Generative Networks

James Hays

Computer Vision

Interesting Illusion: Ames Window

- https://www.youtube.com/watch?v=aHjQe8EuKHc
- https://en.wikipedia.org/wiki/Ames_trapezoid

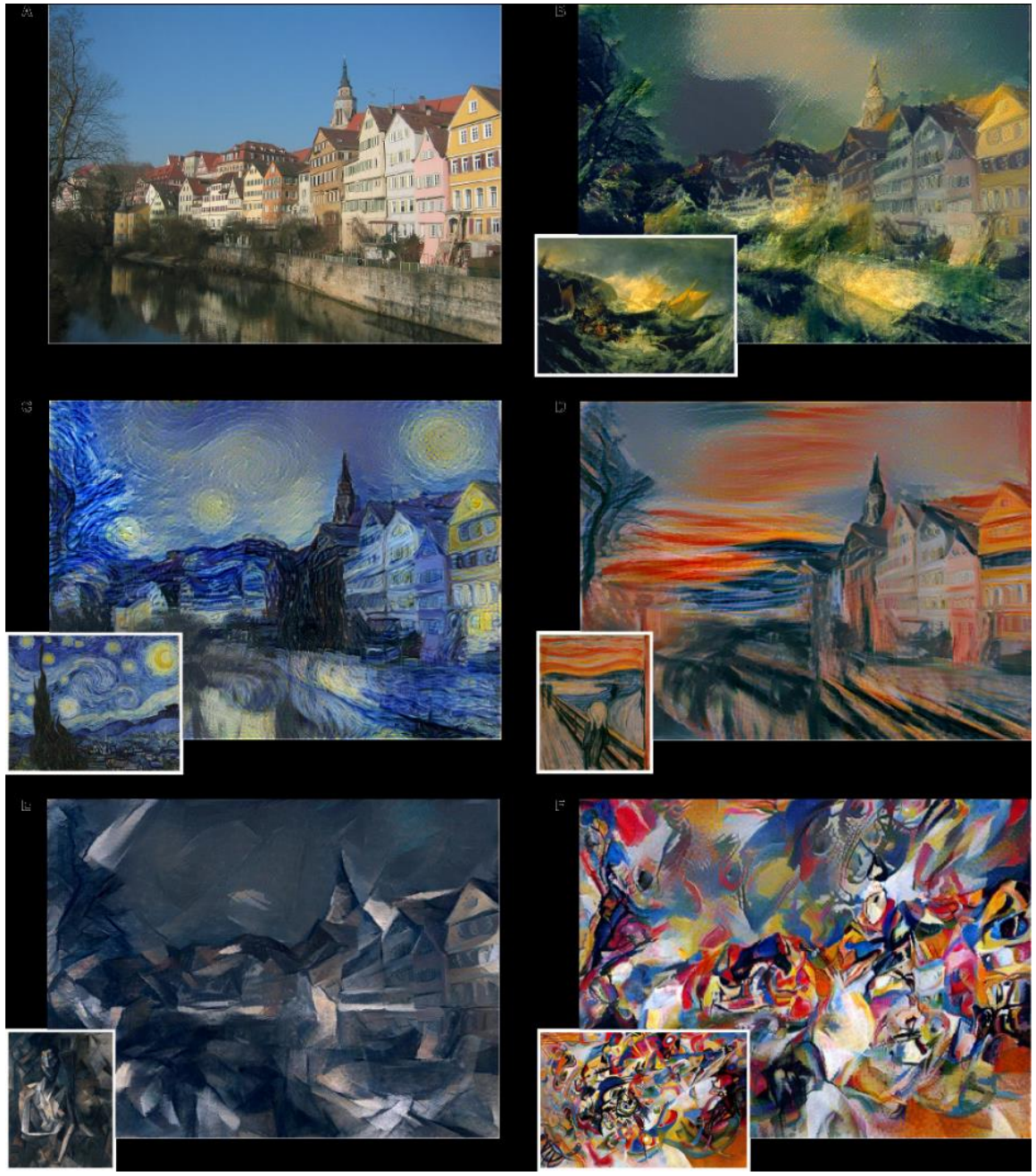
Recap

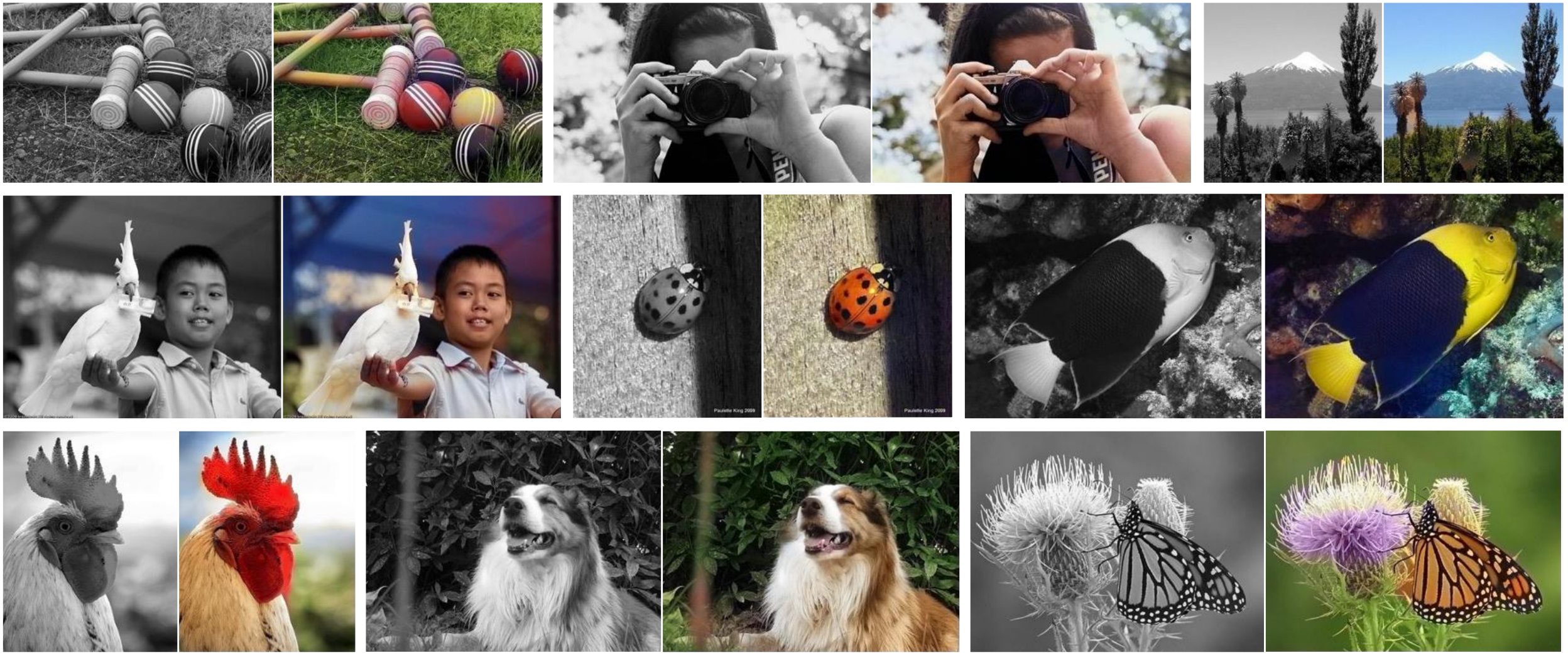
- “Unsupervised Learning”
- Style Transfer

A Neural Algorithm of Artistic Style

Leon A. Gatys, Alexander S. Ecker, Matthias Bethge.

CVPR 2016.

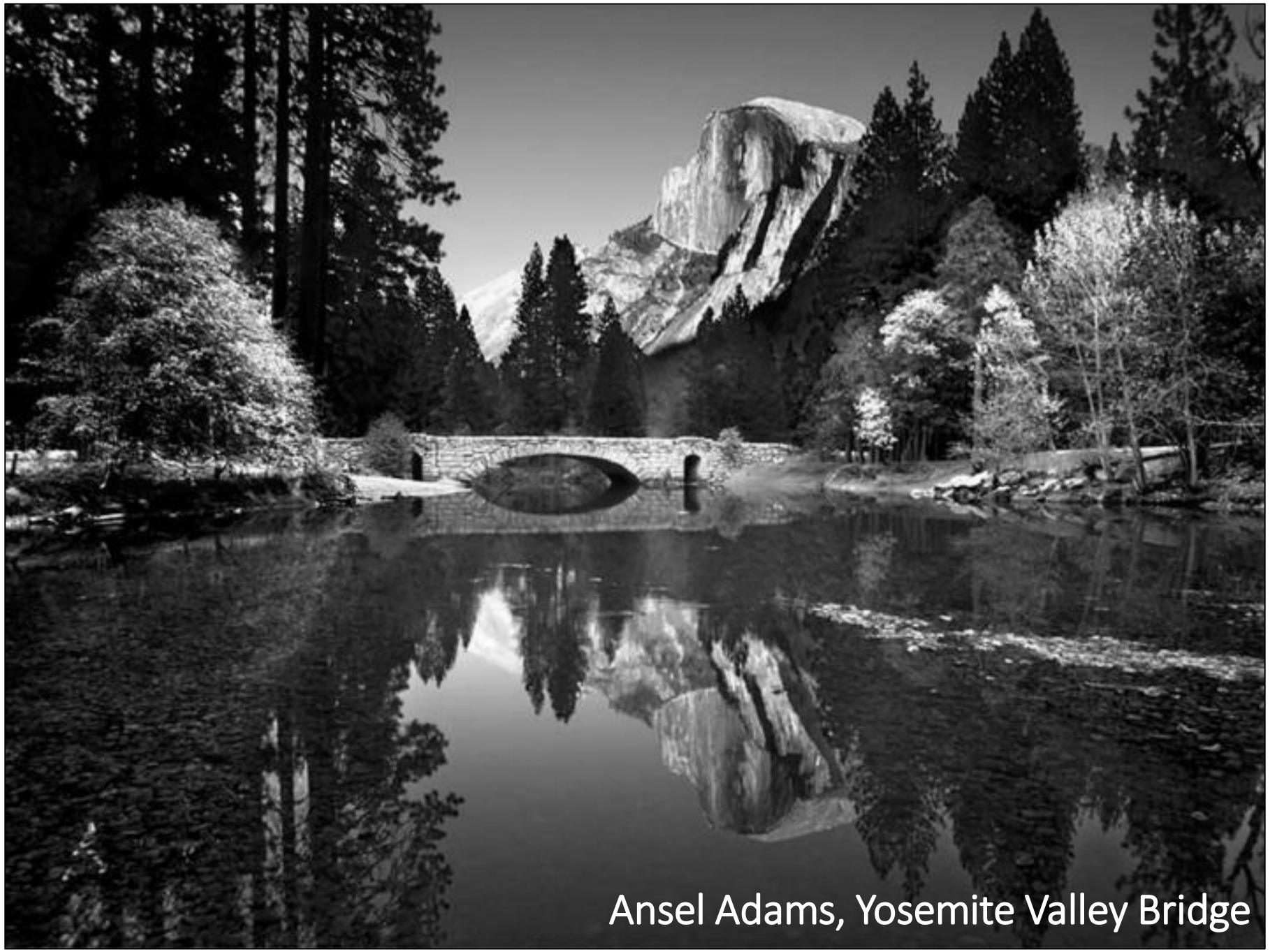


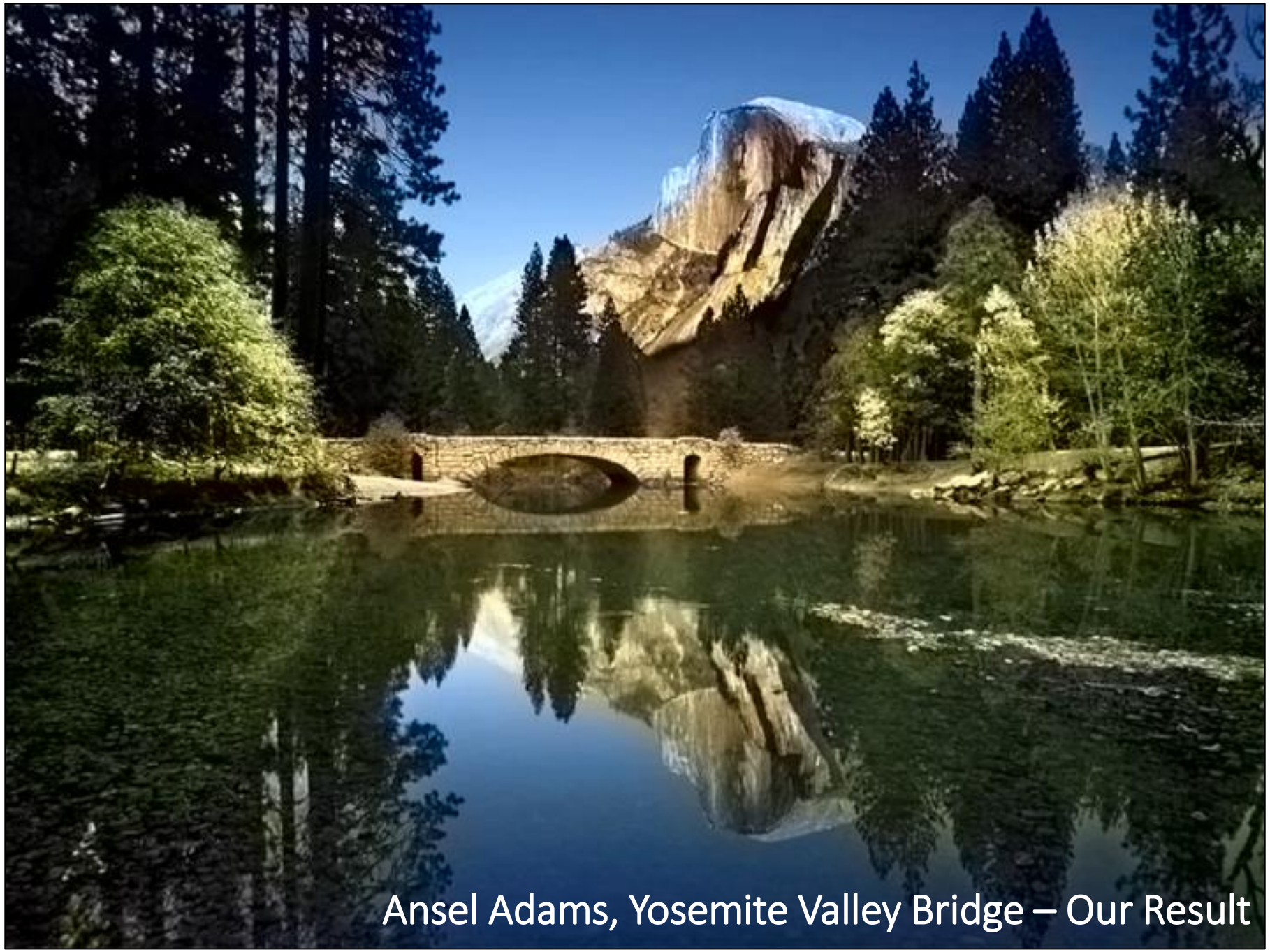


Colorful Image Colorization

Richard Zhang, Phillip Isola, Alexei (Alyosha) Efros

richzhang.github.io/colorization


Ansel Adams, Yosemite Valley Bridge


Ansel Adams, Yosemite Valley Bridge – Our Result

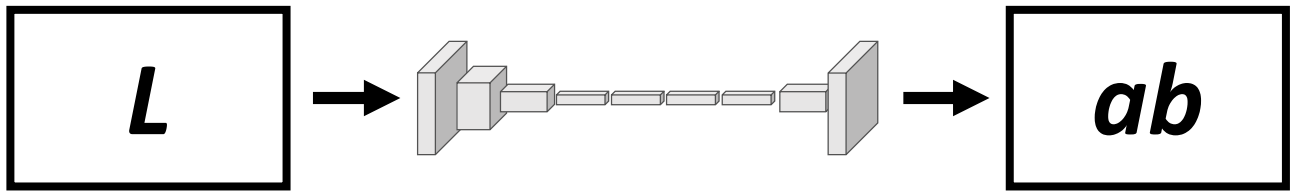
$\mathcal{F}$

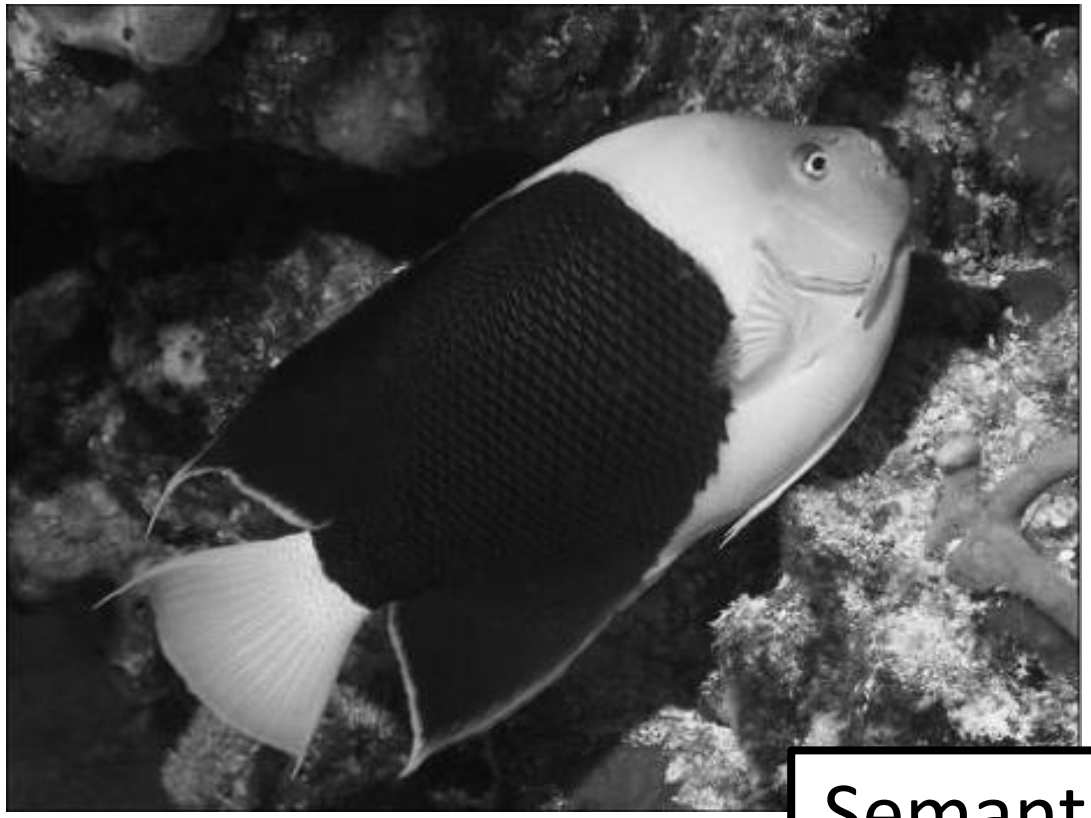
Grayscale image: *L* channel

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

Color information: *ab* channels

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



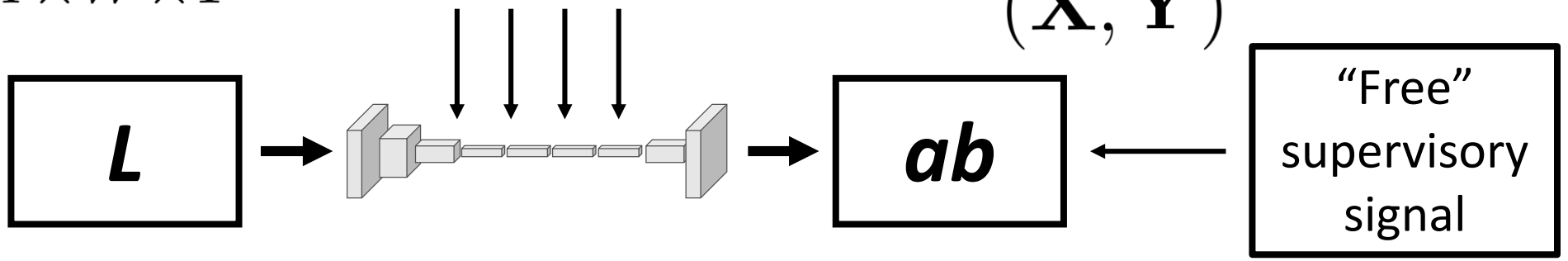


$\mathcal{F}$

Semantics? Higher-level abstraction?

Grayscale image: *L* ch ncatenate (L,ab)

$\mathbf{X} \in \mathbb{R}^{H \times W \times 1}$ $(\mathbf{X}, \hat{\mathbf{Y}})$



Inherent Ambiguity

Grayscale

Inherent Ambiguity

Our Output

Ground Truth

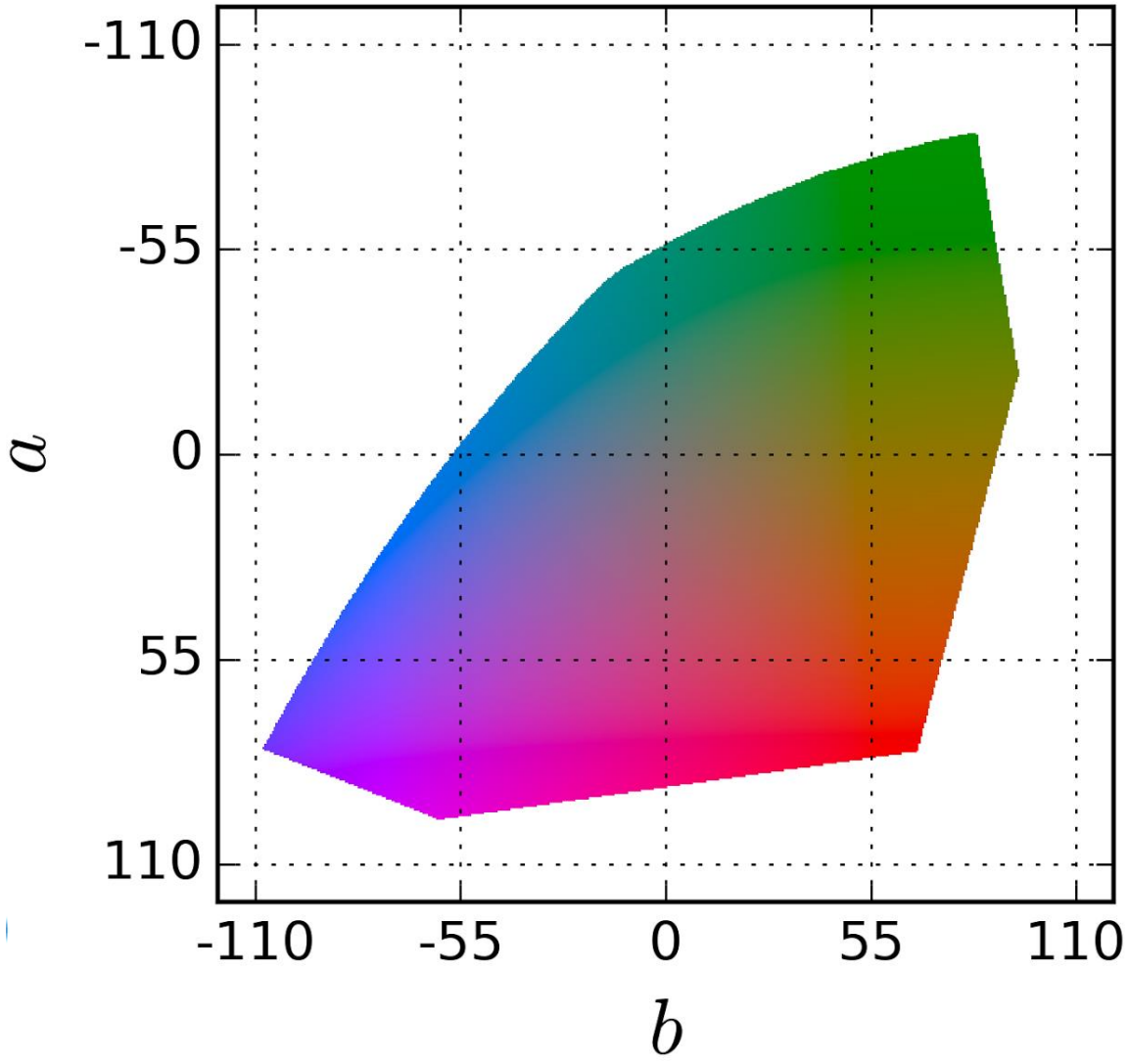
Better Loss Function

- 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

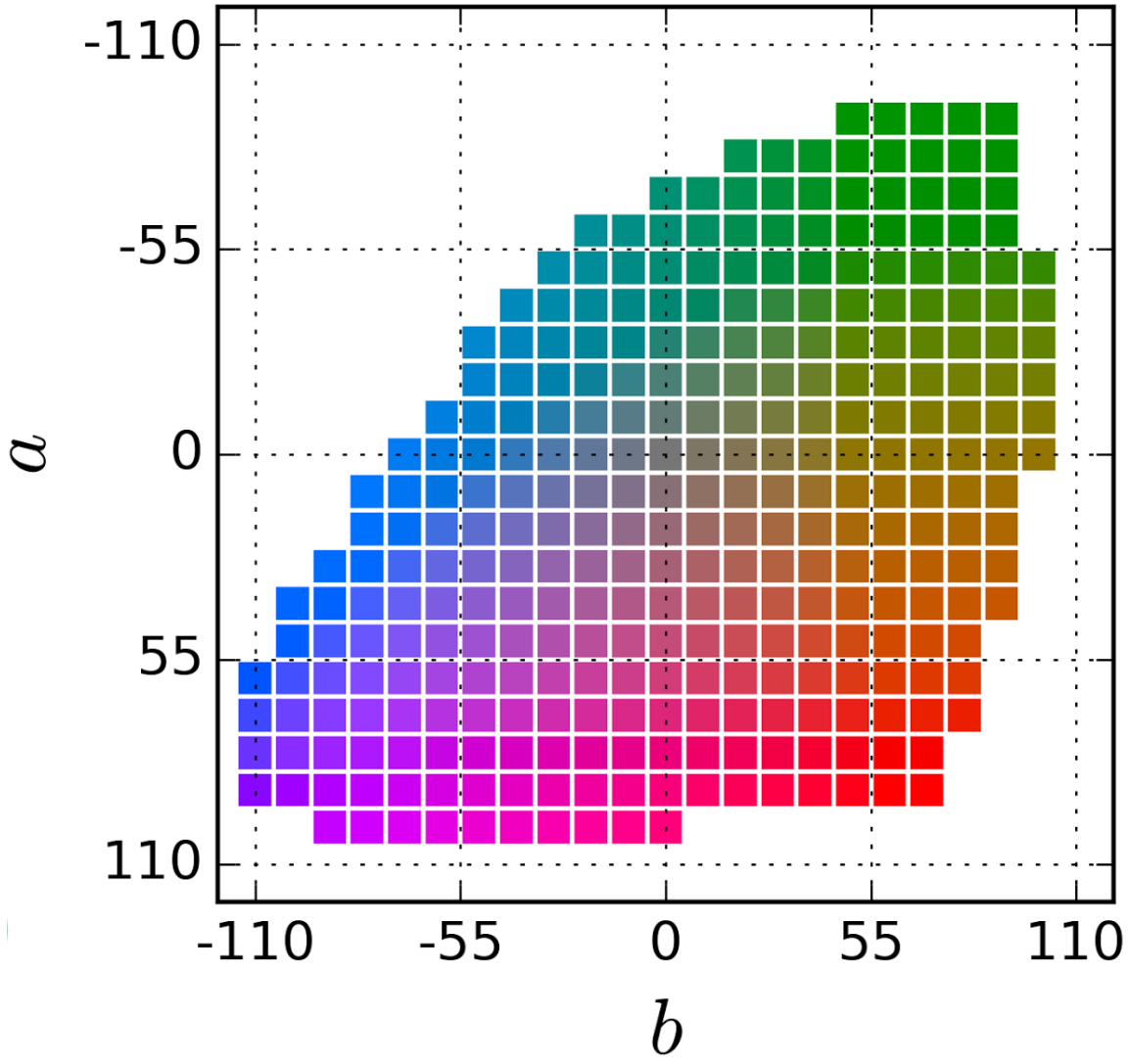
$$\mathrm{L}_2(\widehat{\mathbf{Y}}, \mathbf{Y}) = \frac{1}{2}\sum_{h,w} \|\mathbf{Y}_{h,w} - \widehat{\mathbf{Y}}_{h,w}\|_2^2$$

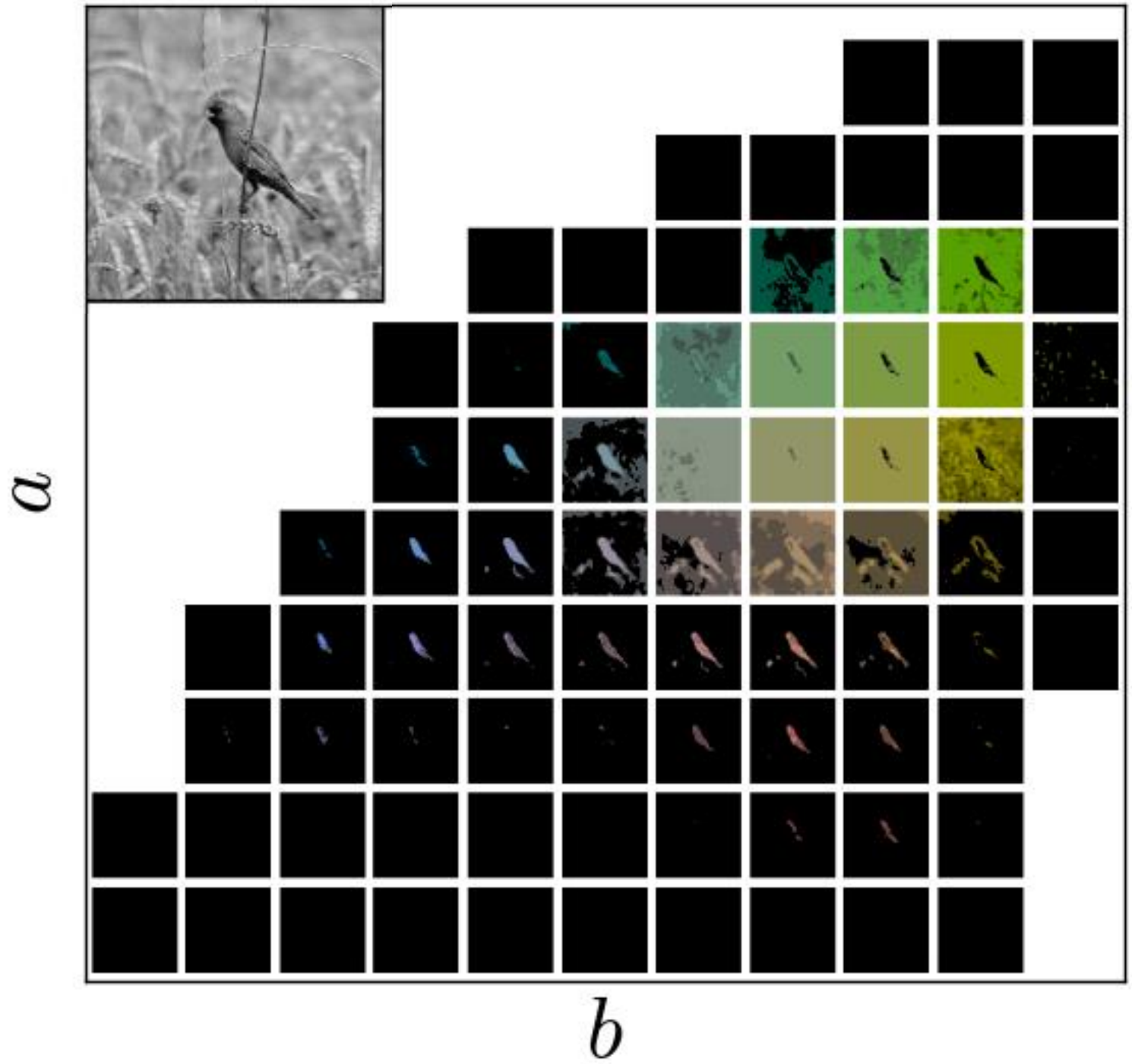
- Use **multinomial classification**

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

Colors in *ab* space

(discrete)




a
b

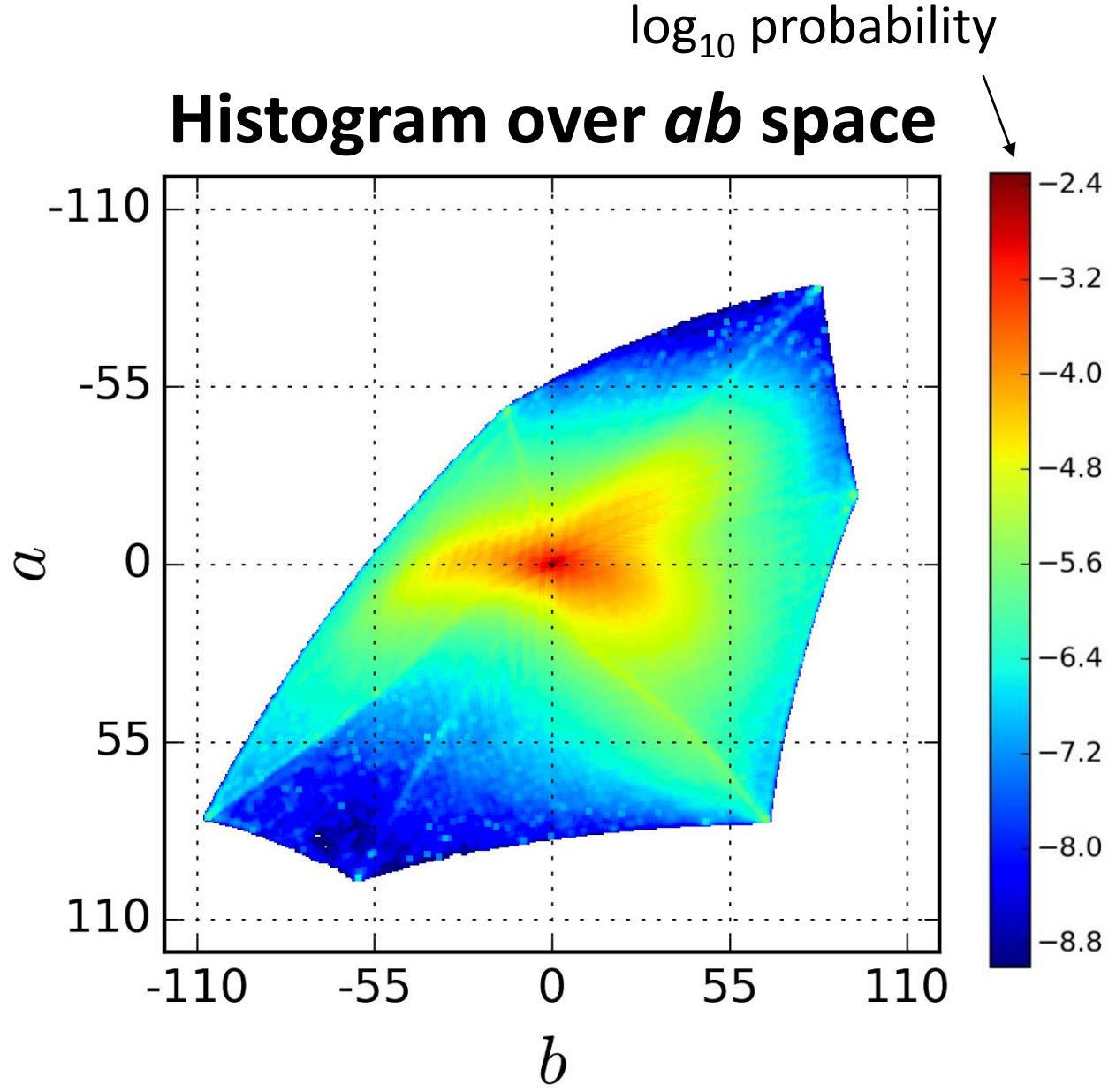
Better Loss Function

- 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

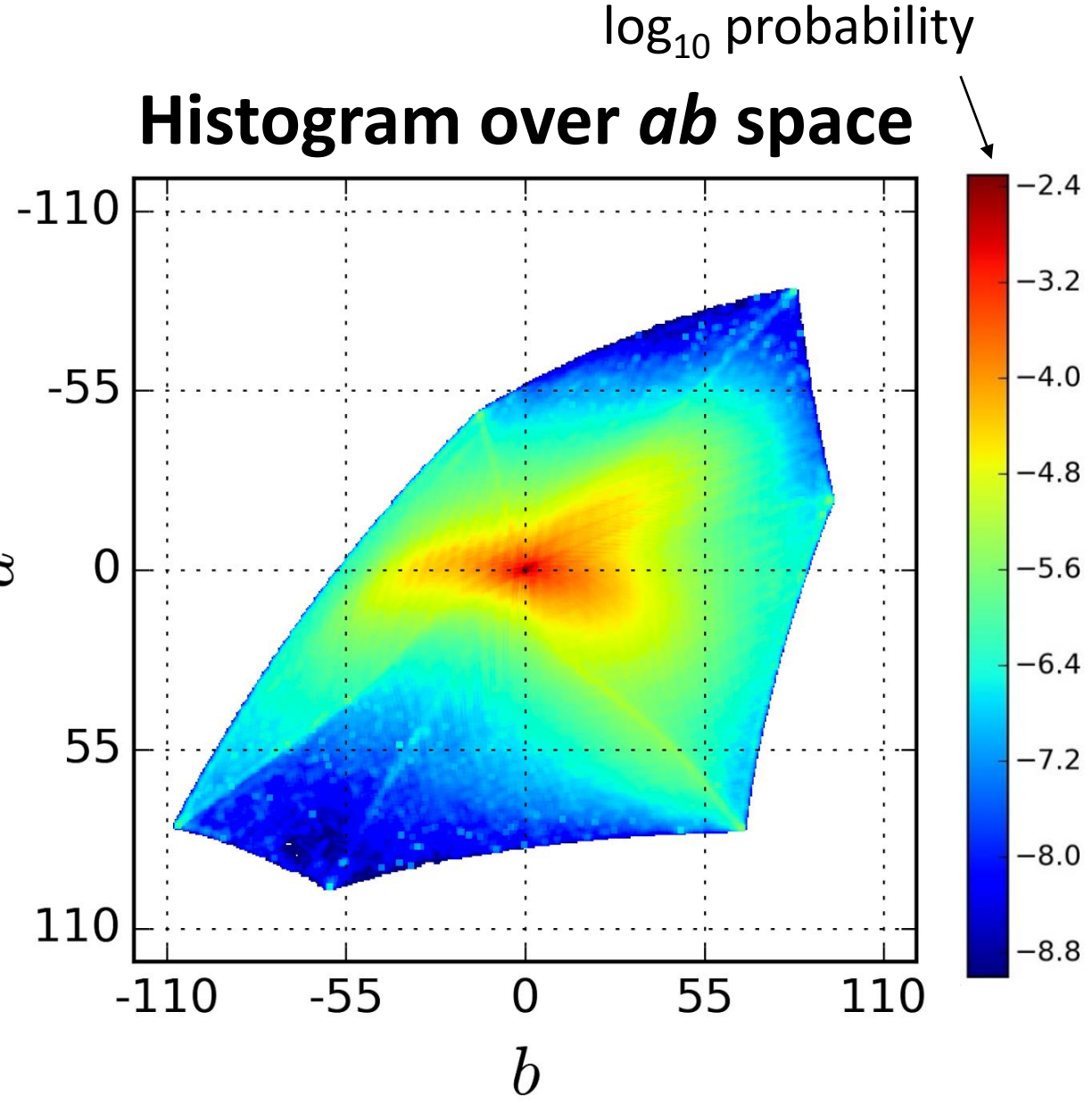
$$\mathrm{L}_2(\widehat{\mathbf{Y}}, \mathbf{Y}) = \frac{1}{2}\sum_{h,w} \|\mathbf{Y}_{h,w} - \widehat{\mathbf{Y}}_{h,w}\|_2^2$$

- Use **multinomial classification**

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

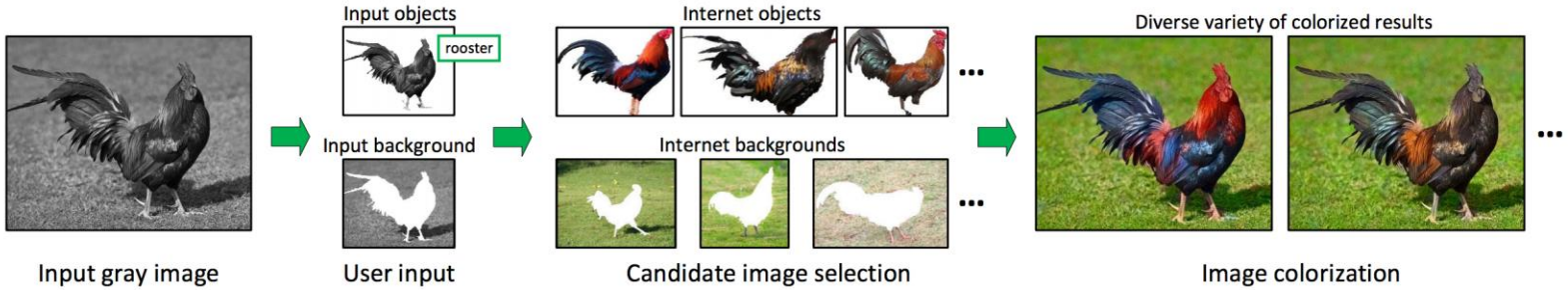
- **Class rebalancing** to encourage learning of *rare* colors

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



Non-parametric

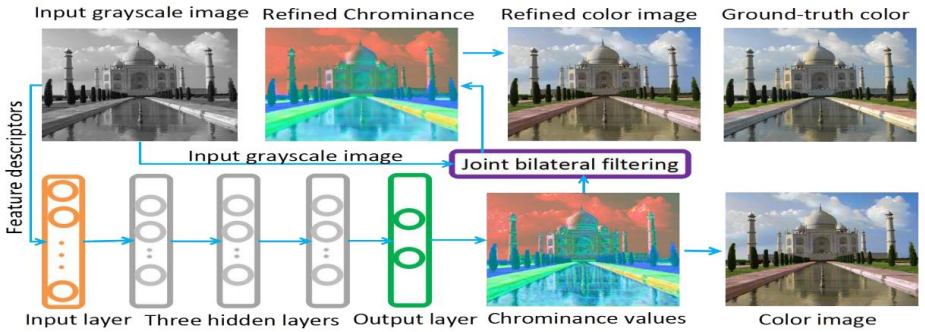
Hertzmann et al. In SIGGRAPH, 2001.
Welsh et al. In TOG, 2002.
Irony et al. In Eurographics, 2005.
Liu et al. In TOG, 2008.
Chia et al. In ACM 2011.
Gupta et al. In ACM, 2012.



Parametric

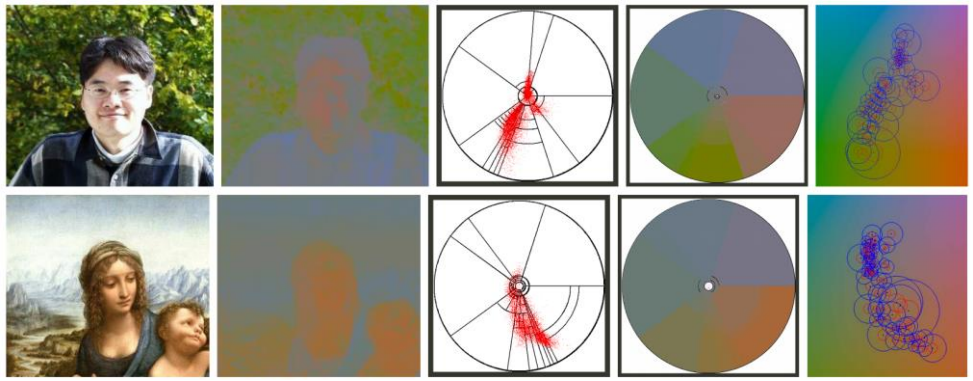
Hand-engineered Features

L2 Regression



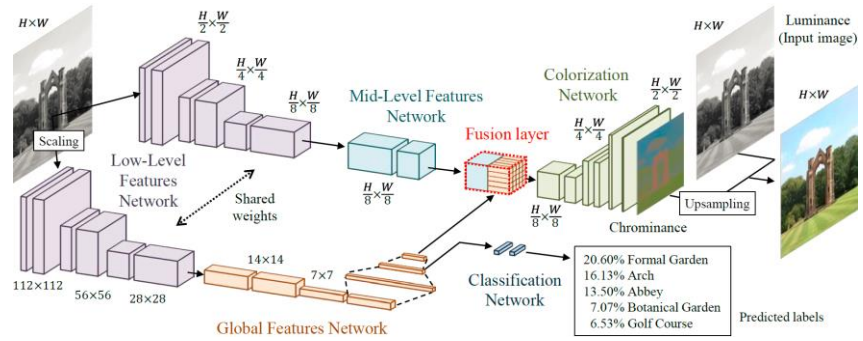
Deshpande et al. Cheng et al. In ICCV 2015.

Classification

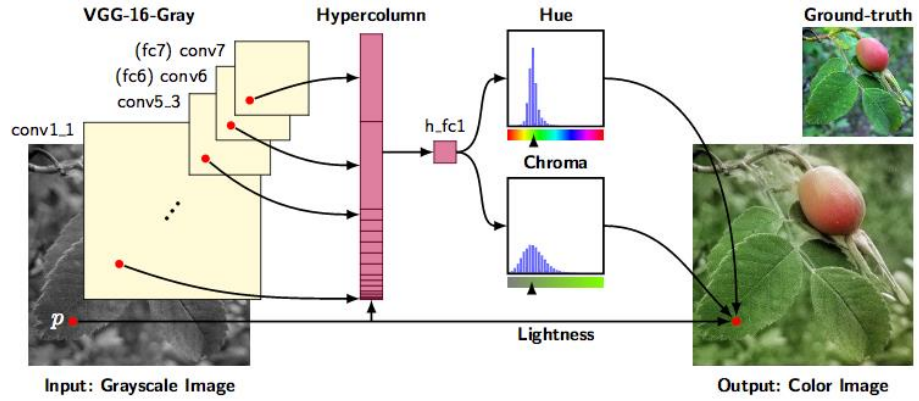


Charpiat et al. In ECCV 2008.

Deep Networks

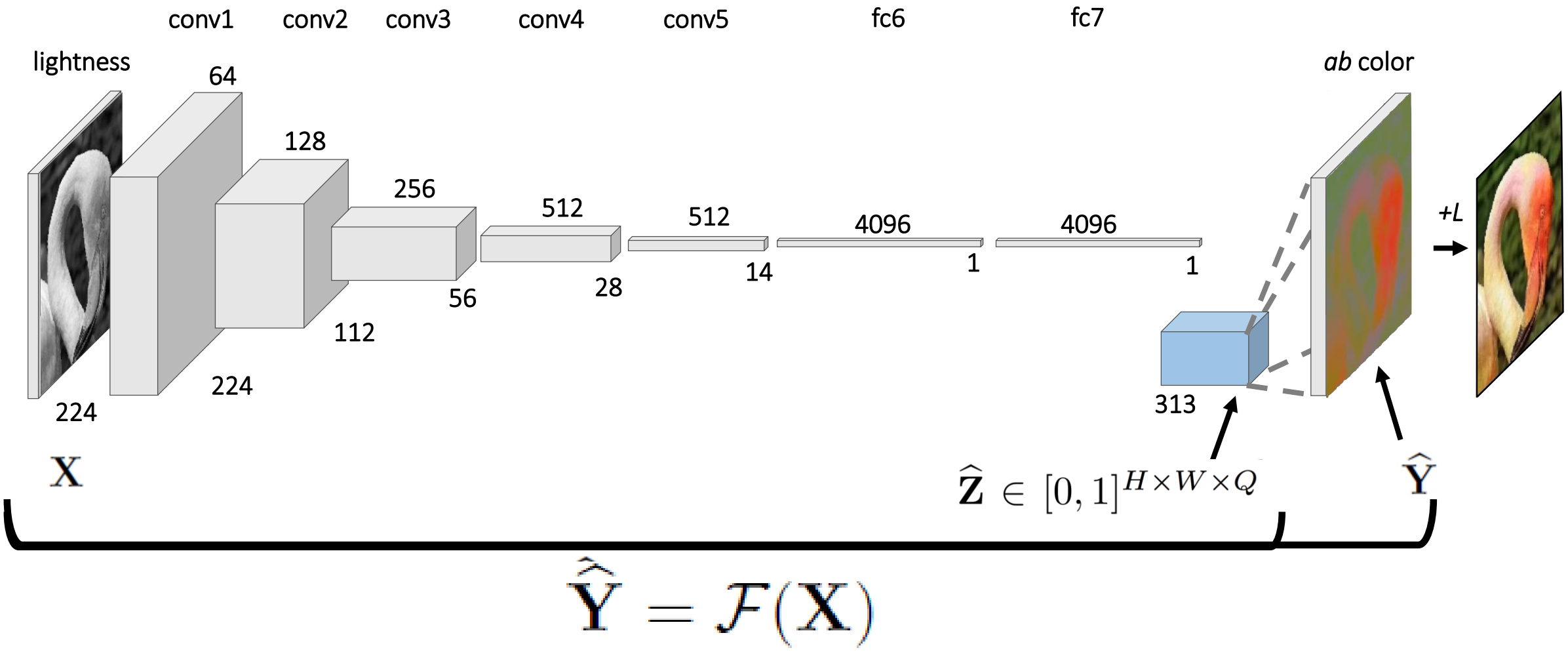


Dahl. Jan 2016. Iizuka et al. In SIGGRAPH, 2016.



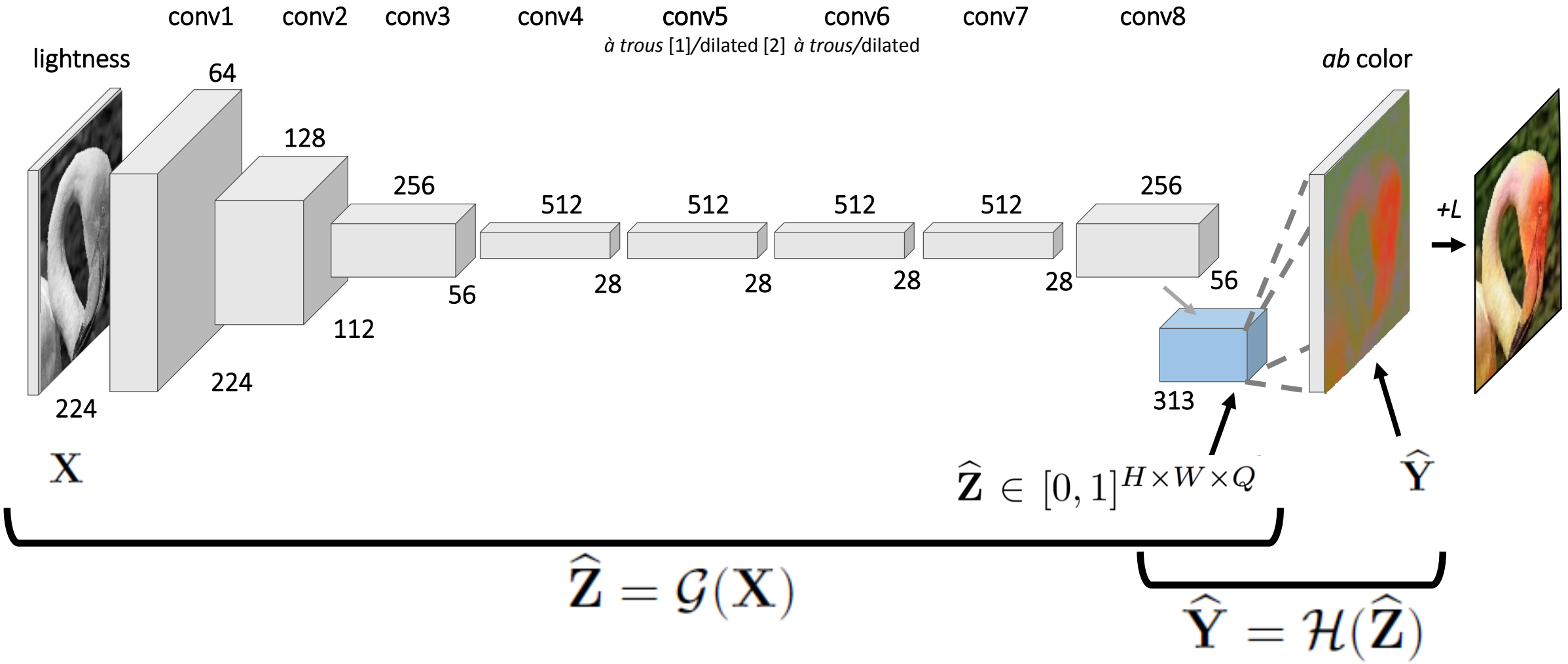
Larsson et al. In ECCV 2016. [Concurrent]

Network Architecture



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

Network Architecture



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

Grayscale input

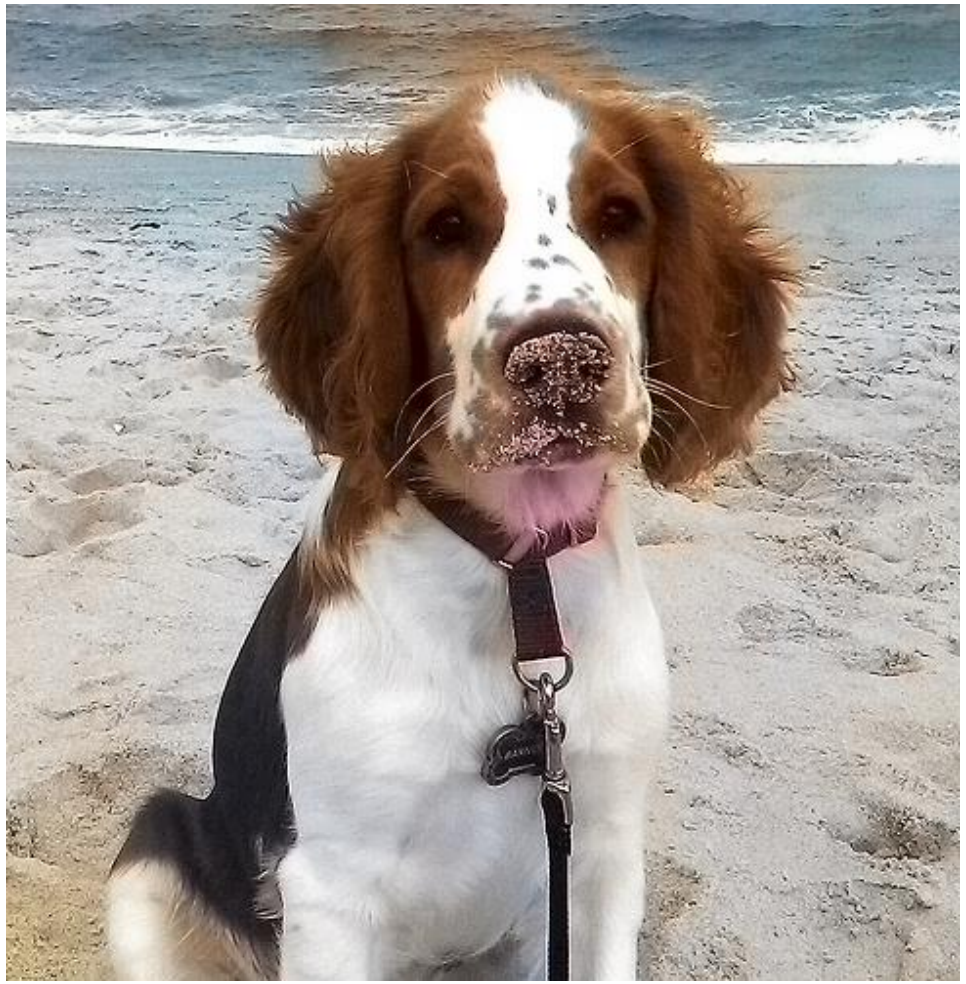
Ground truth

L2 Regression

Class w/ Rebalancing

Failure Cases

Biases



Evaluation

	Visual Quality
Quantitative	Per-pixel accuracy

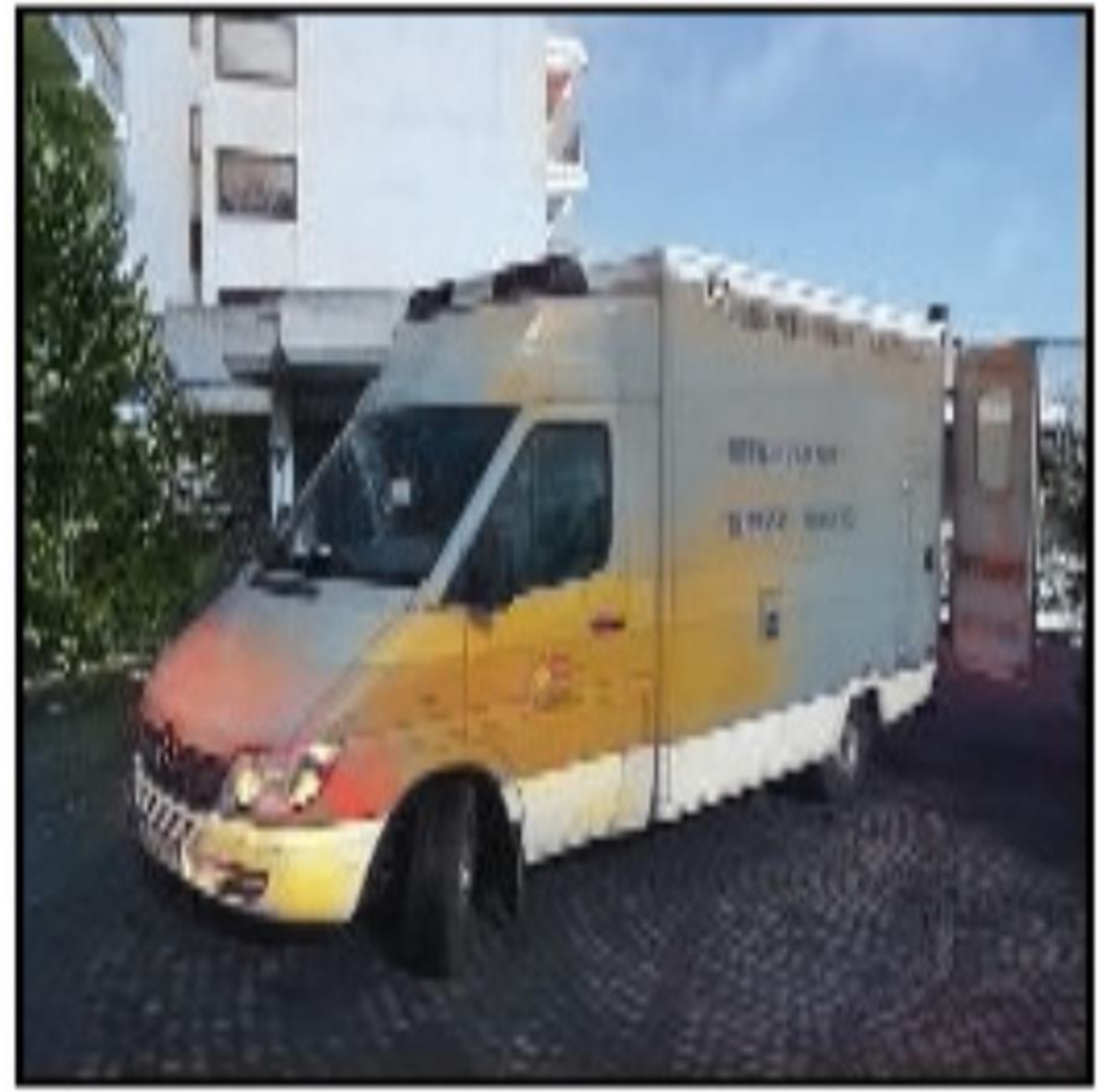
Evaluation

	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

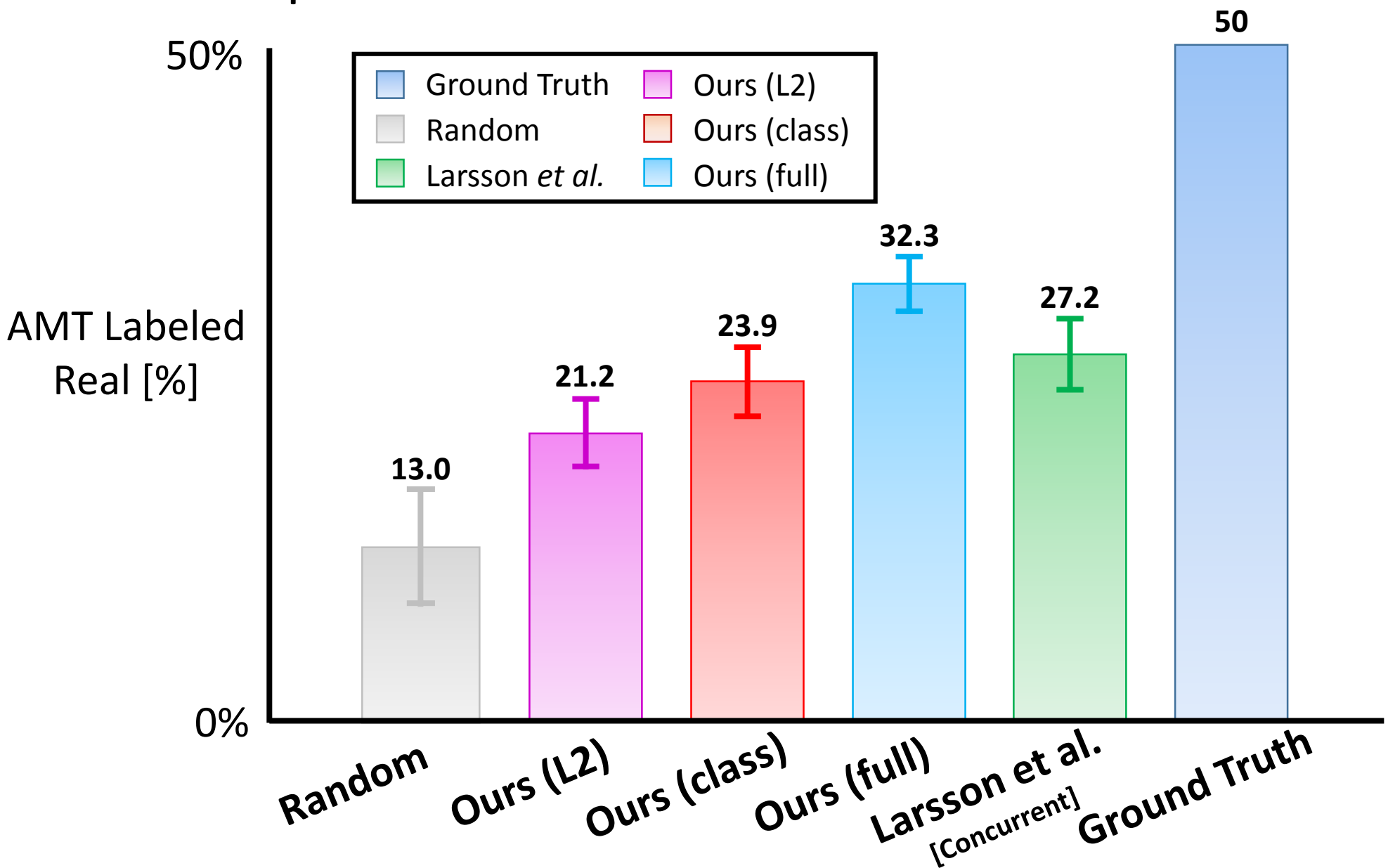
from Reddit /u/SherySantucci

Recolorized by Reddit ColorizeBot

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

Recolorized by
Reddit
ColorizeBot

Perceptual Realism Test

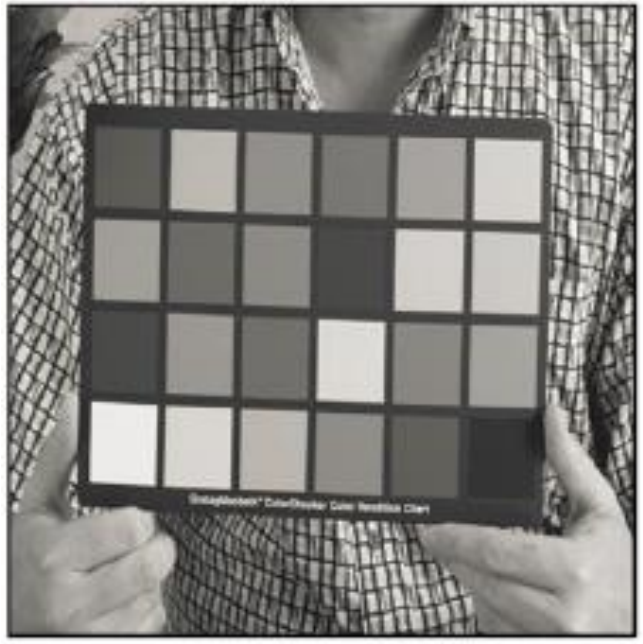


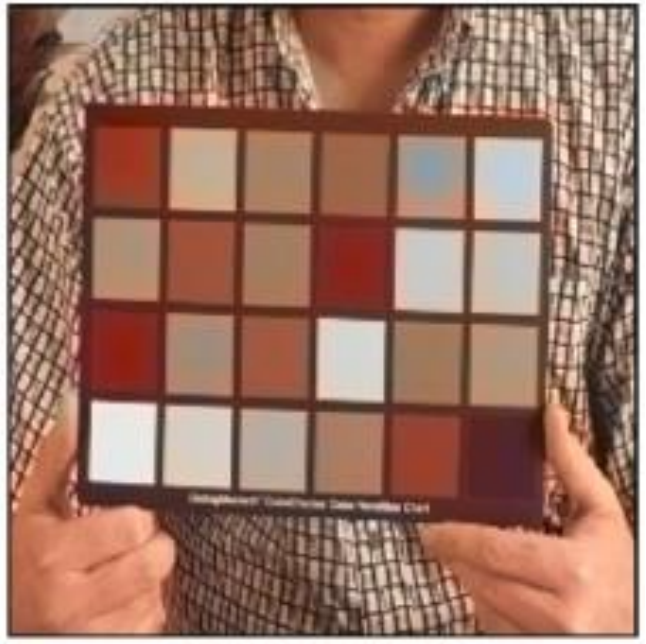
1600 images tested per algorithm

Input

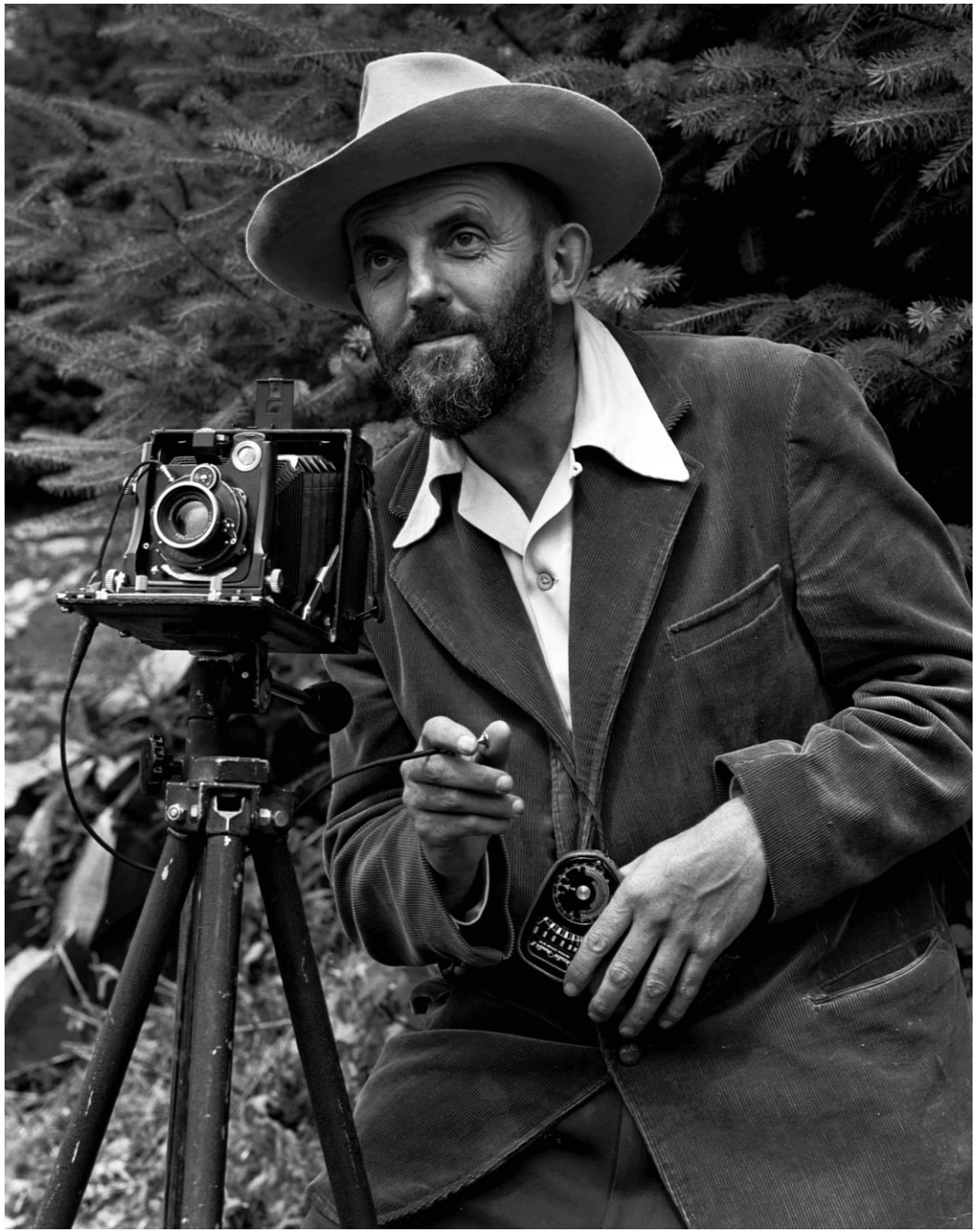
Ground Truth

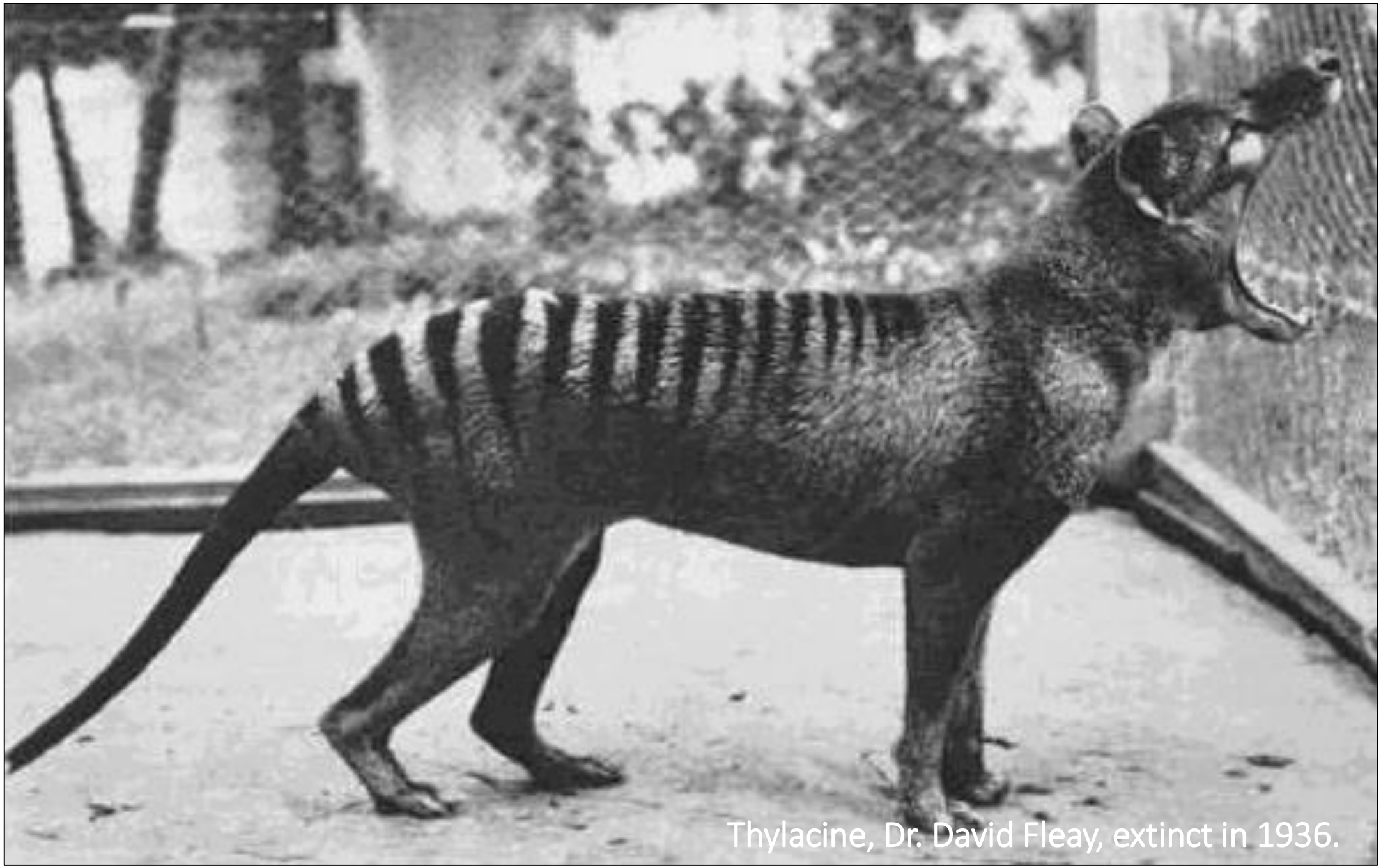
Output





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

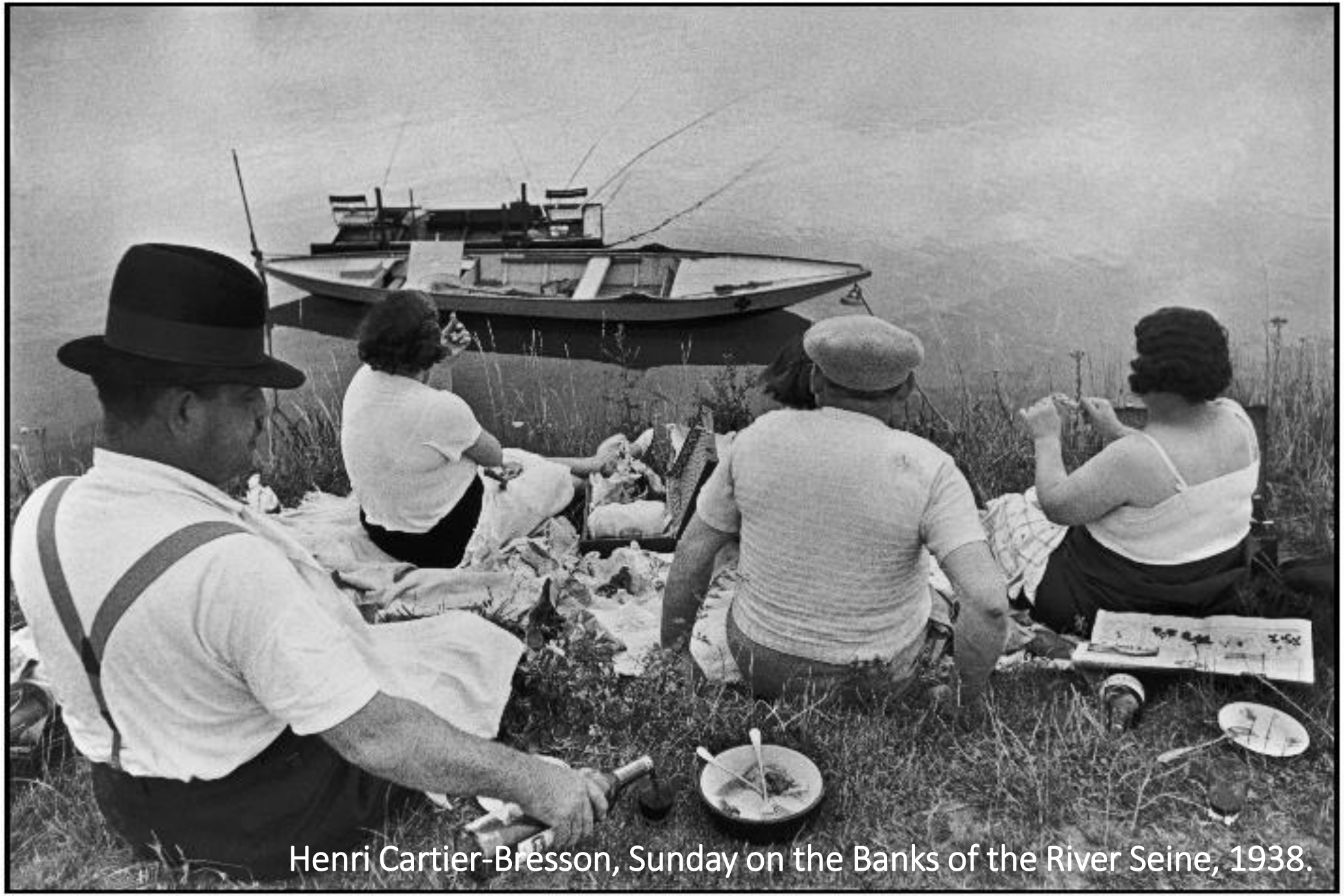



Thylacine, Dr. David Fleay, extinct in 1936.


Thylacine, Dr. David Fleay, extinct in 1936.

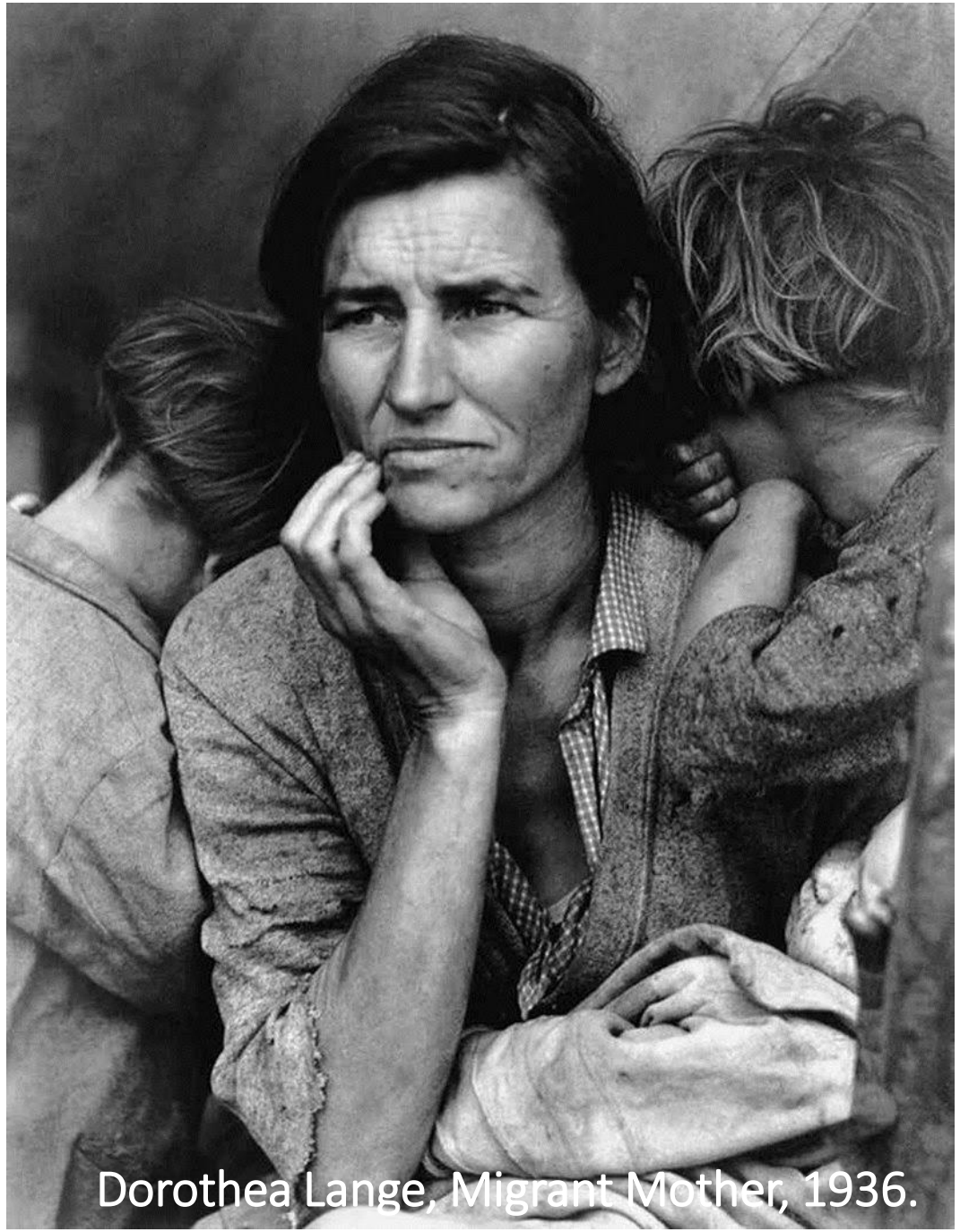
Amateur Family Photo, 1956.

Amateur Family Photo, 1956.

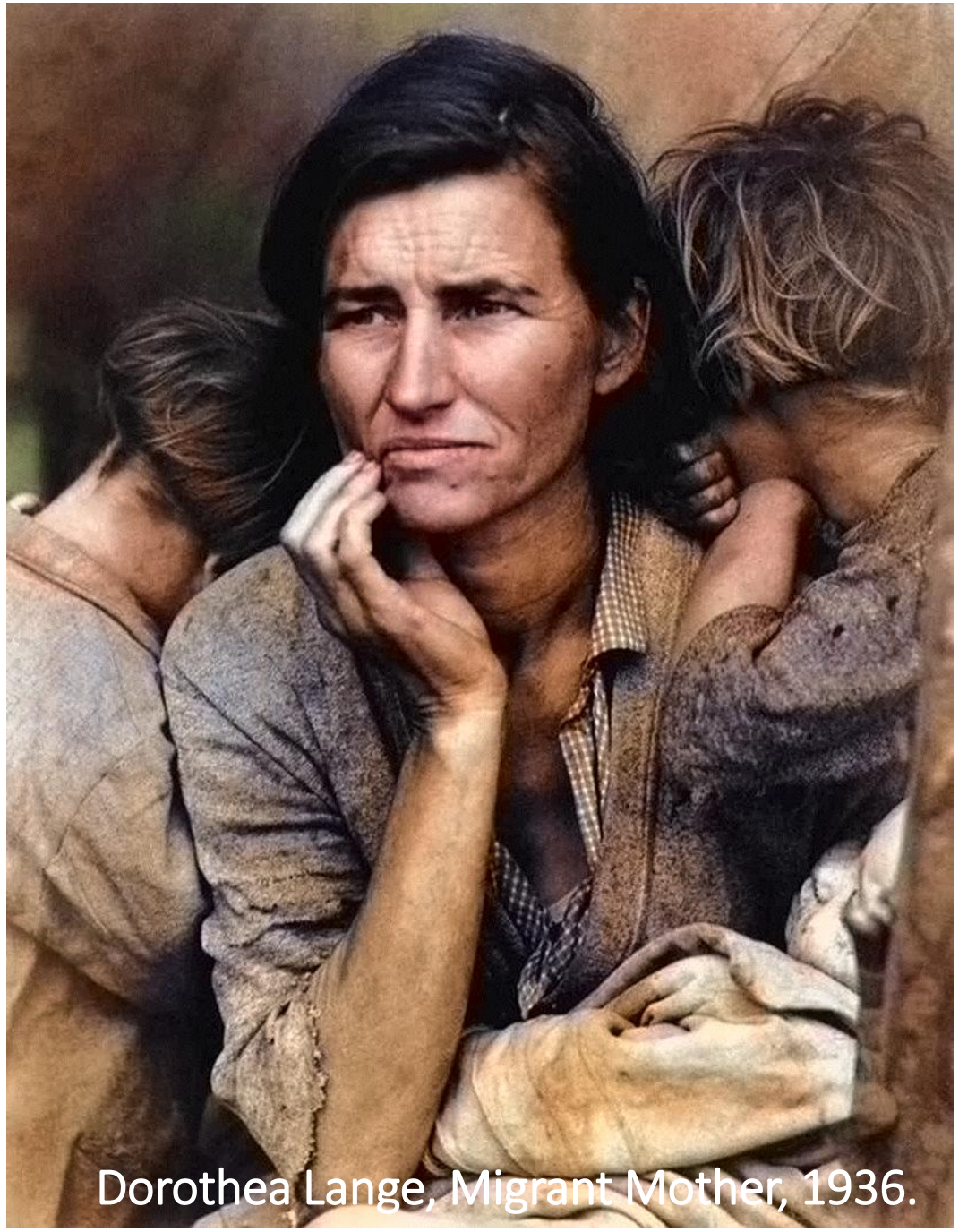


Henri Cartier-Bresson, Sunday on the Banks of the River Seine, 1938.

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



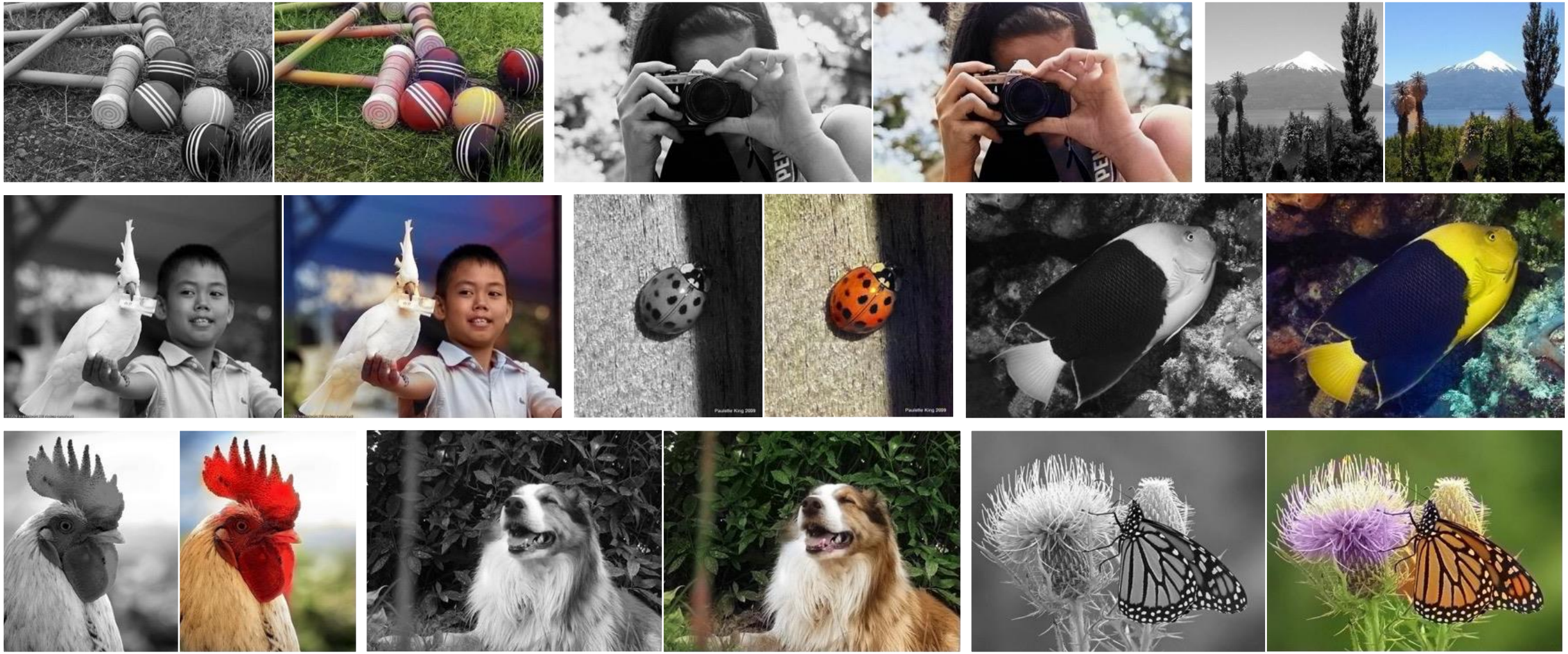
Dorothea Lange, Migrant Mother, 1936.



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



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

Small Sample of other Generative Networks of interest

- **DCGAN. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks**
  Alec Radford, Luke Metz, Soumith Chintala
  https://arxiv.org/abs/1511.06434
- **Generative Adversarial Text-to-Image Synthesis**
  [PDF][Supplement][BibTex][Code]
  Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, Honglak Lee. *ICML* 2016.
- **Pix2Pix: Image-to-Image Translation with Conditional Adversarial Nets.**
  Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros
  https://phillipi.github.io/pix2pix/