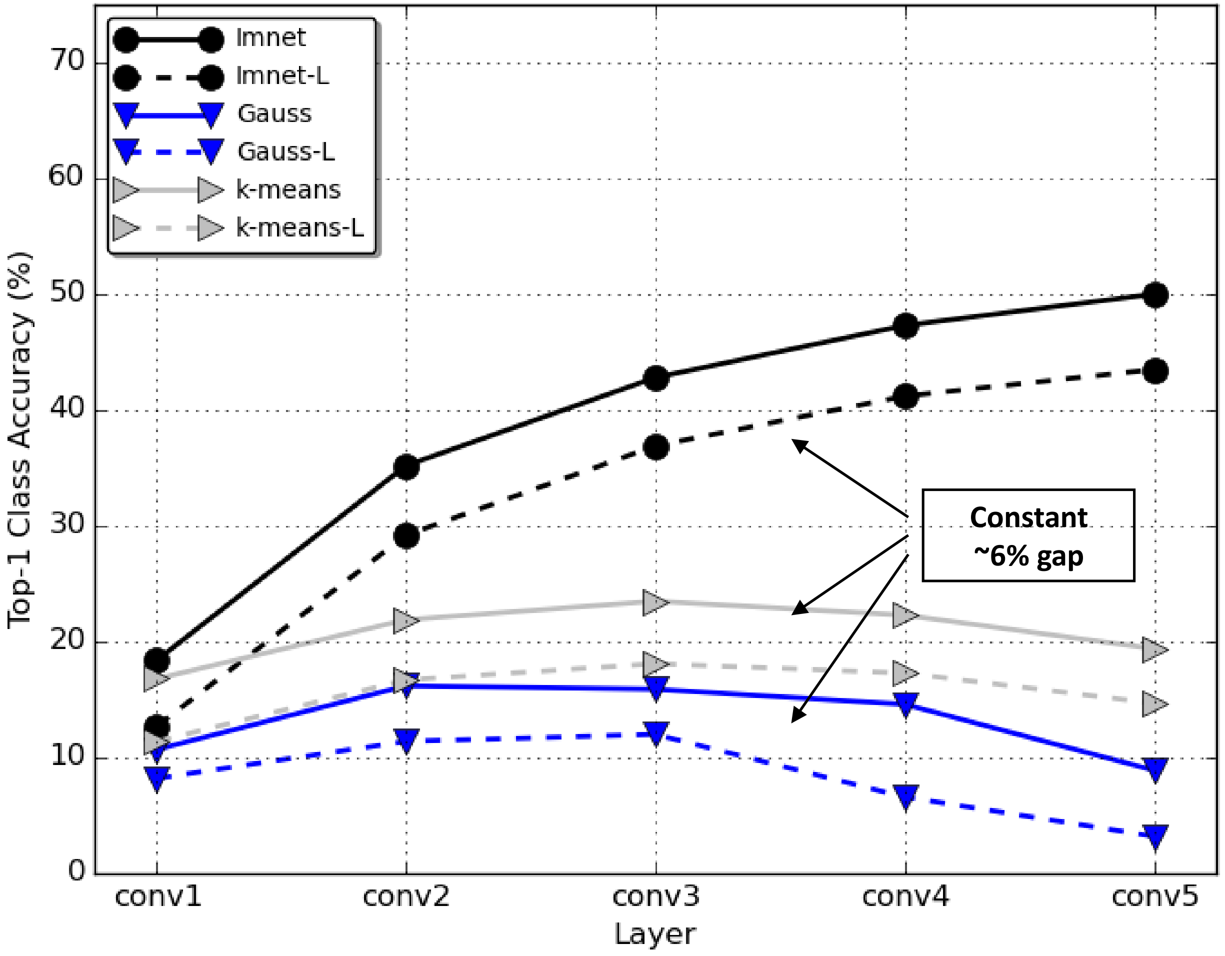
Task Generalization - ILSVRC linear classification

- Directly training on labels provides “oracle”
- Stacked k-means provides strong baseline



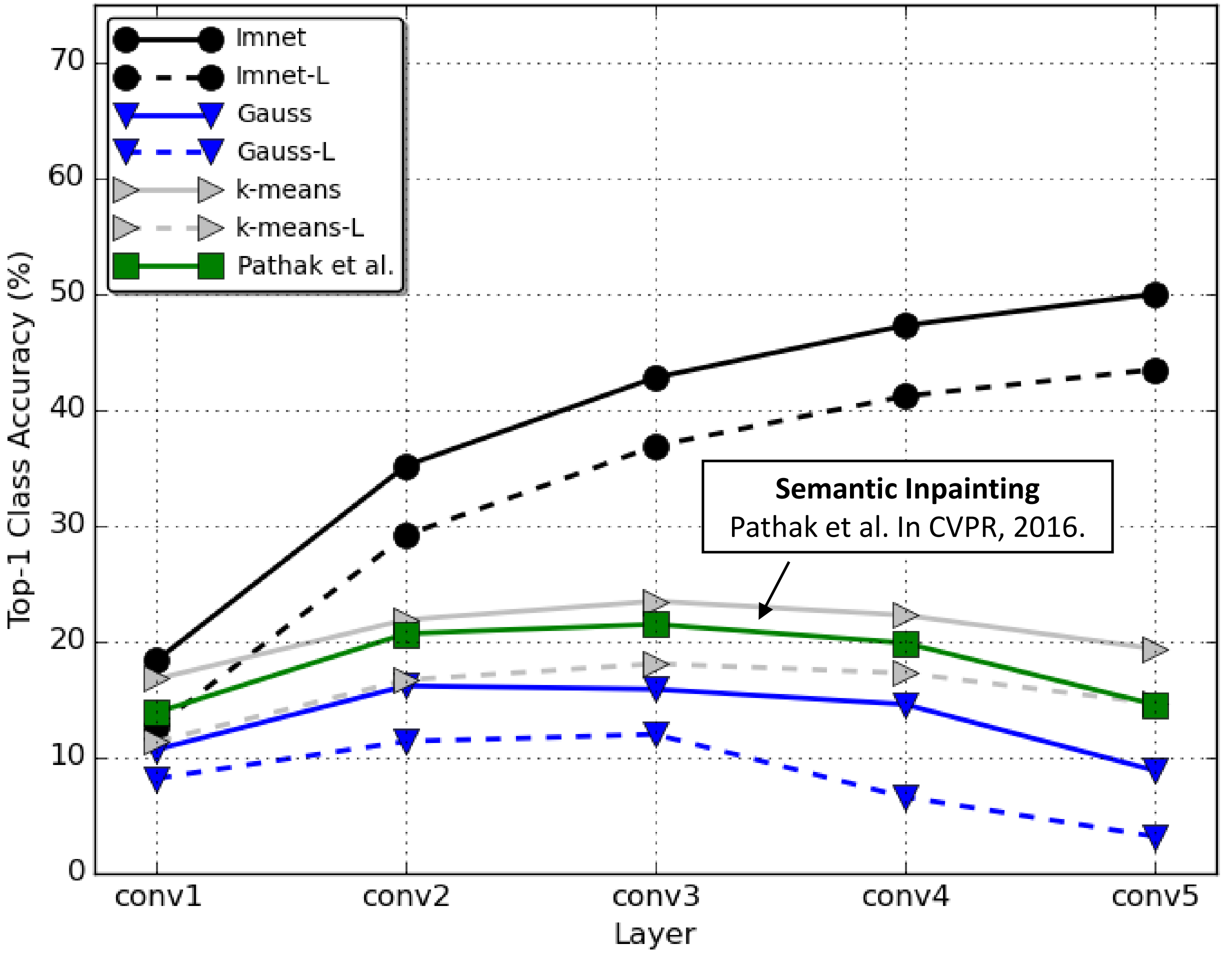
Task Generalization - ILSVRC linear classification

- Directly training on labels provides “oracle”
- Stacked k-means provides strong baseline
- Constant 6% gap from grayscale handicap



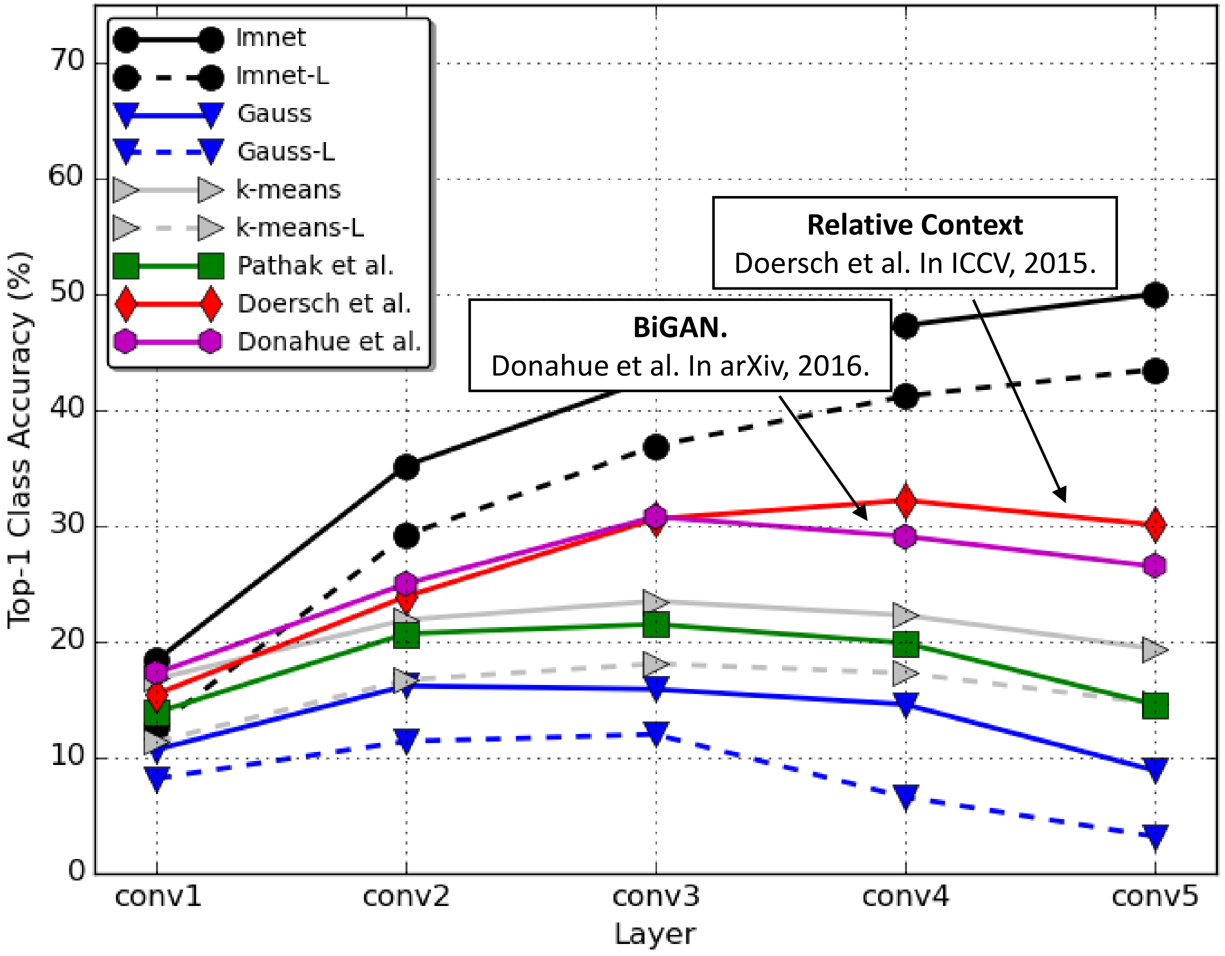
Task Generalization - ILSVRC linear classification

- Directly training on labels provides “oracle”
- Stacked k-means provides strong baseline
- Constant 6% gap from grayscale handicap



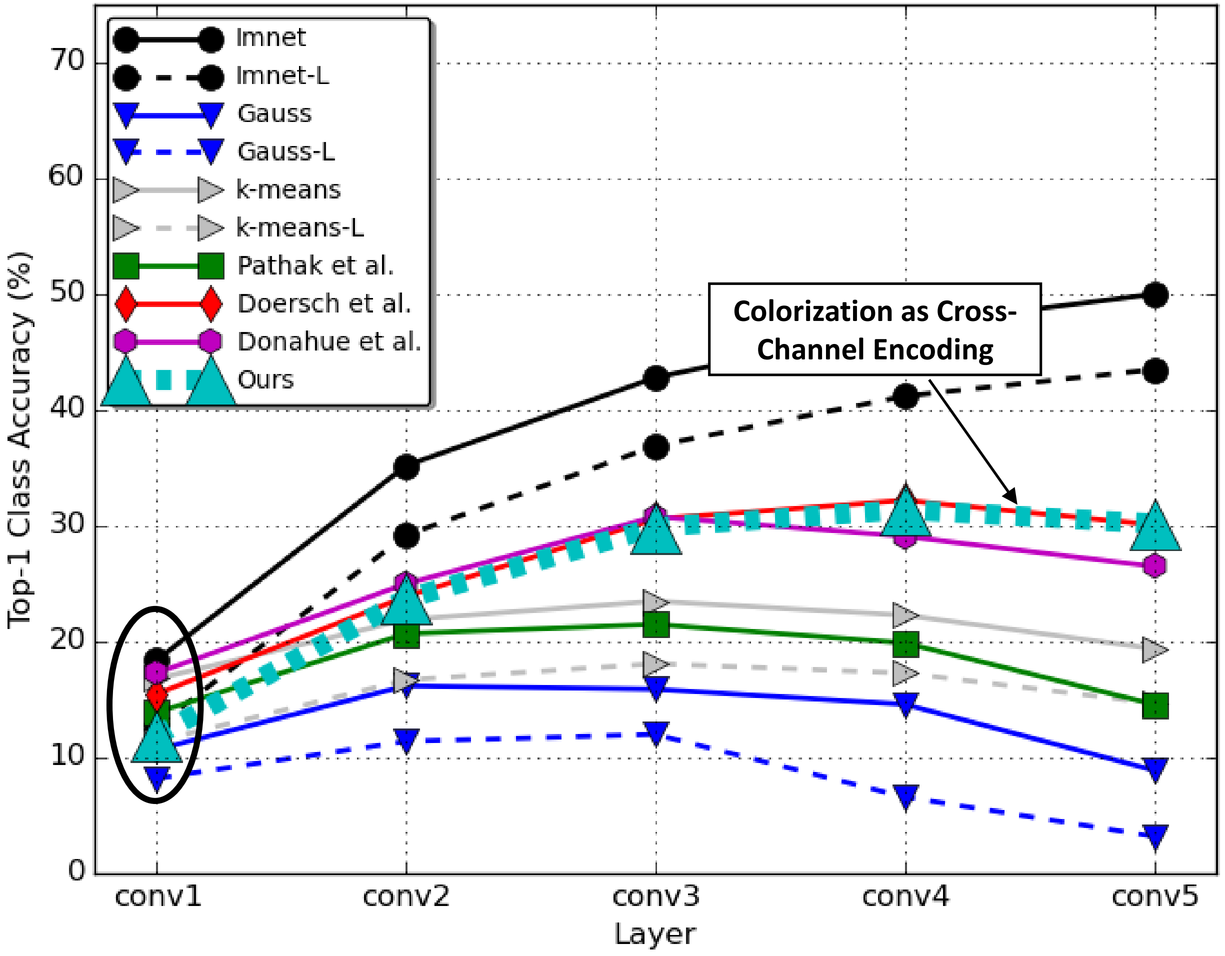
Task Generalization - ILSVRC linear classification

- Directly training on labels provides “oracle”
- Stacked k-means provides strong baseline
- Constant 6% gap from grayscale handicap



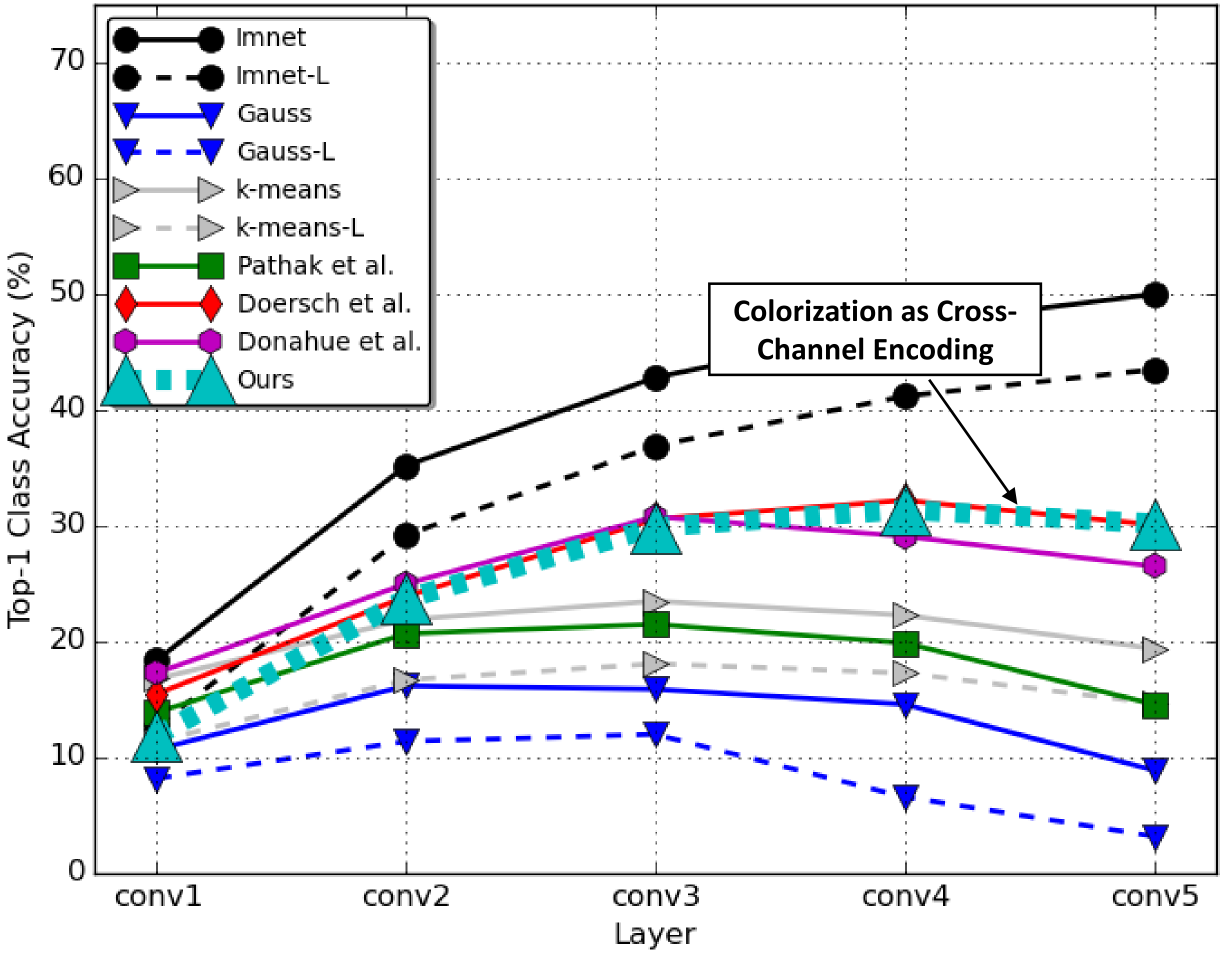
Task Generalization - ILSVRC linear classification

- Directly training on labels provides “oracle”
- Stacked k-means provides strong baseline
- Constant 6% gap from grayscale handicap
    - Our *conv1* suffers from input handicap



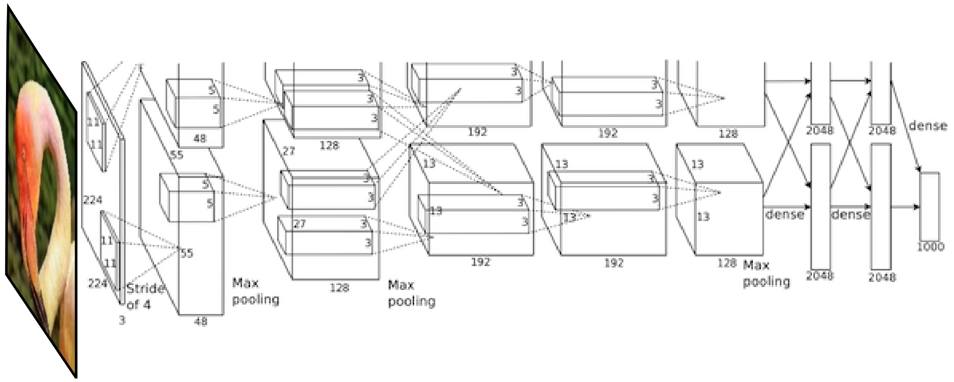
Task Generalization - ILSVRC linear classification

- Directly training on labels provides “oracle”
- Stacked k-means provides strong baseline
- Constant 6% gap from grayscale handicap
  - Our *conv1* suffers from input handicap
- Our *conv2-5* performs competitively throughout



Predicting Labels from Data

Image **Label**

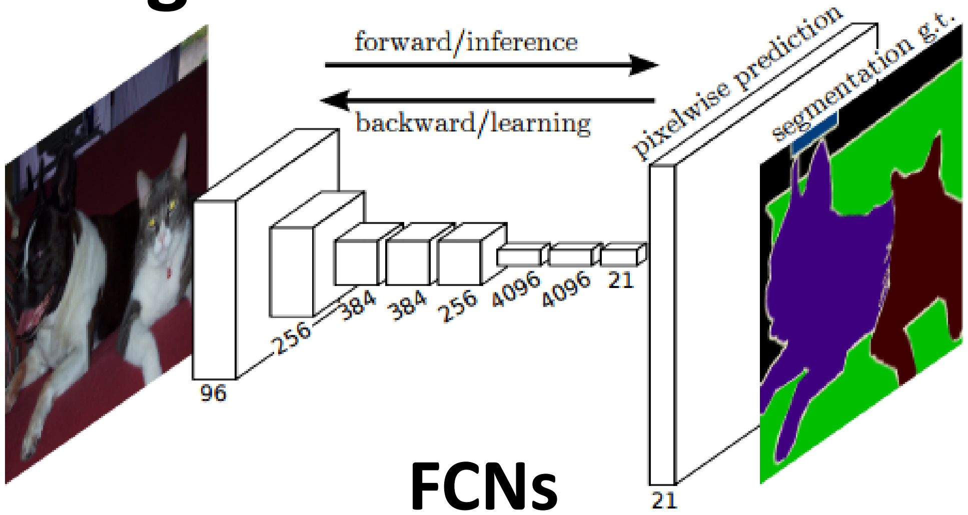


“flamingo”

Alexnet

Krizhevsky et al. In *NIPS*, 2012.

Image **Dense Labels**



FCNs

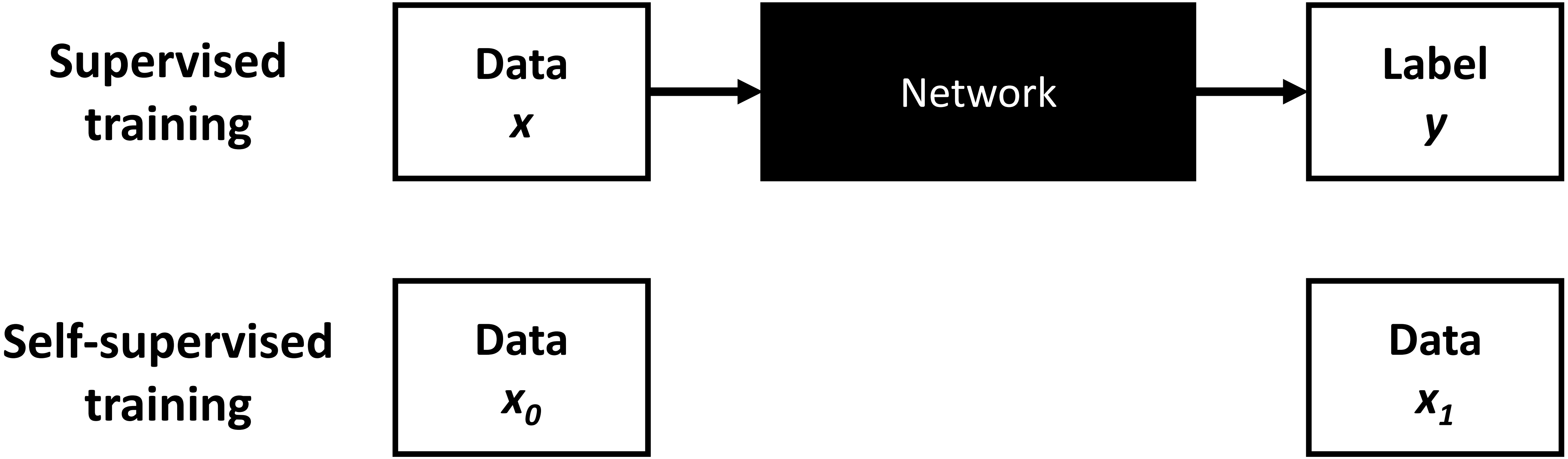
Long et al. In *CVPR*, 2015.

Data *x* → Network → **Label *y***

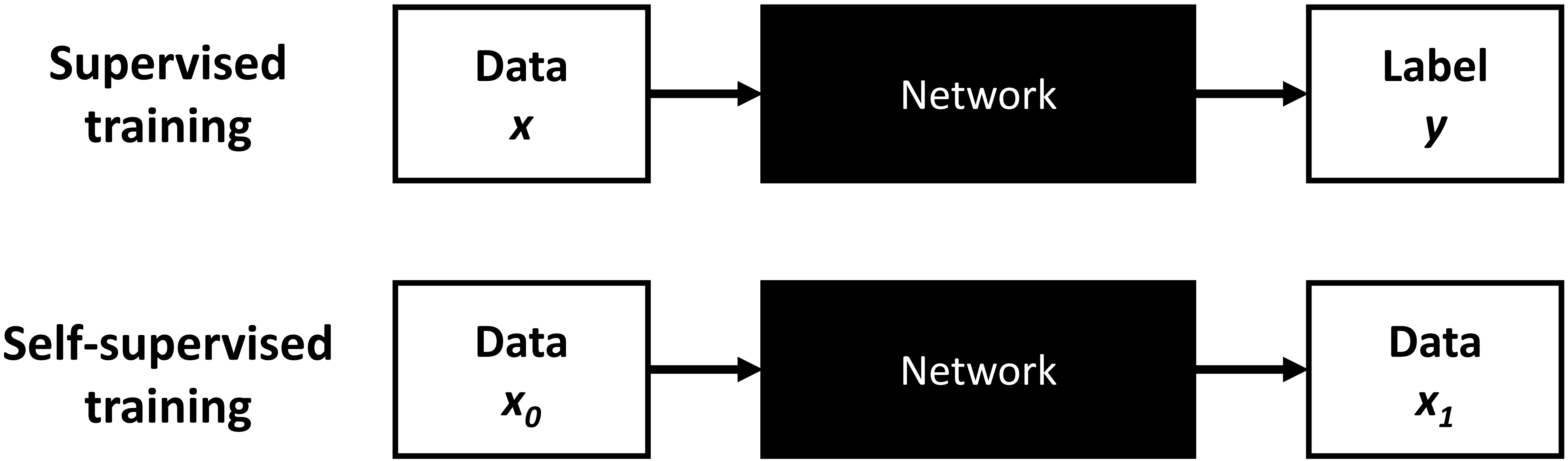
Predicting Labels from Data

Supervised training

Predicting Data from Data



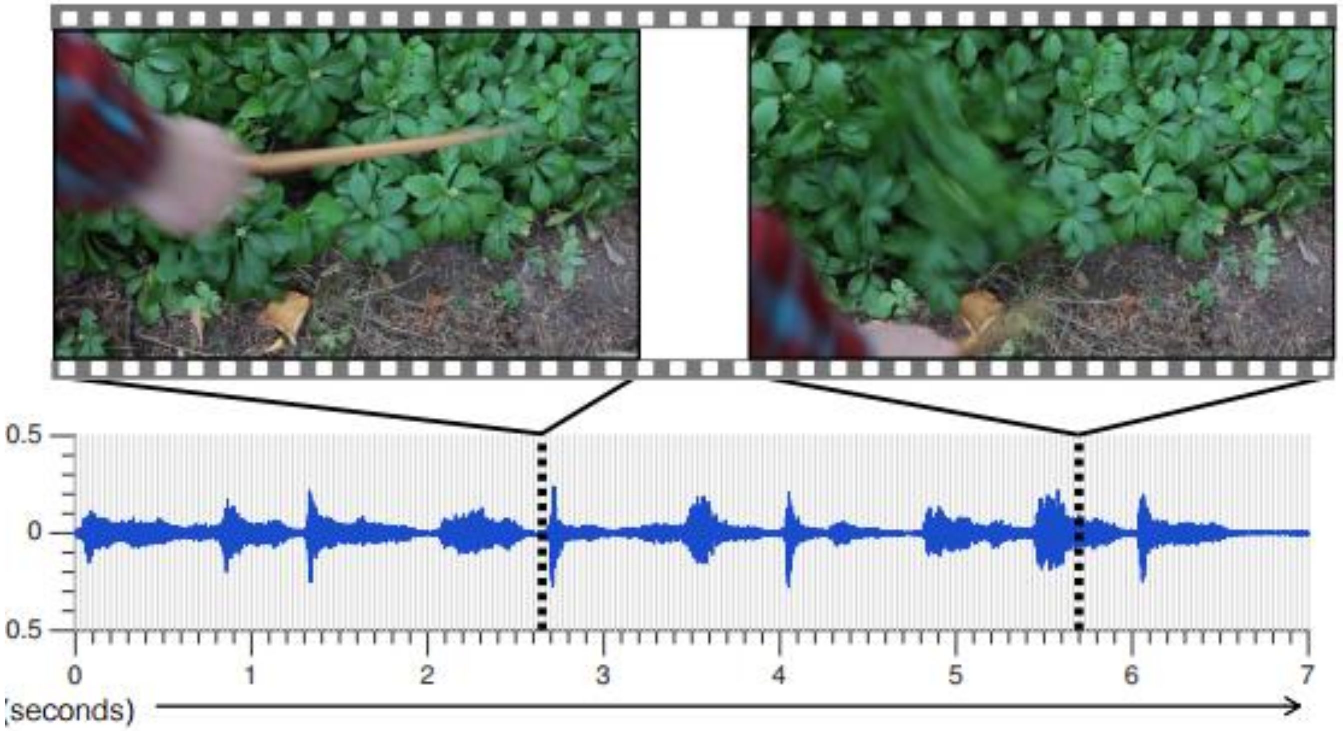
Predicting Data from Data



Visually Indicated Sounds

Sound

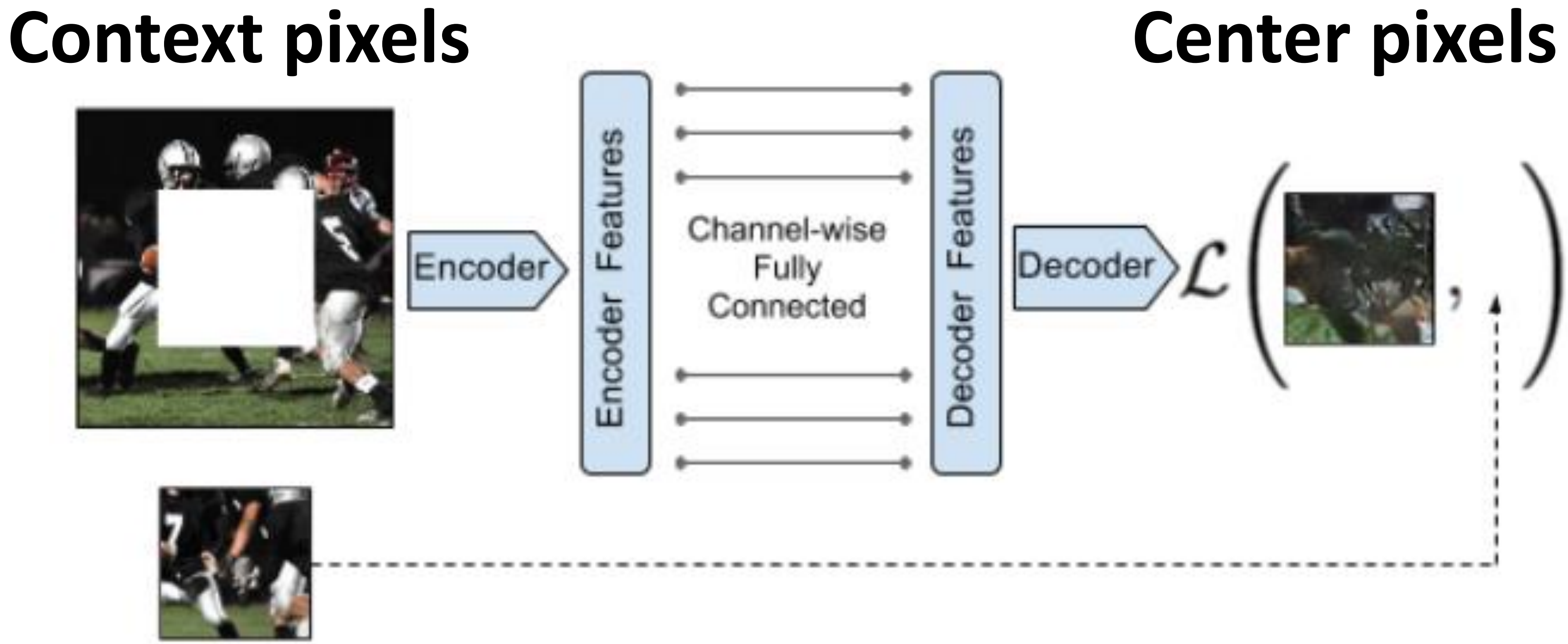
Video



Owens et al. **Visually Indicated Sounds.** In *CVPR*, 2016.

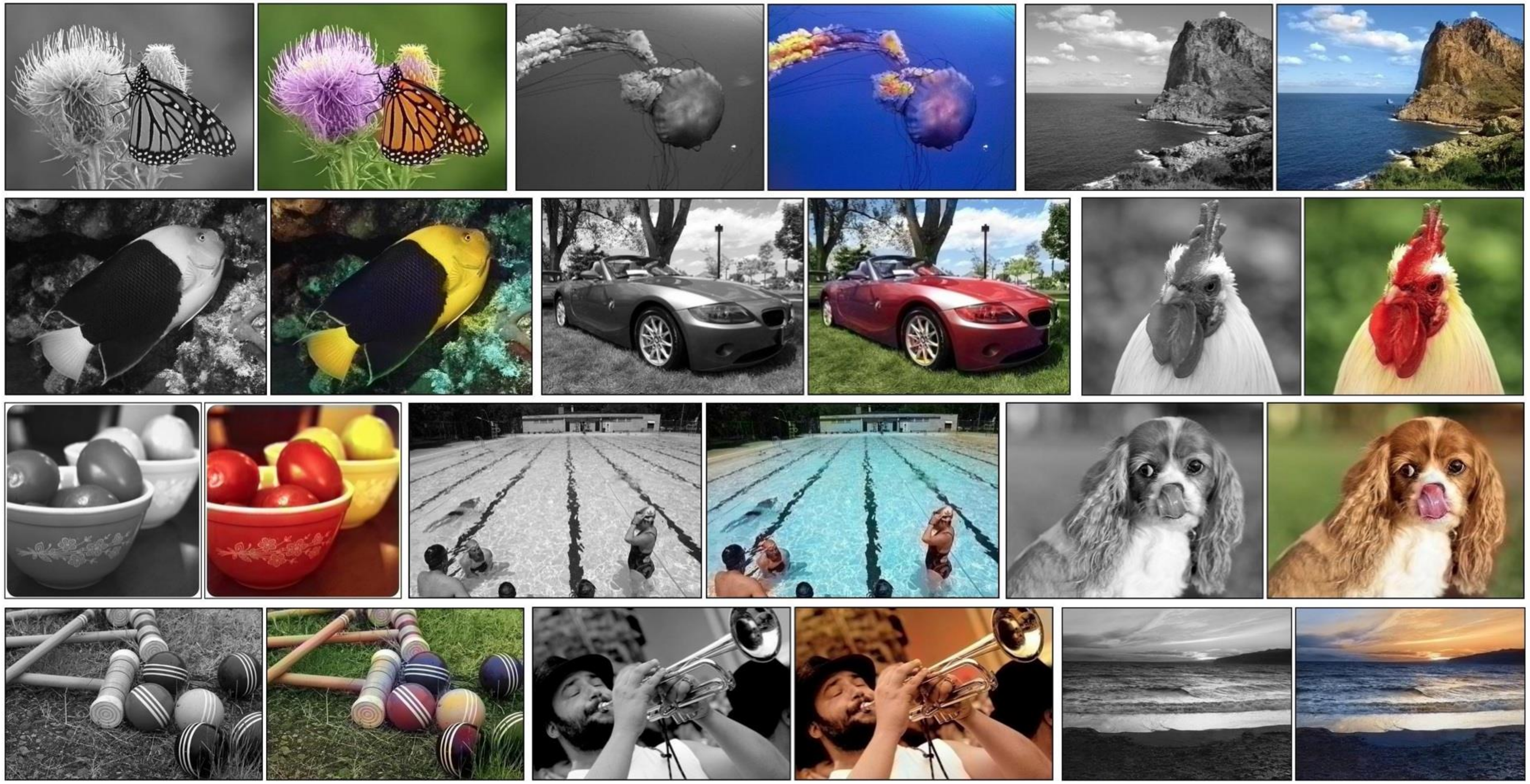
Video → Network → Sound

Context Encoders



Pathak et al. **Context Encoders: Feature Learning by Inpainting**. In *CVPR*, 2016.

Context pixels → Network → Center pixels



Colourful Image Colouri~~z~~sation

Richard Zhang, Phillip Isola, Alexei A. Efros

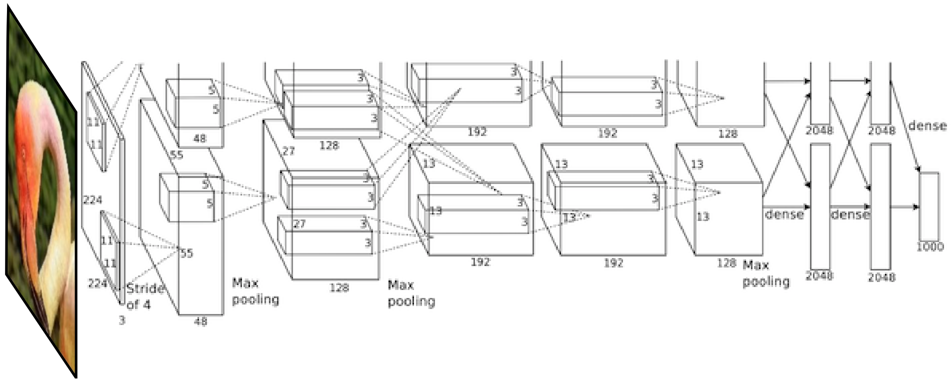
In ArXiv, March 2016.

richzhang.github.io/colorization

Predicting Labels from Data

Image

Label



“flamingo”

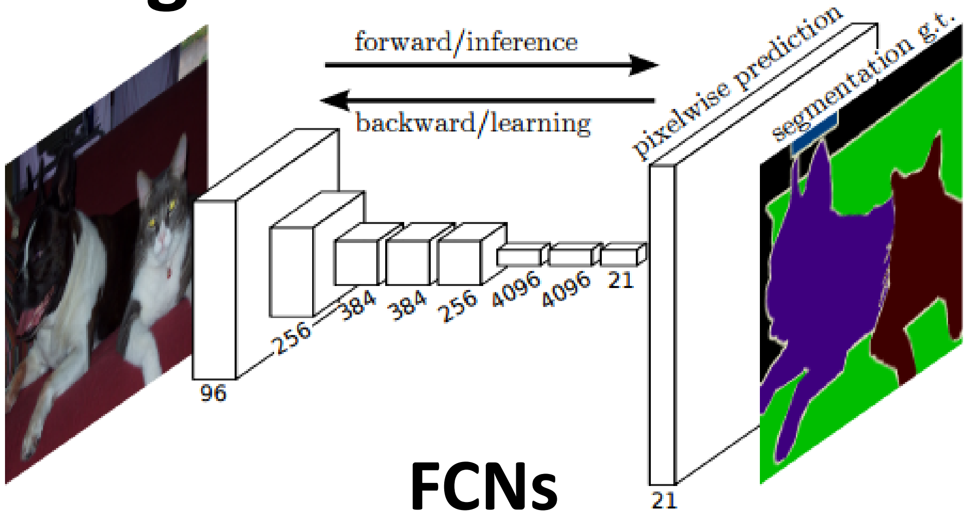
Alexnet

Krizhevsky et al. In *NIPS*, 2012.

Image → **Network** → **Image Label**

Predicting Labels from Data

Image **Dense Label**

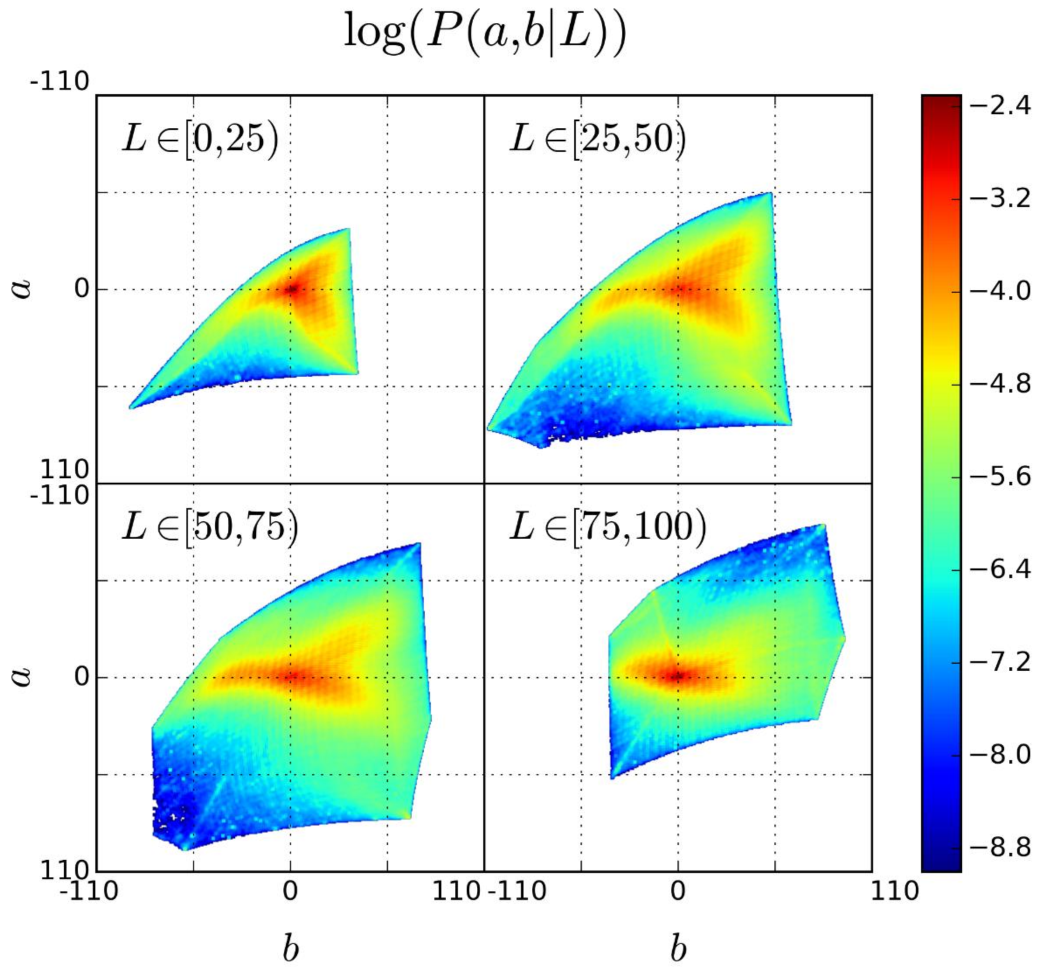


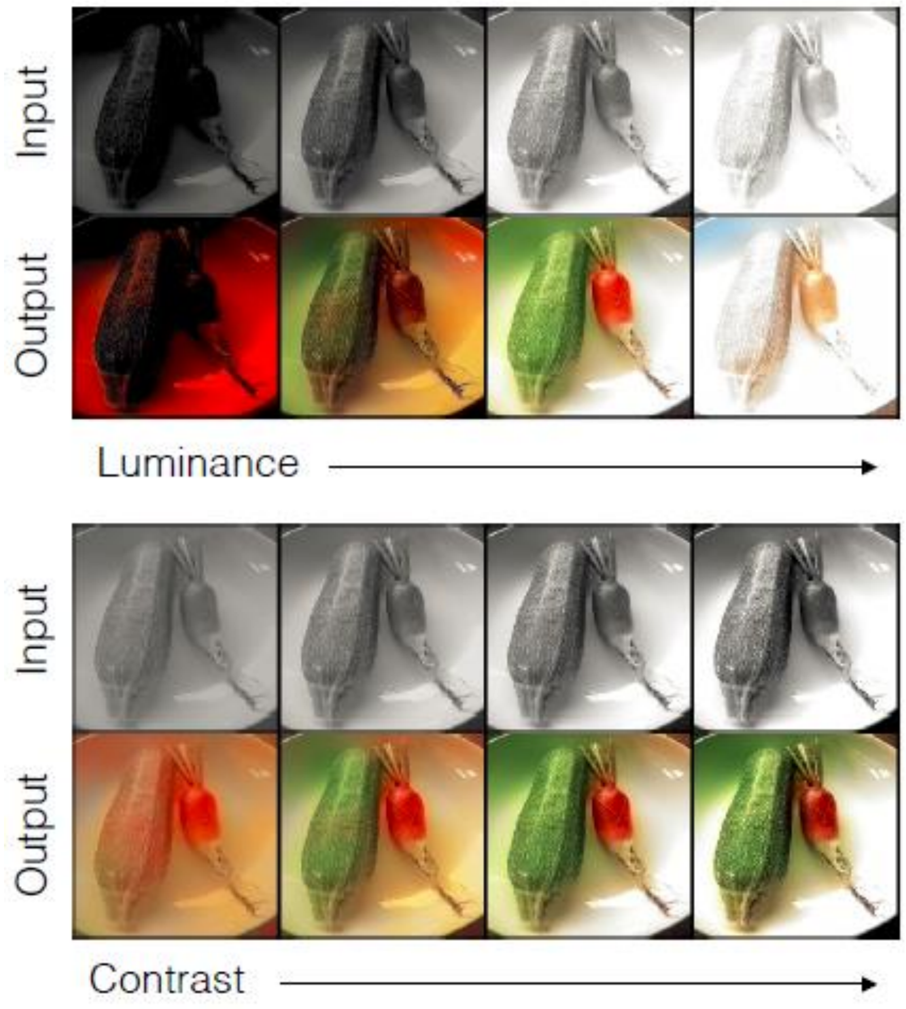
FCNs

Long et al. In *CVPR*, 2015.

Image → Network → **Pixel Labels**

Low-level Perturbations





Common Confusions

jacamar

Ground truth

standard schnauzer

Ground truth

Common Confusions

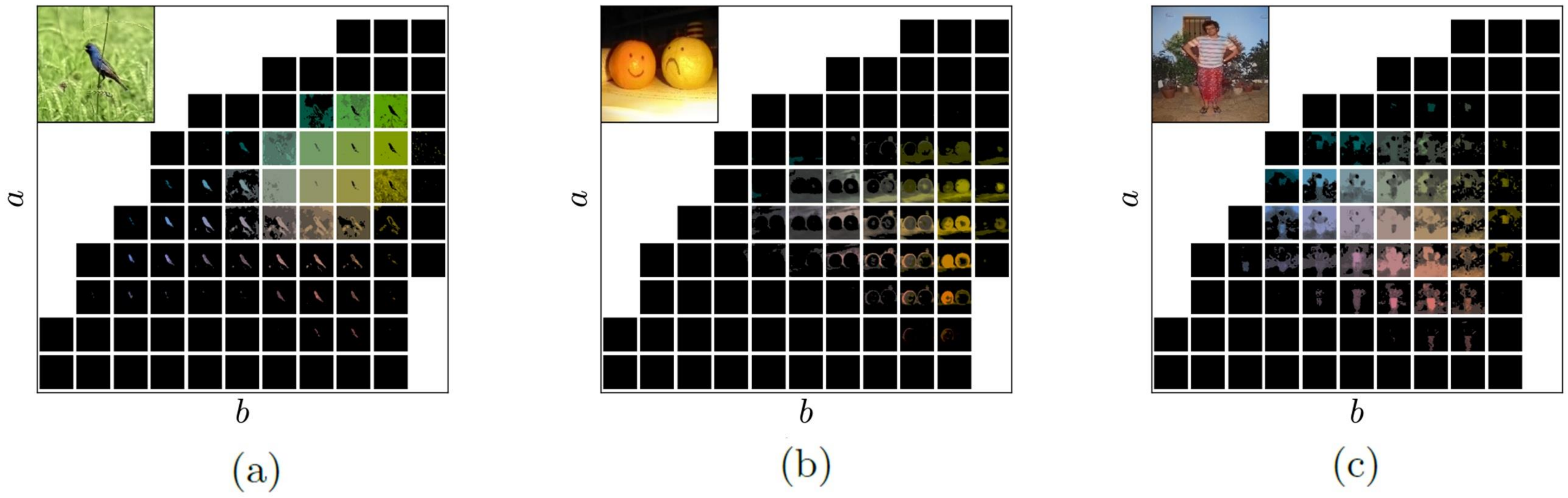
jacamar
bulbul

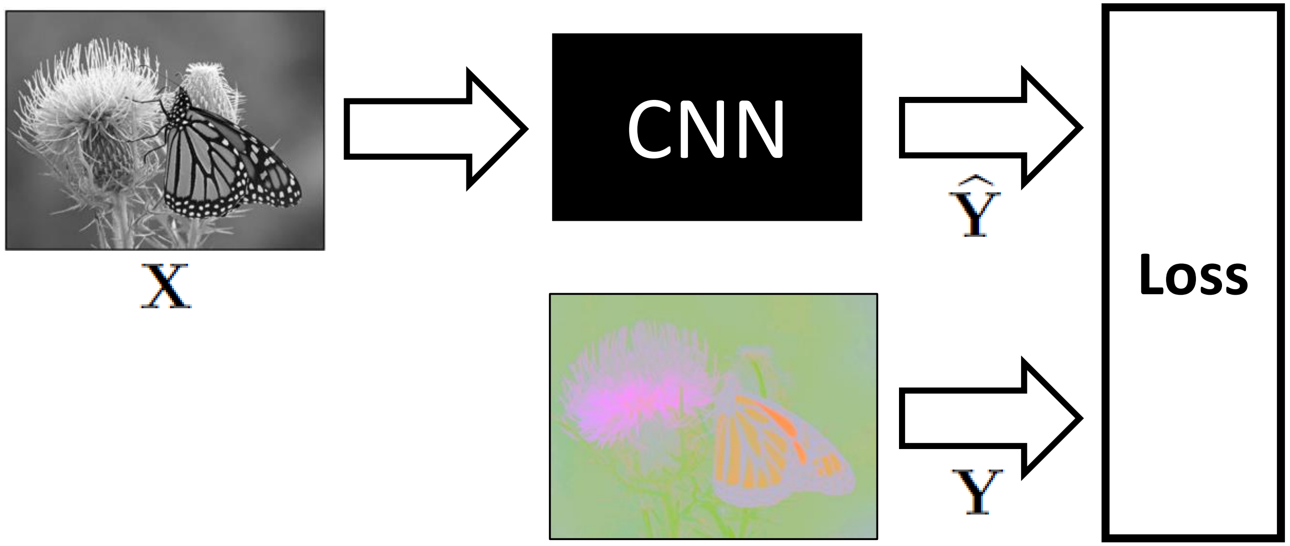
standard schnauzer
irish terrier

Future steps

- Perceptual Losses
- Back-propagate end-to-end
- Train on “infinite” data
- Domain gap to legacy black and white images

Example Output distribution

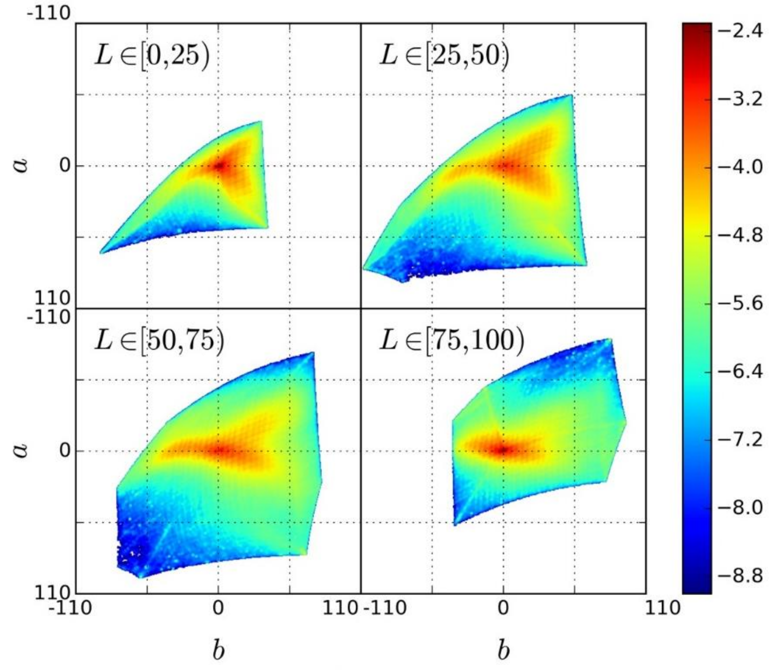




X
CNN
$\hat{Y}$
Loss
Y

Color statistics

Histogram over *ab* space
Conditioned on *L*



Loss

- Regression with L2 loss will not address inherent ambiguity
- Use ***multinomial classification***
  - quantize *ab* space into grid size 10
  - cross entropy loss
- ***Class rebalancing*** at train time to encourage learning of *rare* colors

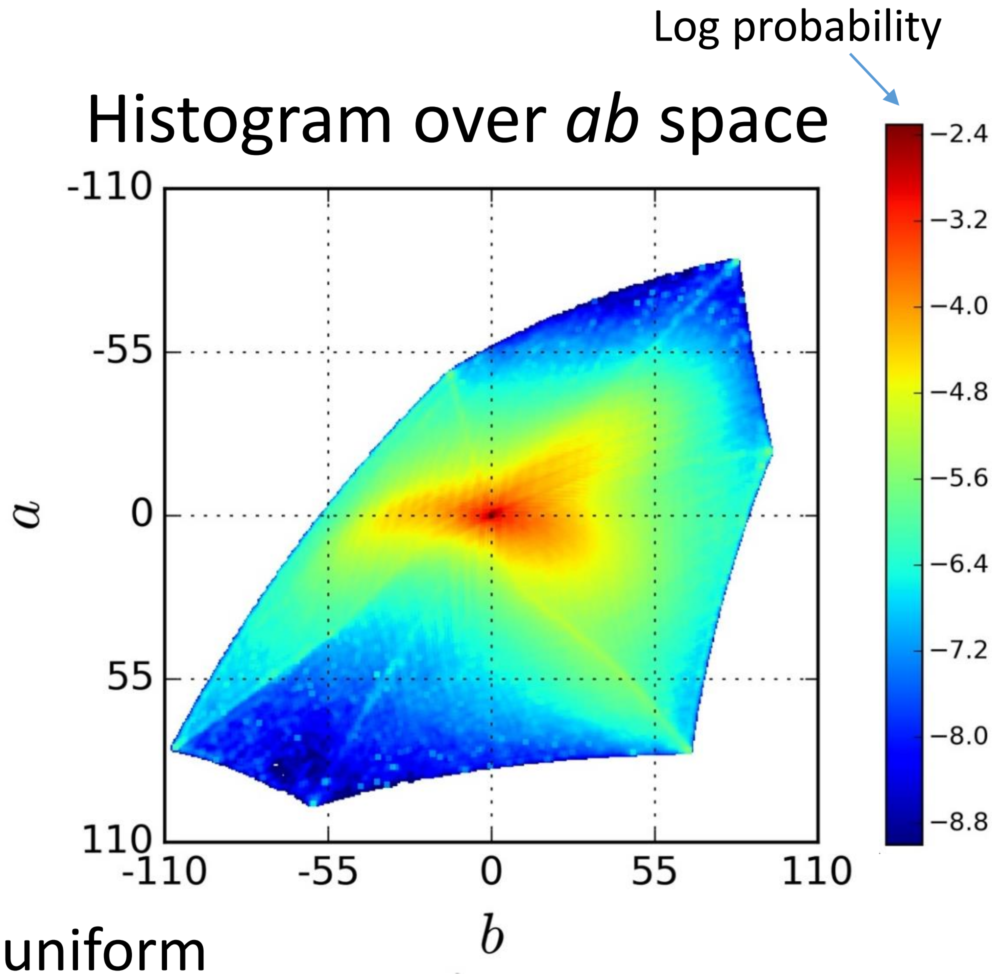
$$\mathbf{w} \propto \left((1-\lambda)(\mathbf{G}_\sigma \circ \mathbf{p}) + \lambda\right)^{-1}$$

Reweighting factor

empirical distribution

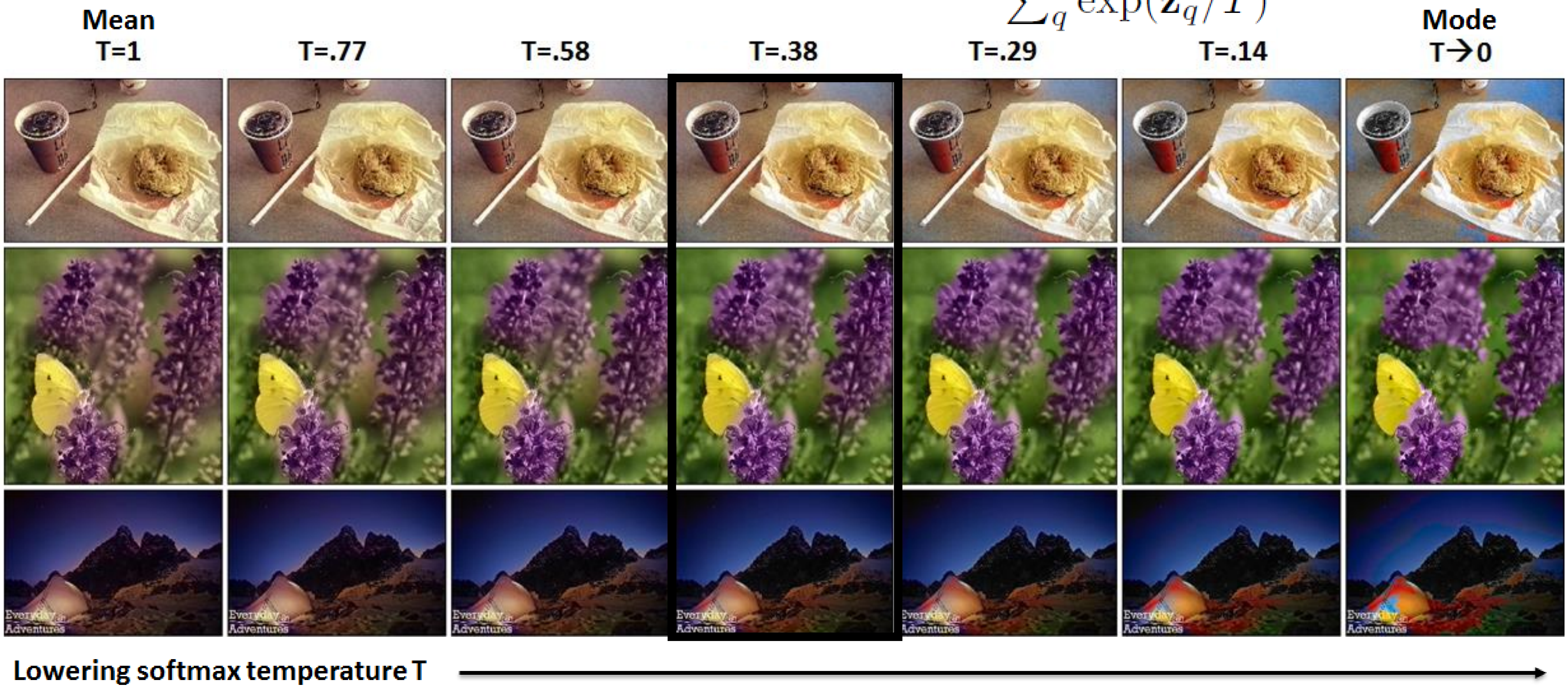
combine with uniform

Histogram over *ab* space

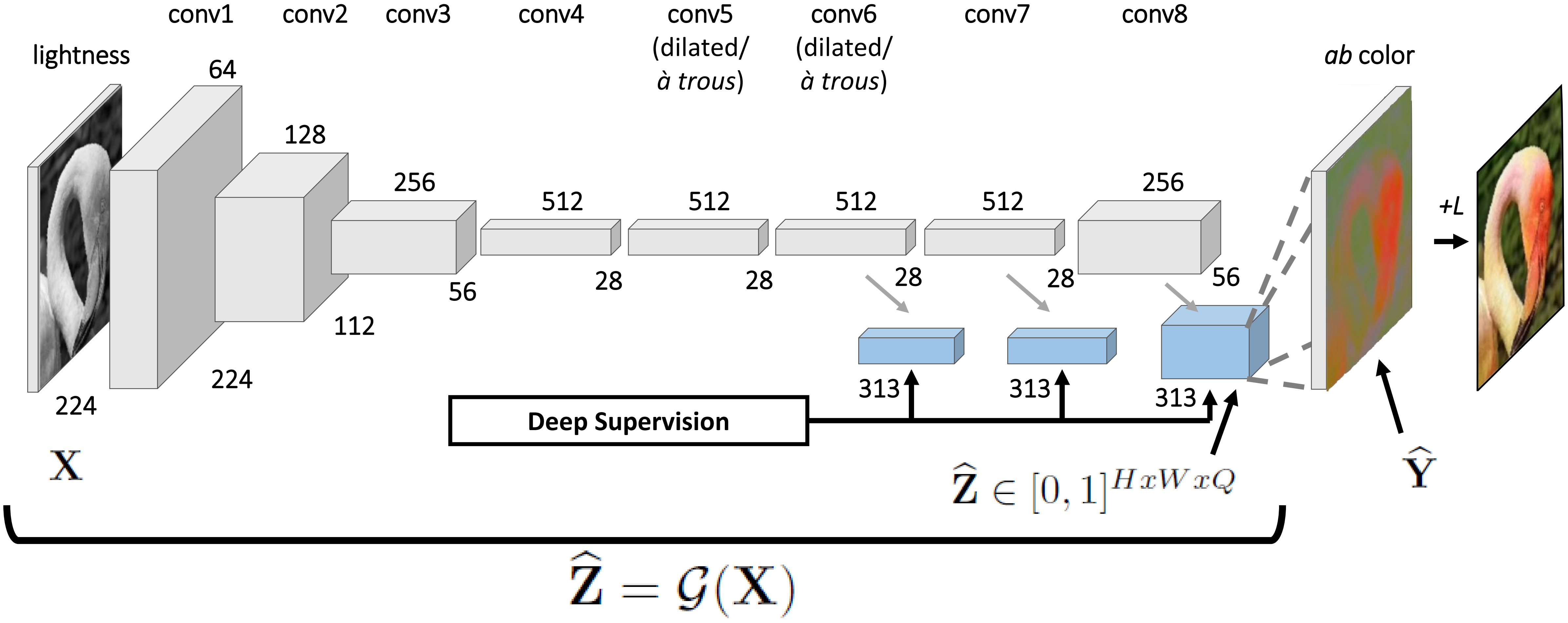


Probability Distribution to Point Estimate

$$\mathcal{H}(\mathbf{Z}_{h,w}) = \mathbb{E}\big(f_T(\log \mathbf{Z}_{h,w})\big), \quad f_T(\mathbf{z}) = \frac{\exp(\mathbf{z}/T)}{\sum_q \exp(\mathbf{z}_q/T)}$$



Network Architecture



Network Architecture