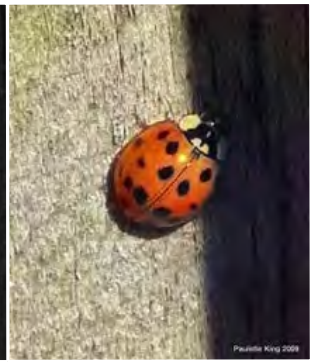
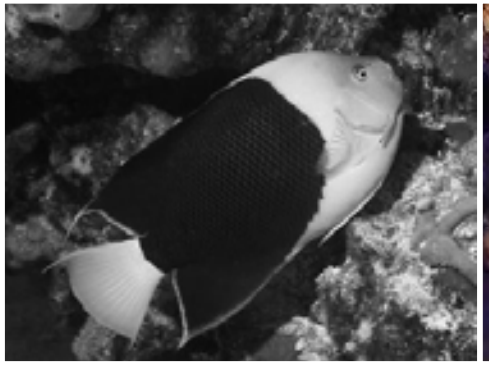
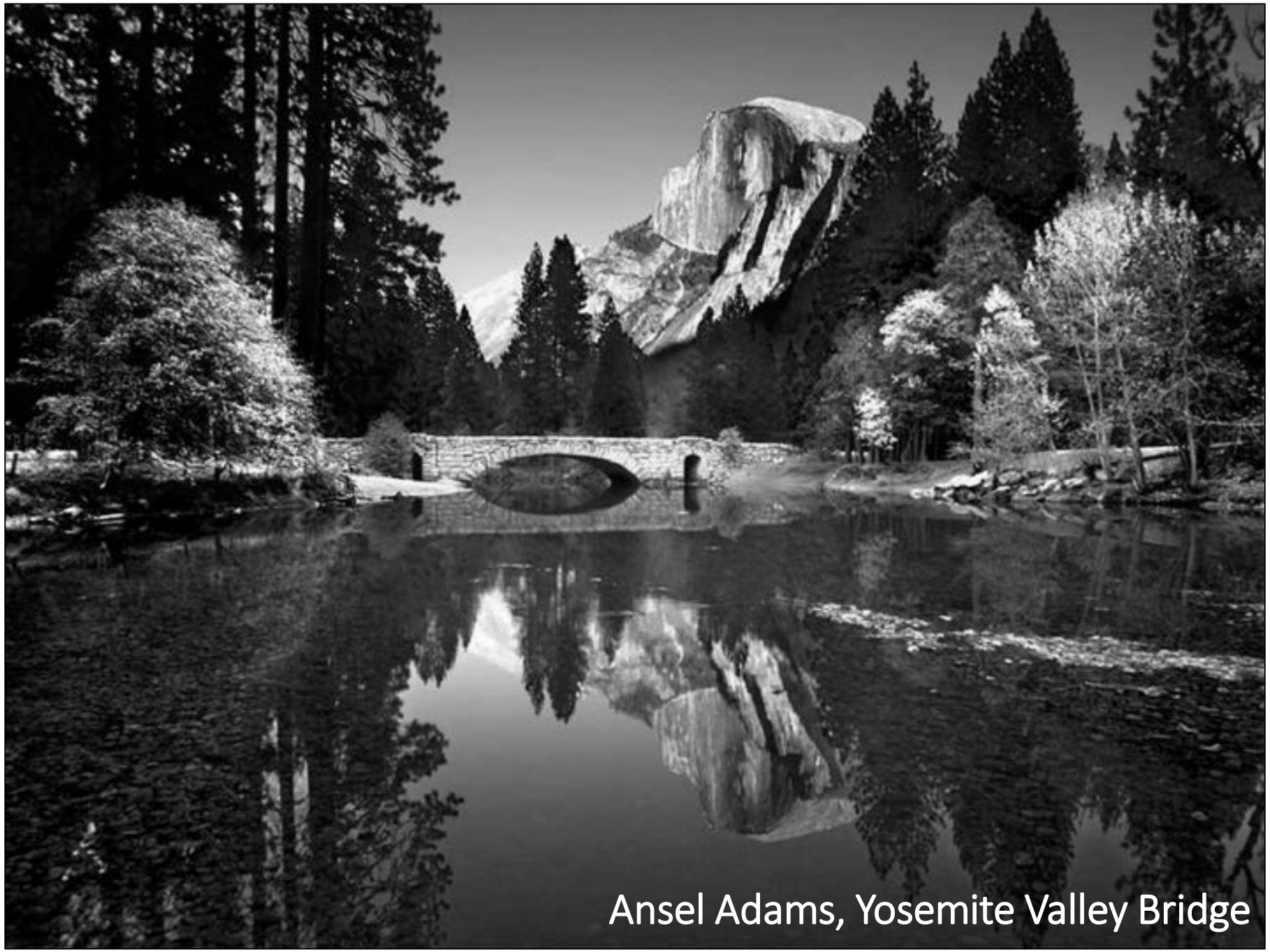



Colorization with Deep Convolutional Networks

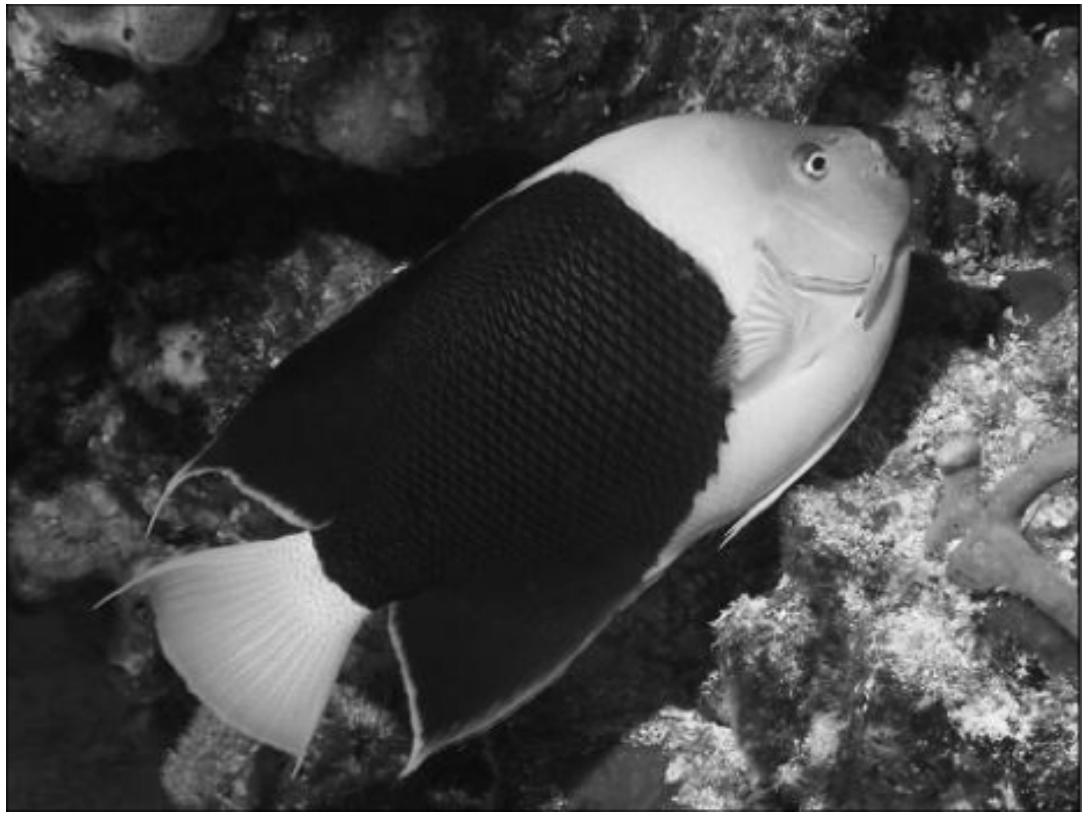
Richard Zhang (章睿嘉)
University of California, Berkeley

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH



Ansel Adams, Yosemite Valley Bridge


Ansel Adams, Yosemite Valley Bridge – Our Result



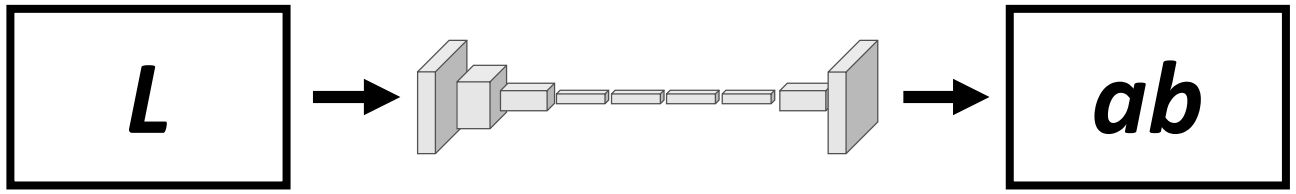
$\mathcal{F}$

Grayscale image: *L* channel

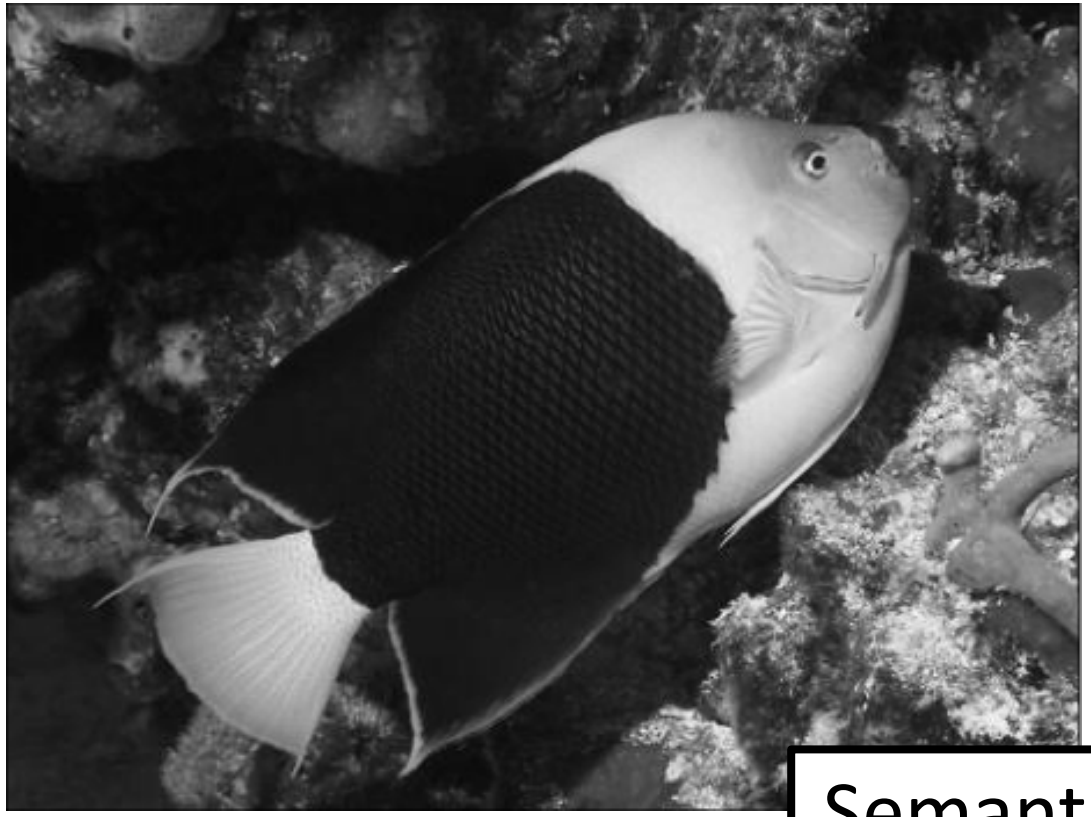
$$\mathbf{X} \in \mathbb{R}^{H \times W \times 1}$$

Color information: *ab* channels

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



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

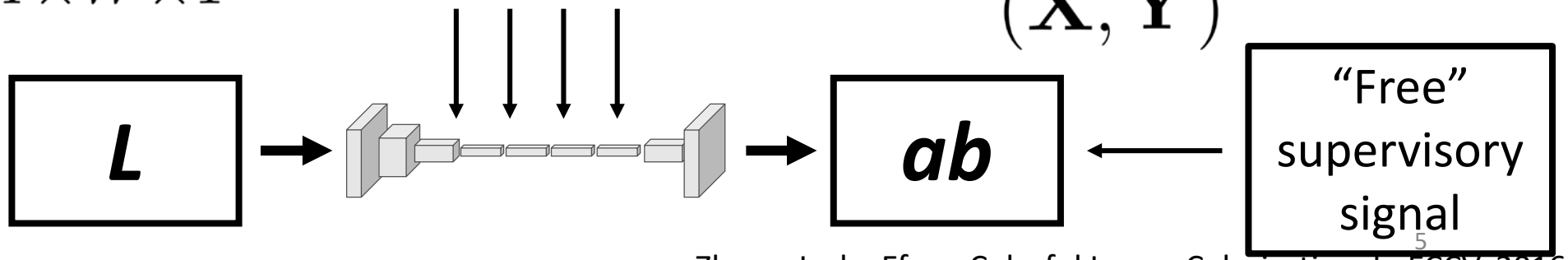


$\mathcal{F}$

Semantics? Higher-level abstraction?

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

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



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

Training a Deep Network

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

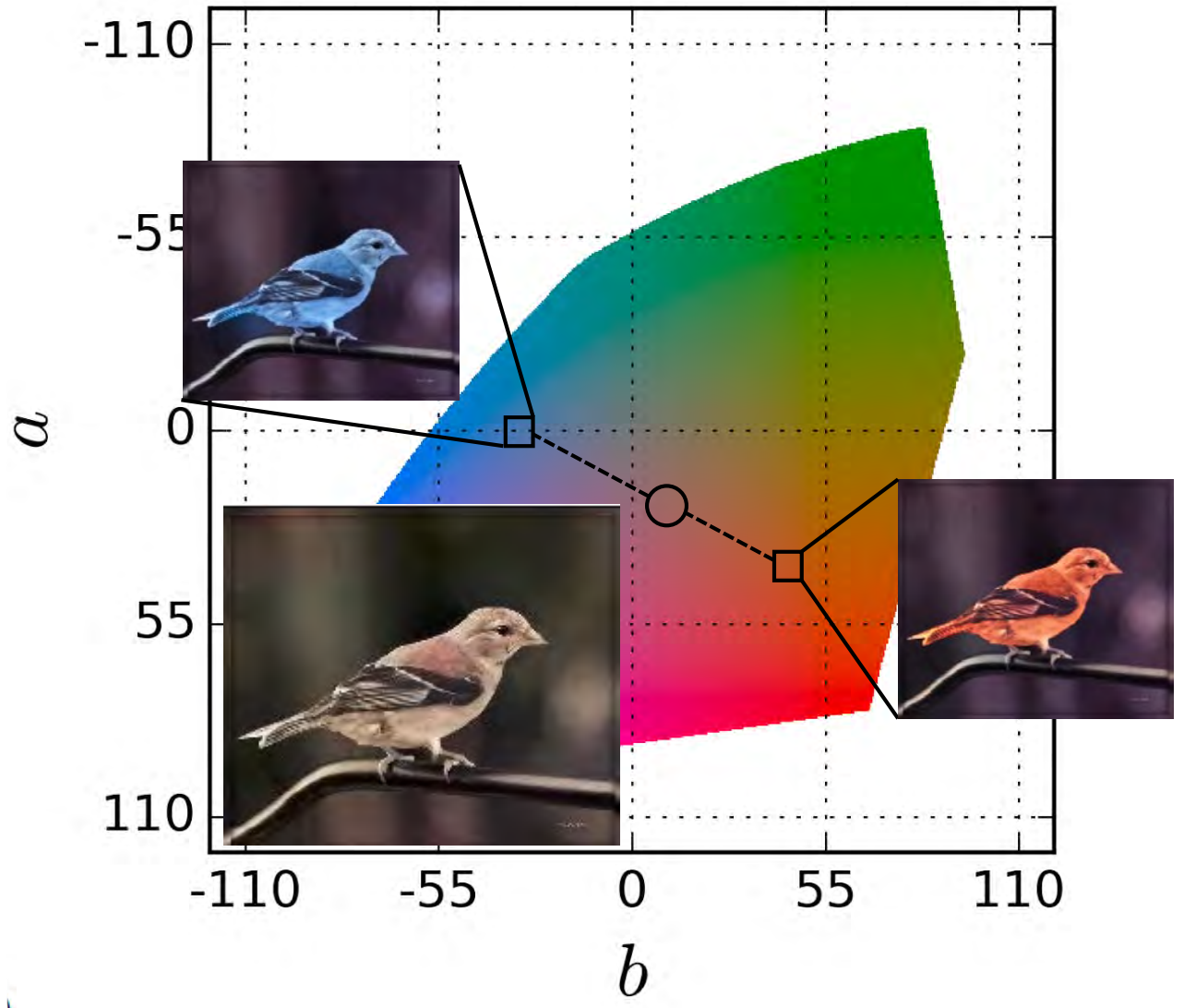
Training data

Loss Function

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

Colors in *ab* space

(continuous)



Loss Function

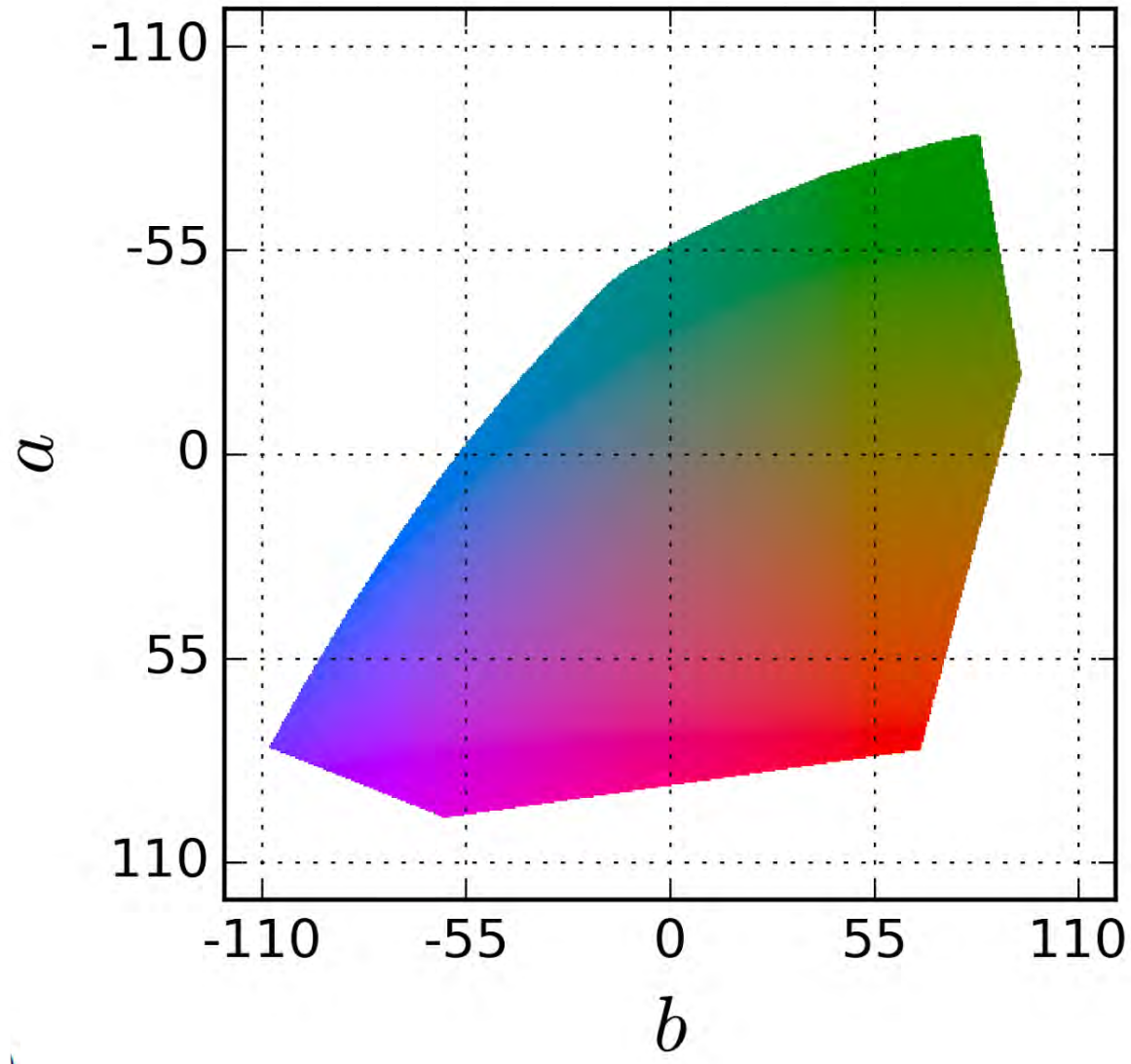
$$\theta^* = \arg\min_{\theta} \ell(\mathcal{F}_\theta(\mathbf{X}), \mathbf{Y})$$

- Regression with L2 loss inadequate

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

Colors in *ab* space

(continuous)



Better Loss Function

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

- Regression with L2 loss inadequate

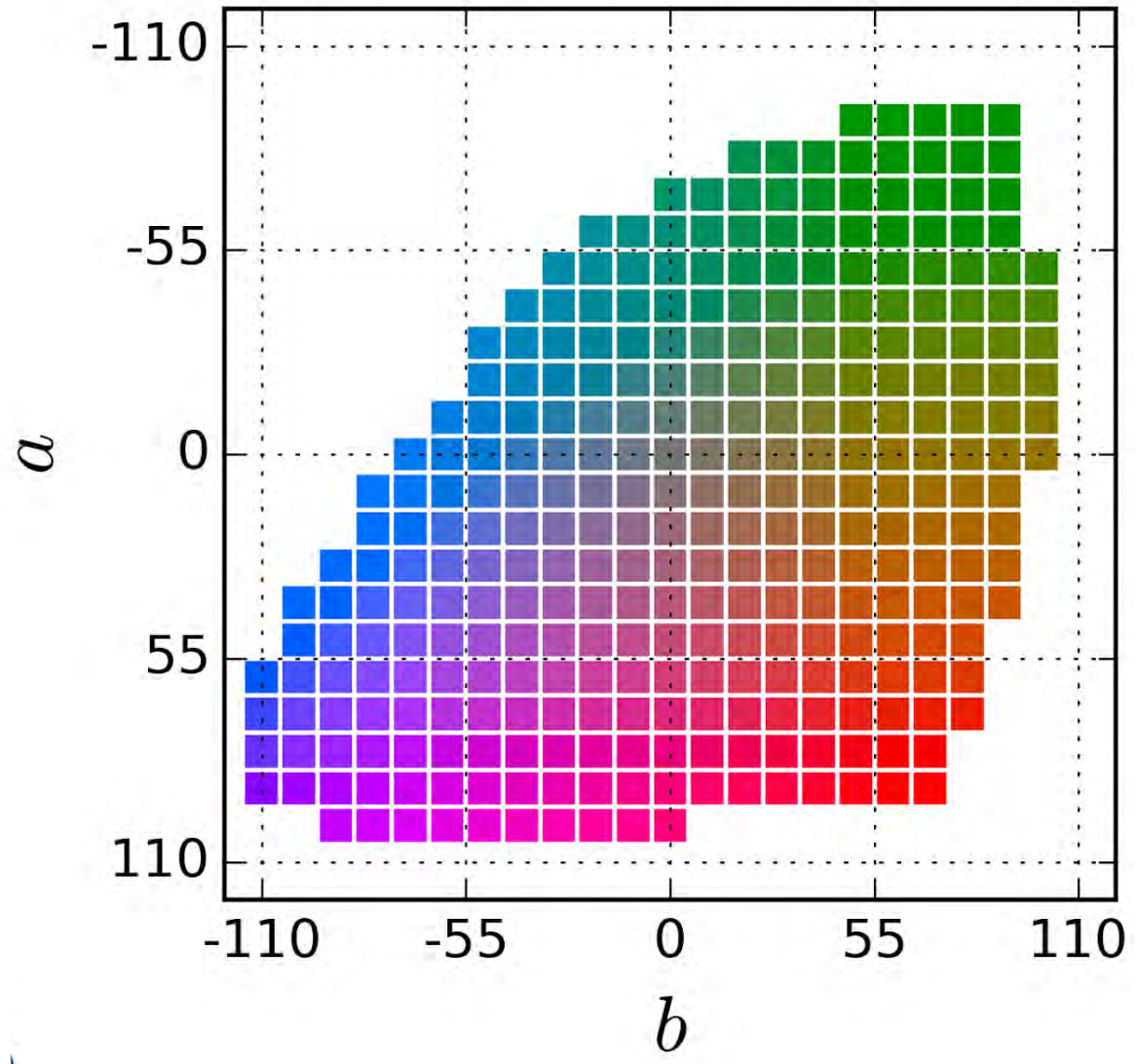
$$\mathrm{L}_2(\hat{\mathbf{Y}}, \mathbf{Y}) = \frac{1}{2}\sum_{h,w} \|\mathbf{Y}_{h,w} - \hat{\mathbf{Y}}_{h,w}\|_2^2$$

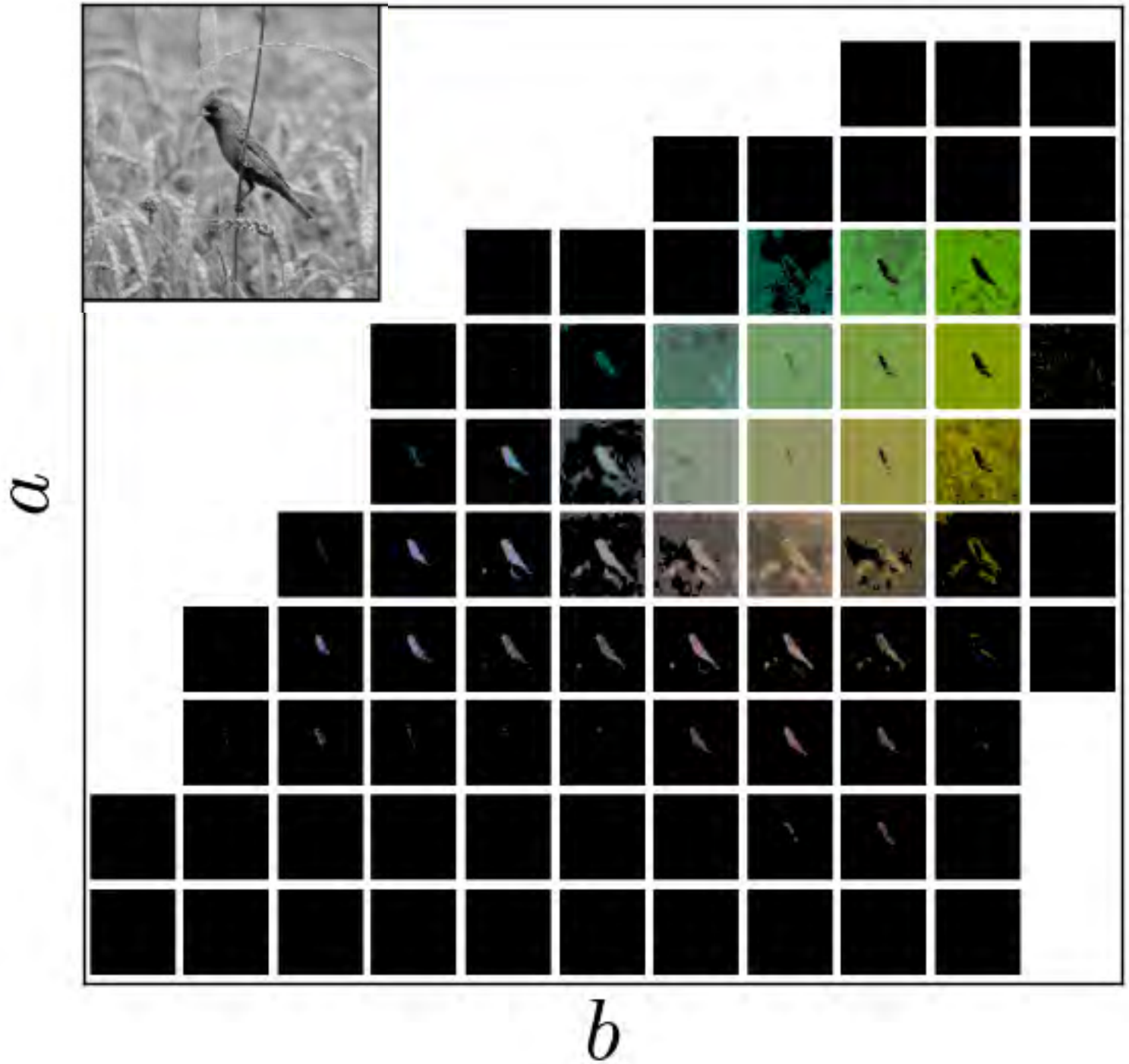
- Use **multinomial classification**

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

Colors in *ab* space

(discrete)




a
b

Better Loss Function

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

- Regression with L2 loss inadequate

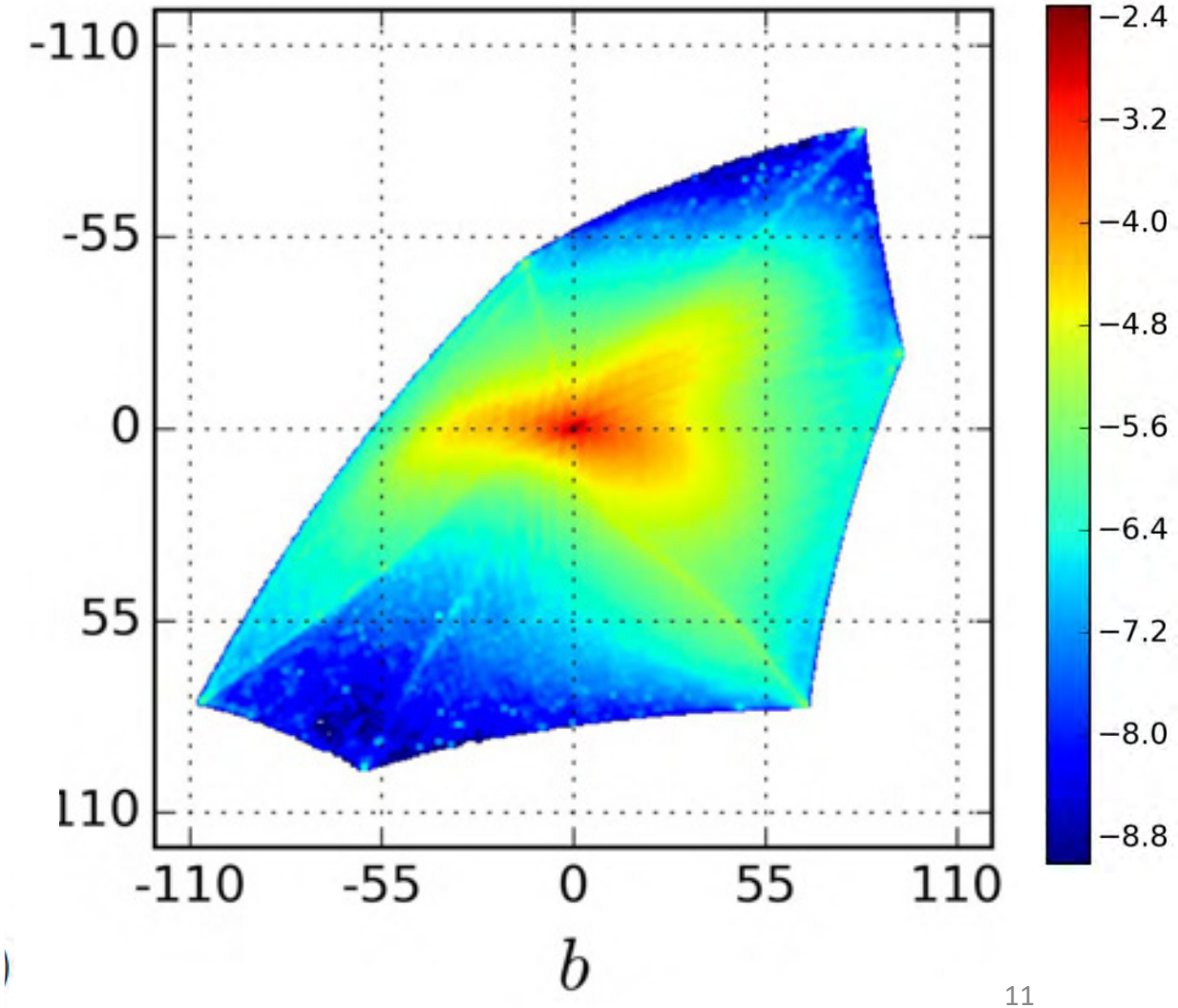
$$\mathrm{L}_2(\widehat{\mathbf{Y}}, \mathbf{Y}) = \frac{1}{2}\sum_{h,w} \|\mathbf{Y}_{h,w} - \widehat{\mathbf{Y}}_{h,w}\|_2^2$$

- Use **multinomial classification**

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

$\log_{10}$ probability

Histogram over *ab* space



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 **multinomial classification**

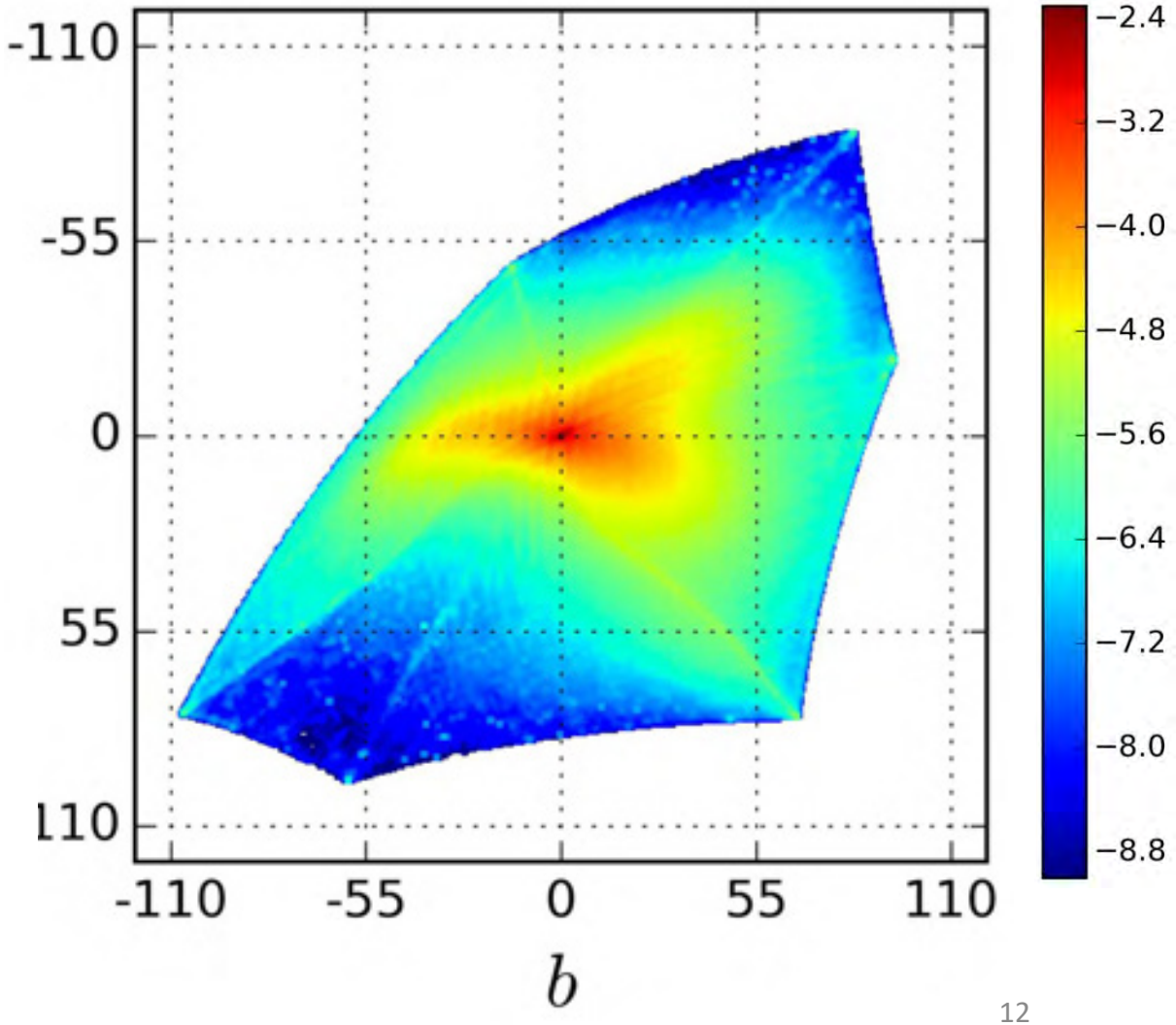
$$\mathrm{L}(\widehat{\mathbf{Z}}, \mathbf{Z}) = -\frac{1}{HW} \sum_{h,w} \sum_{q} \mathbf{Z}_{h,w,q} \log(\widehat{\mathbf{Z}}_{h,w,q})$$

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

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

$\log_{10}$ probability

Histogram over *ab* space



Ground truth

Input

L2 Regression

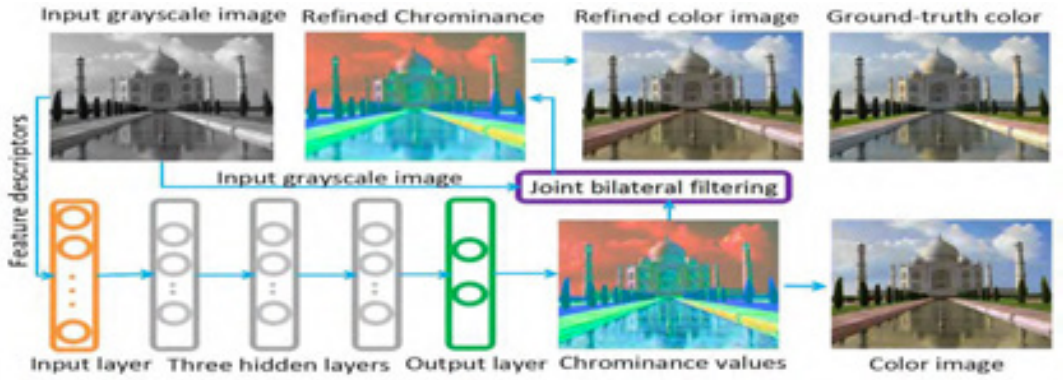
Class w/ Rebalancing

Prior Work

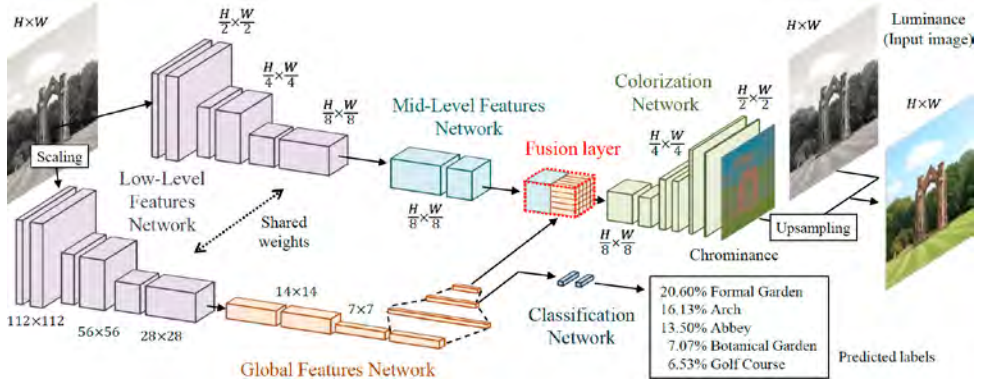
Hand-engineered Features

Deep Networks

L2 Regression

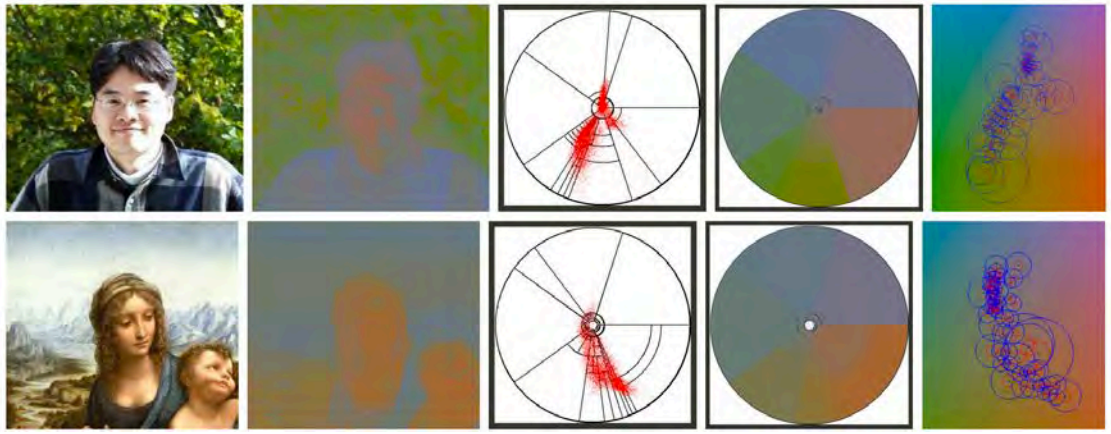


Deshpande et al. Cheng et al. In ICCV 2015.

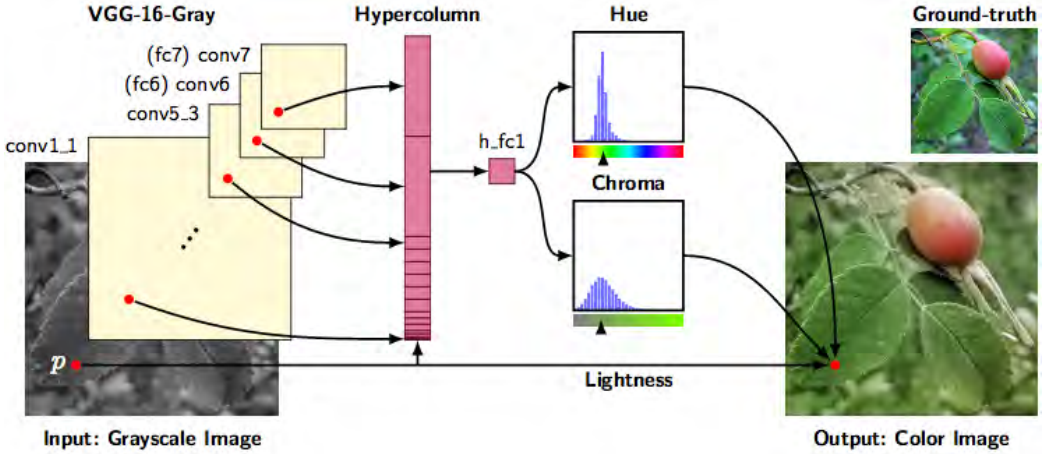


Dahl. Jan 2016. Iizuka et al. In SIGGRAPH, 2016.

Classification



Charpiat et al. In ECCV 2008.



Larsson et al. In ECCV 2016.

Perceptual Realism / Amazon Mechanical Turk Test

Fake, 0% fooled

Fake, 55% fooled

Fake, 58% fooled

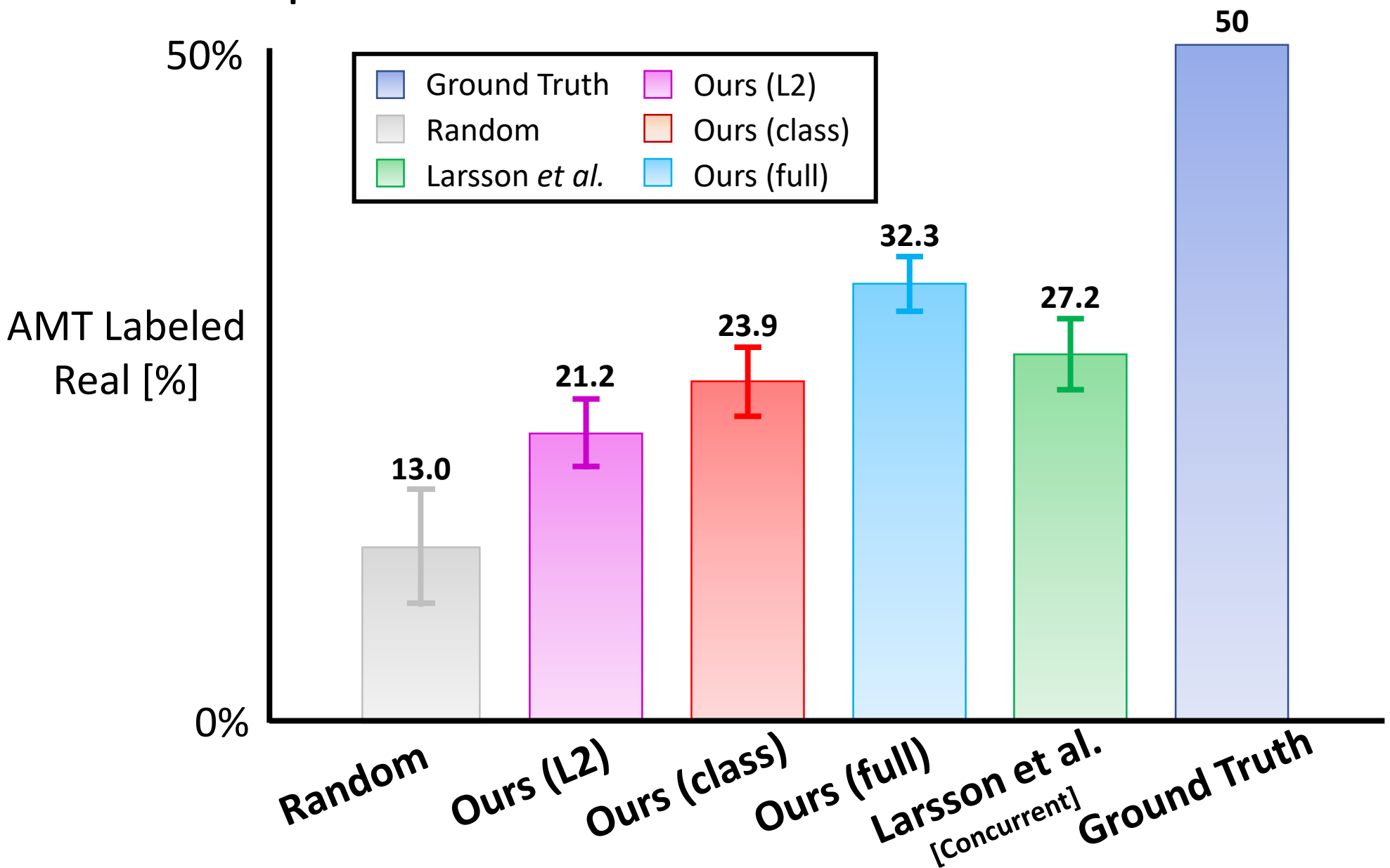
from Reddit /u/SherySantucci

Recolorized by Reddit ColorizeBot

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

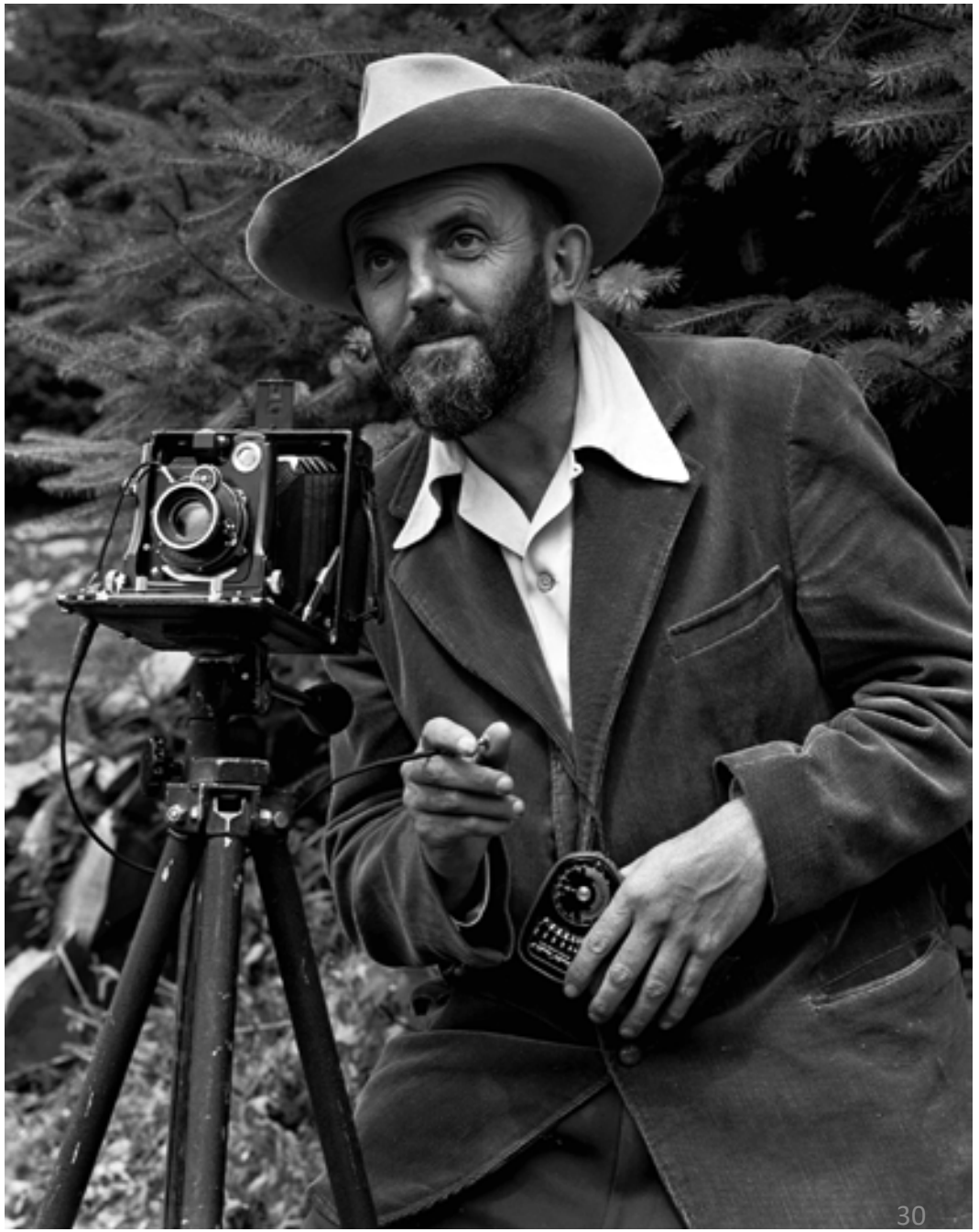
Recolorized by Reddit ColorizeBot

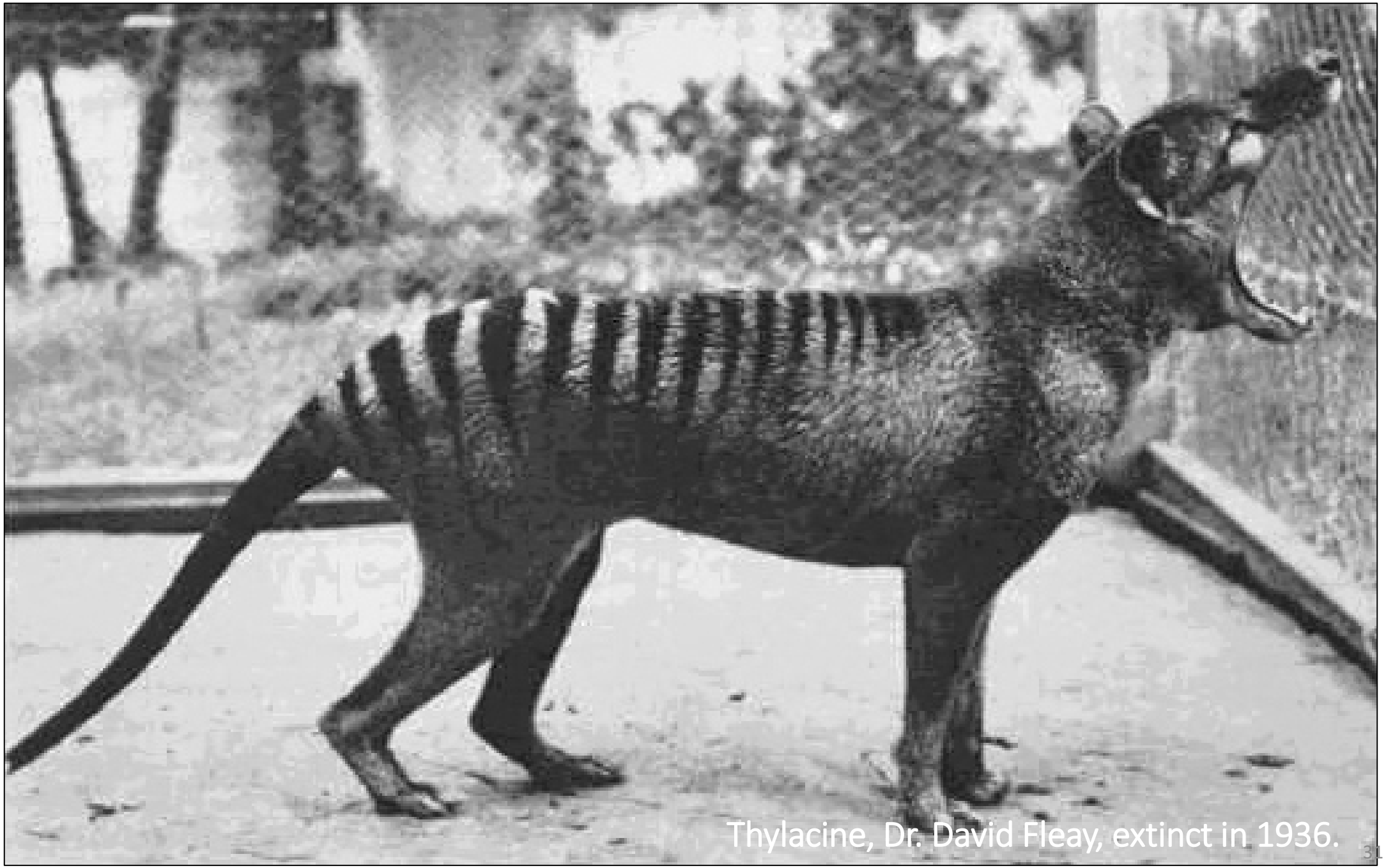
Perceptual Realism Test



1600 images tested per algorithm

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



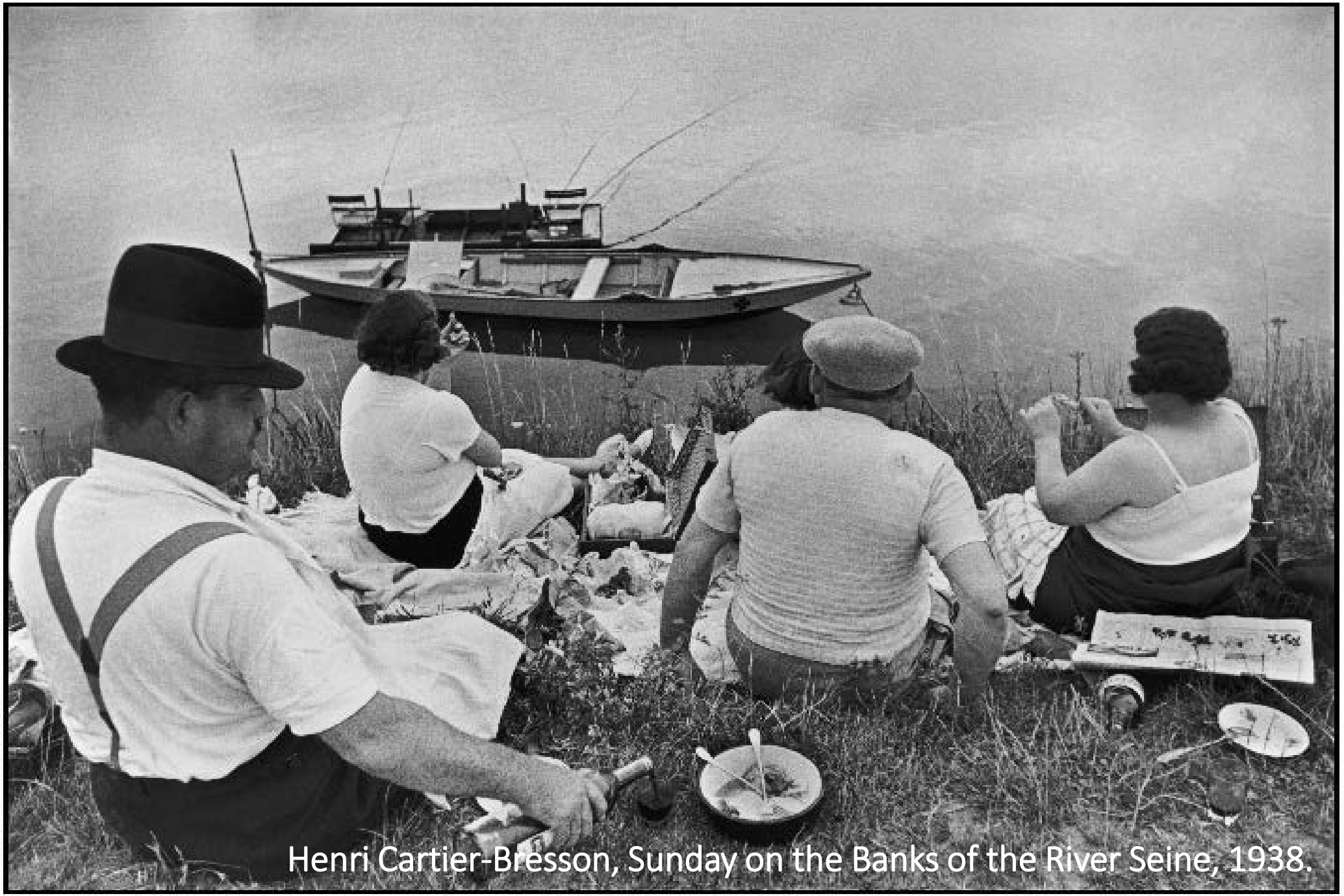


Thylacine, Dr. David Fleay, extinct in 1936.

Thylacine, Dr. David Fleay, extinct in 1936.

Amateur Family Photo, 1956.

Amateur Family Photo, 1956.



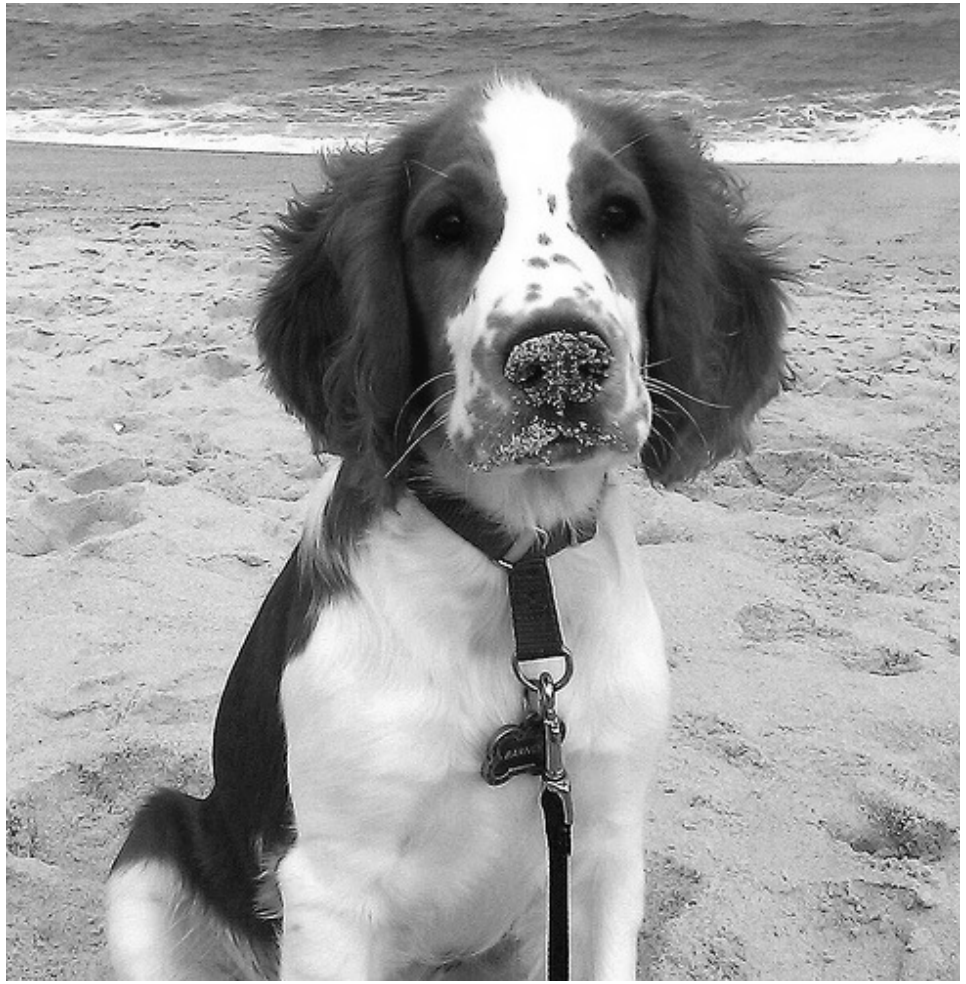
Henri Cartier-Bresson, Sunday on the Banks of the River Seine, 1938.

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

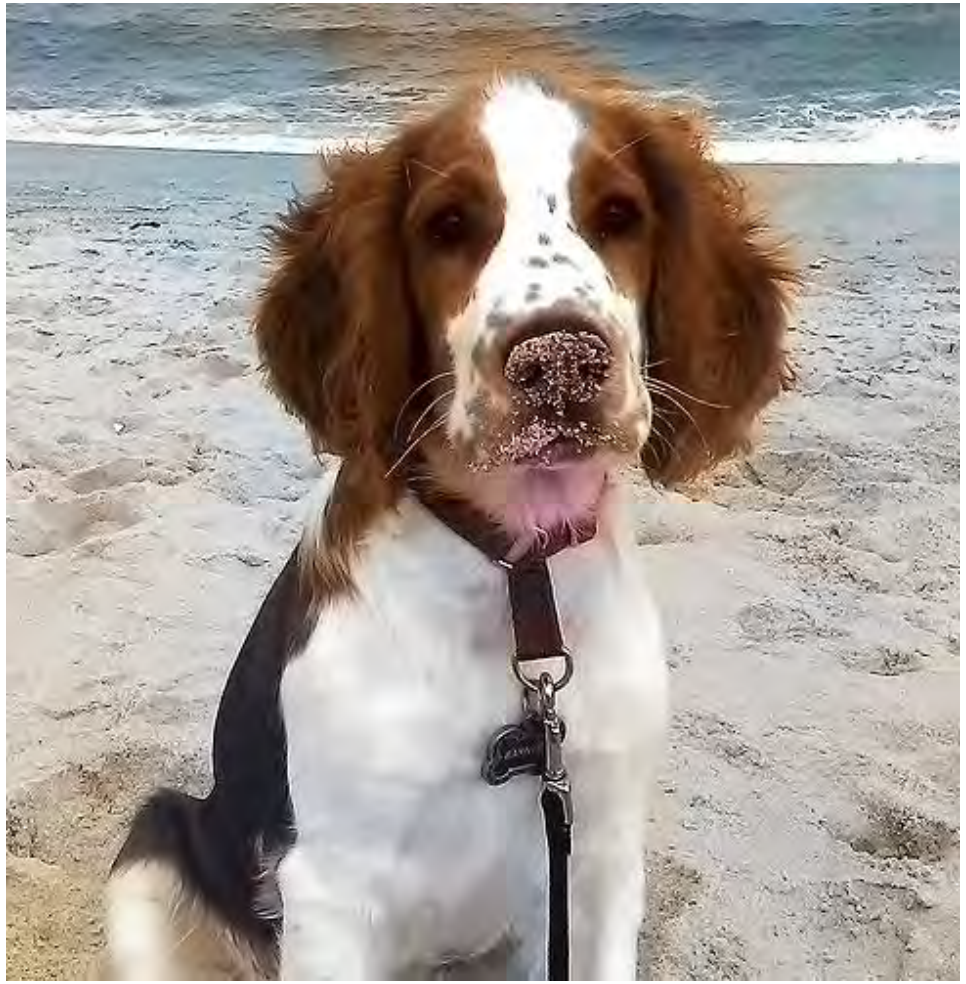
Failure Cases

Failure Cases

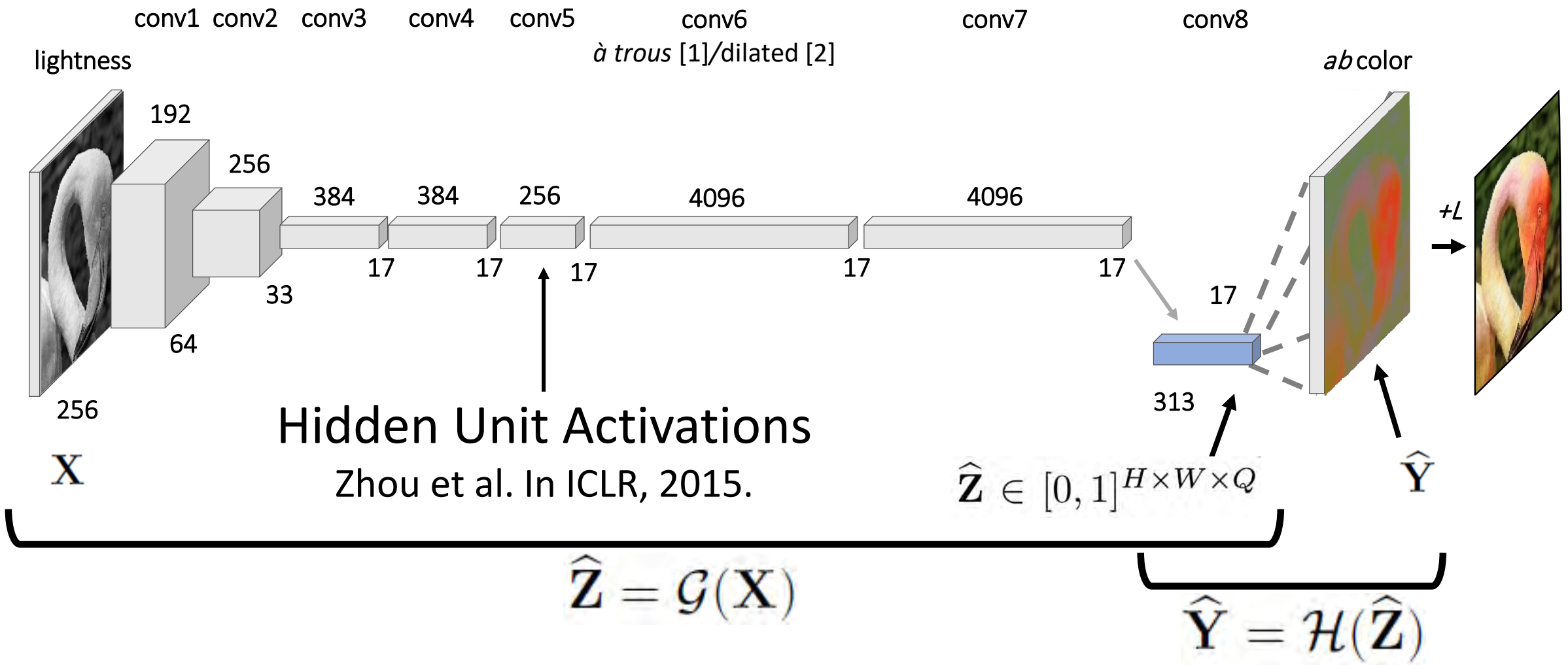
Informative Mistakes



Informative Mistakes



Cross-Channel Encoder



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

Hidden Unit (conv5) Activations

sky

trees

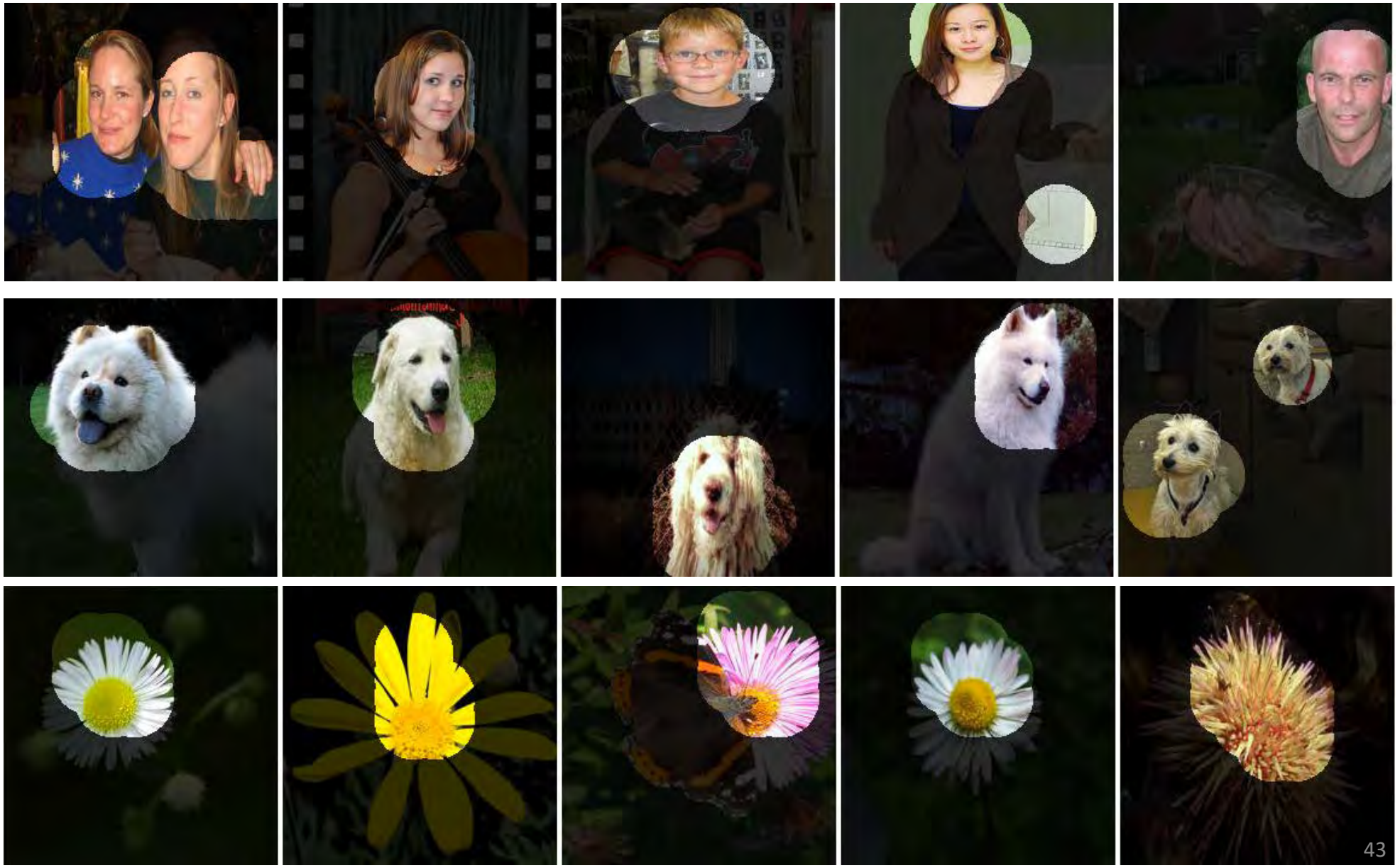
water

Hidden Unit (conv5) Activations

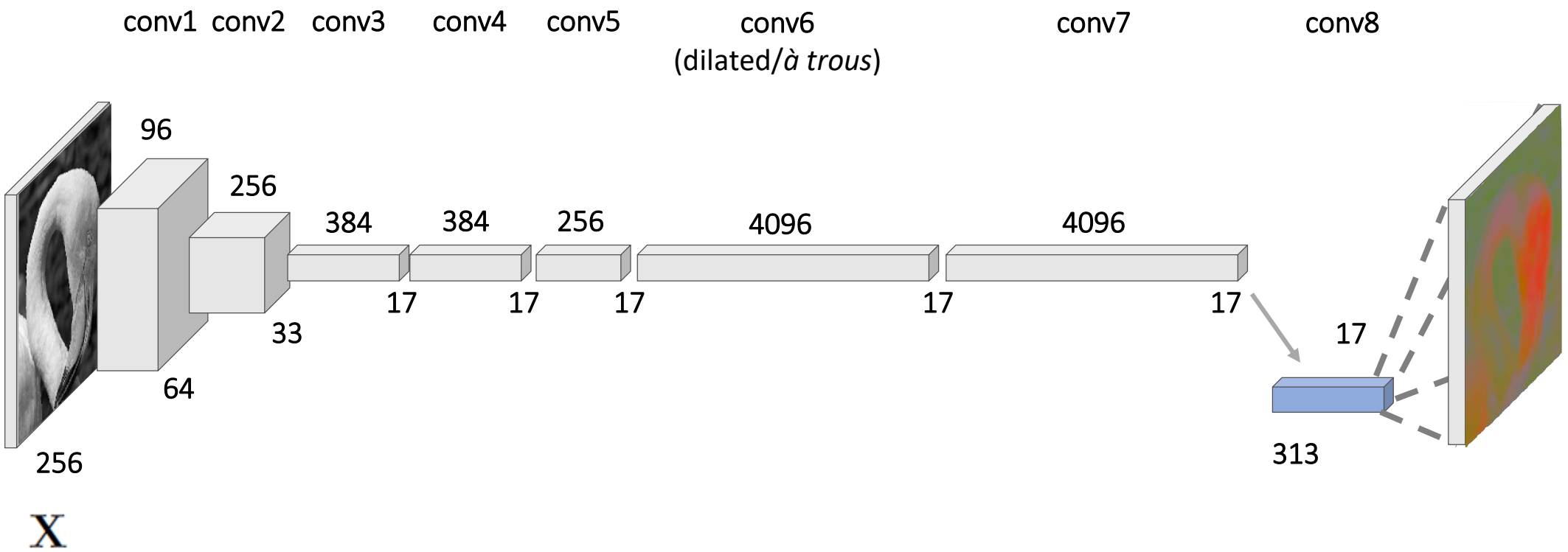
faces

dog faces

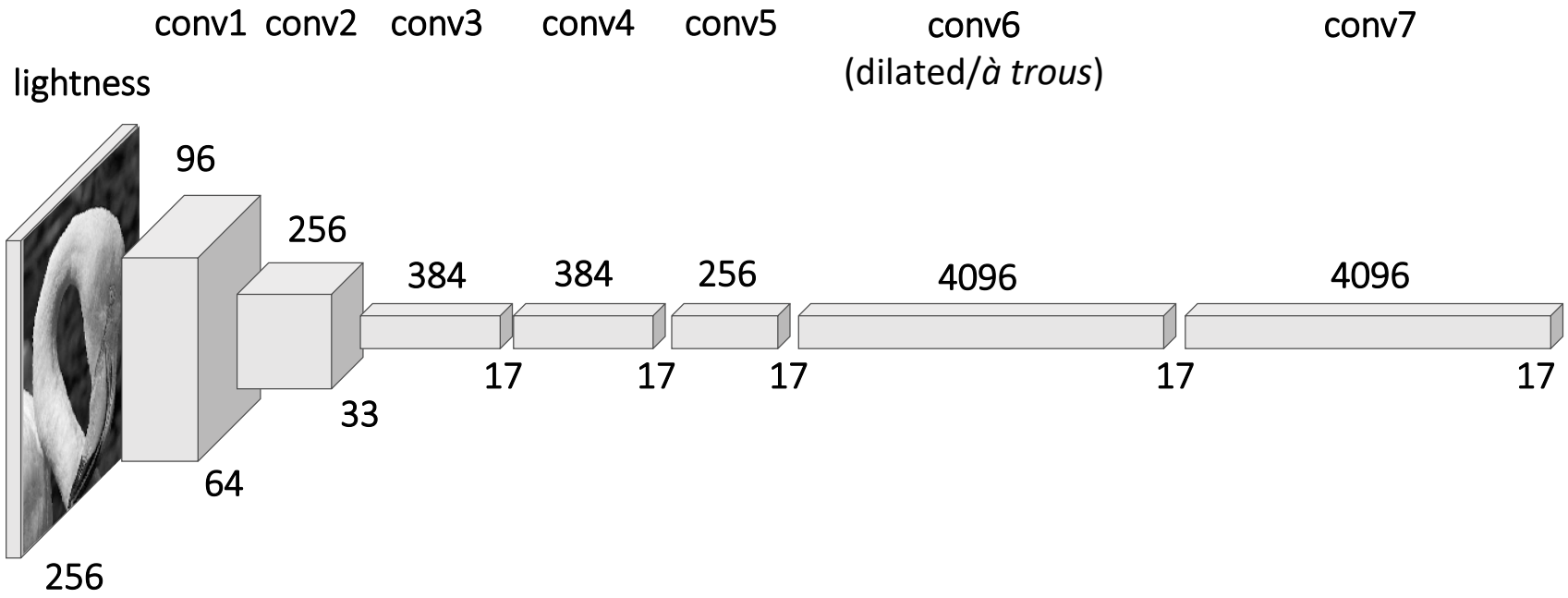
flowers



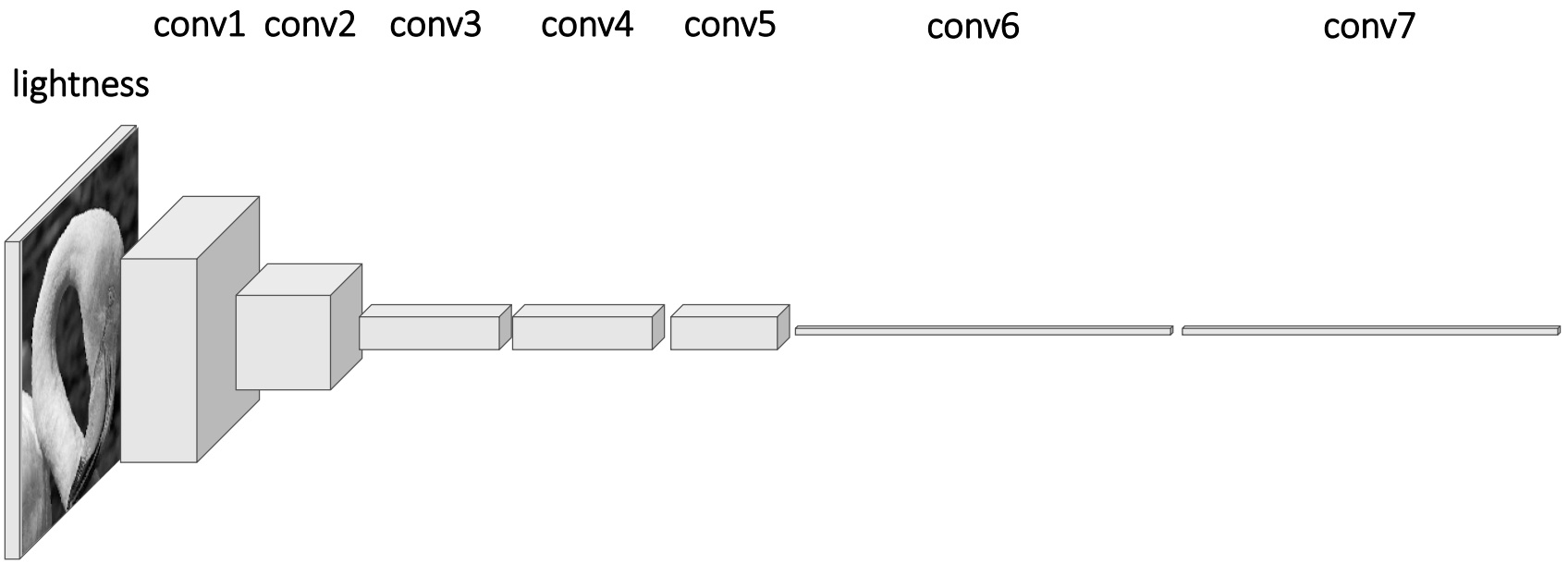
Pre-training on Colorization



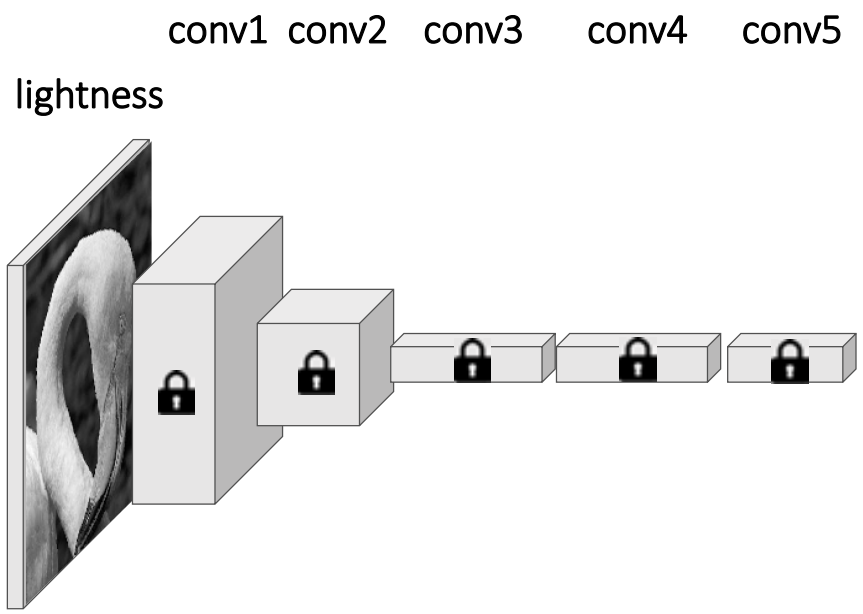
Pre-Trained Model (Fully Convolutional)



Pre-Trained Model

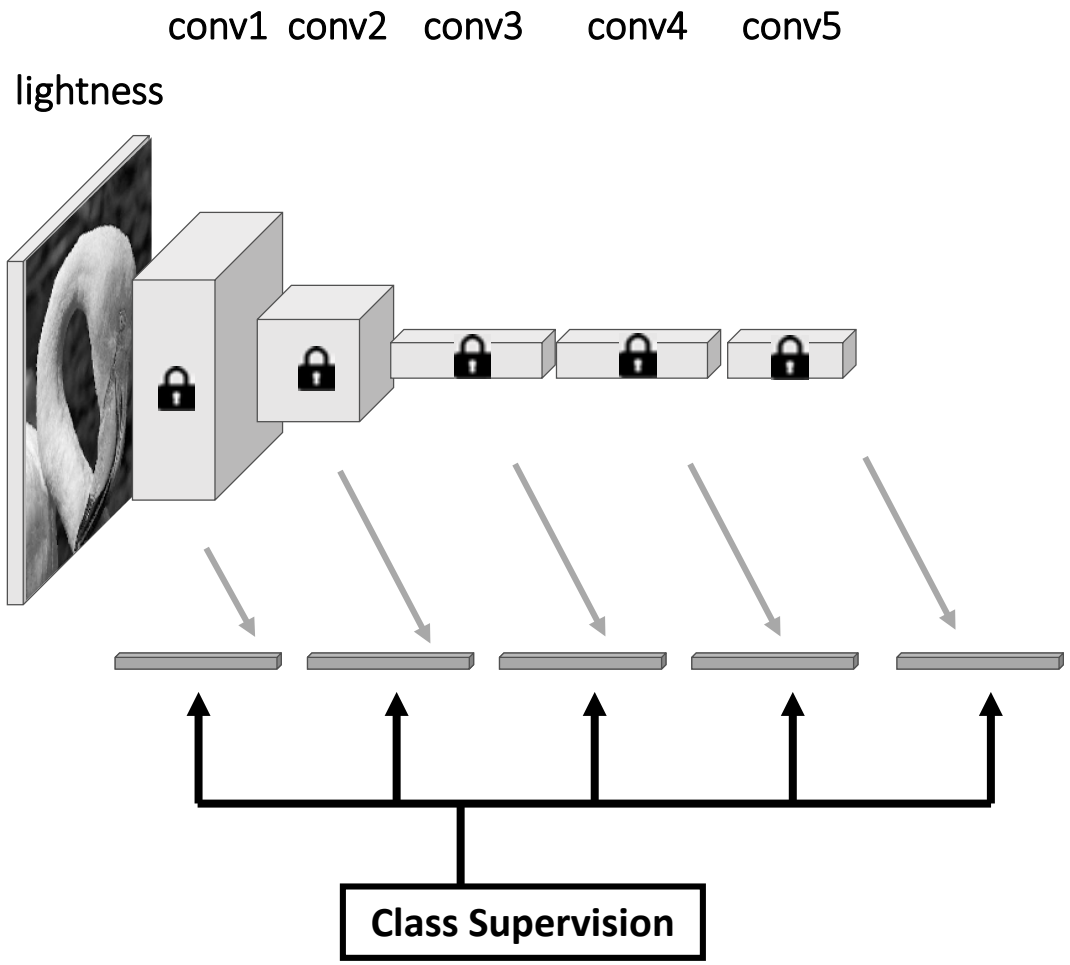


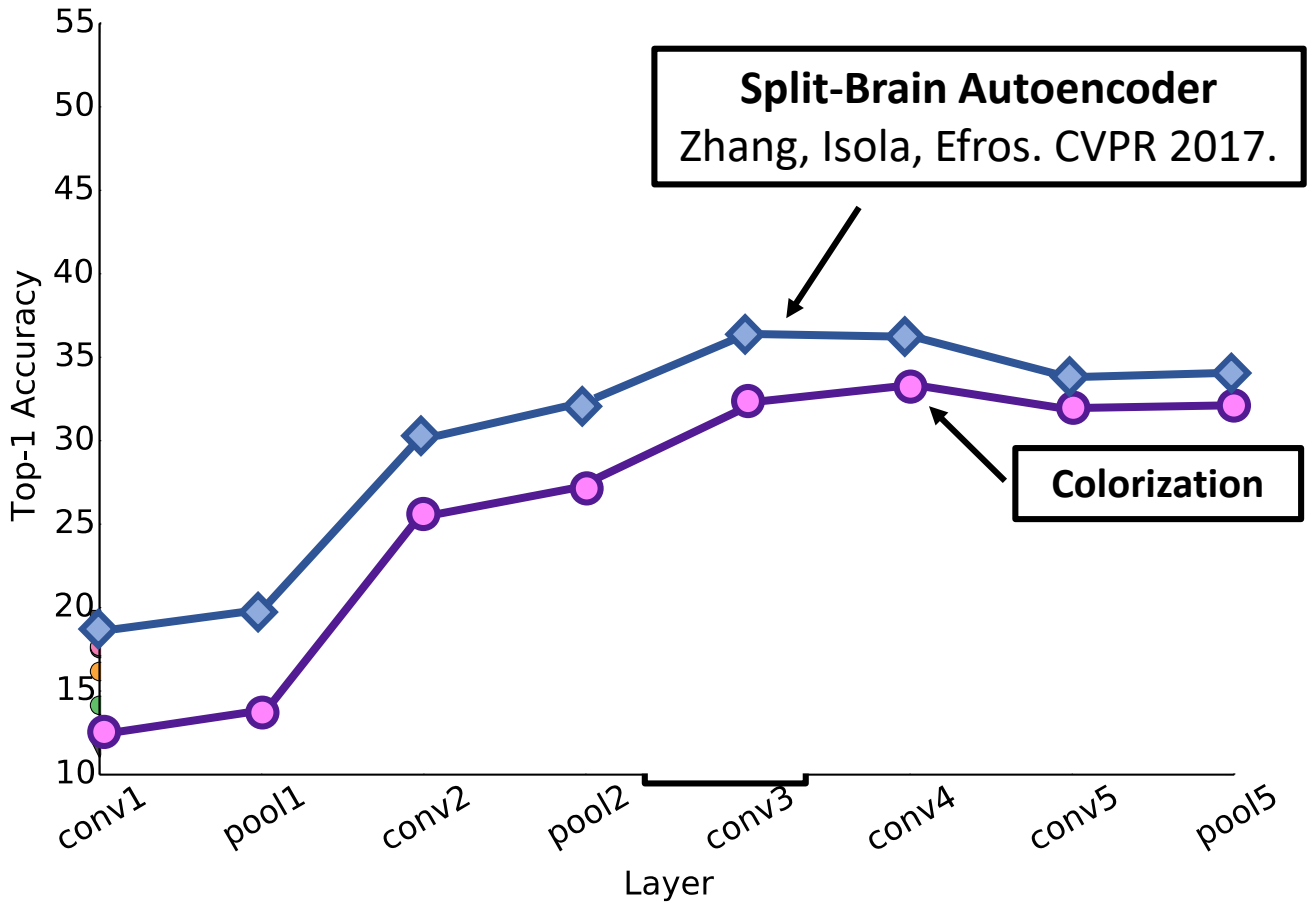
Pre-Trained Model



Pre-Trained Model

How does solving the pre-text task organize the feature space?

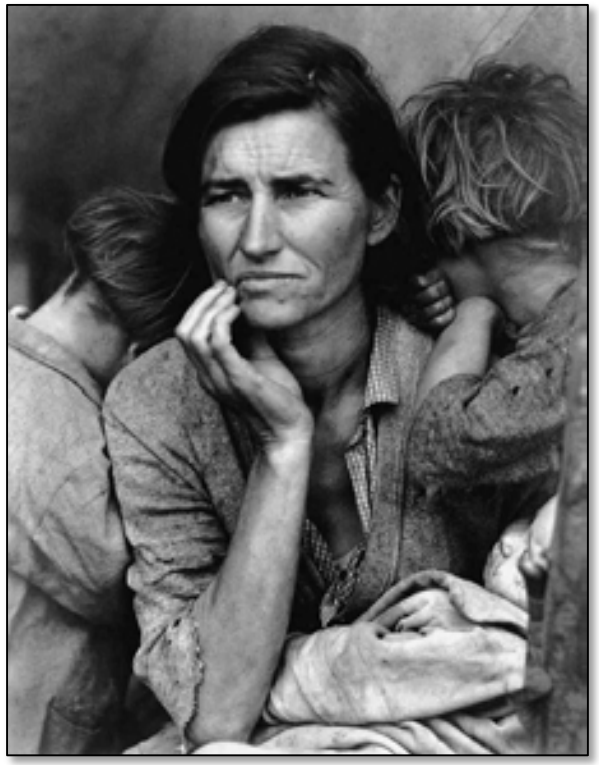




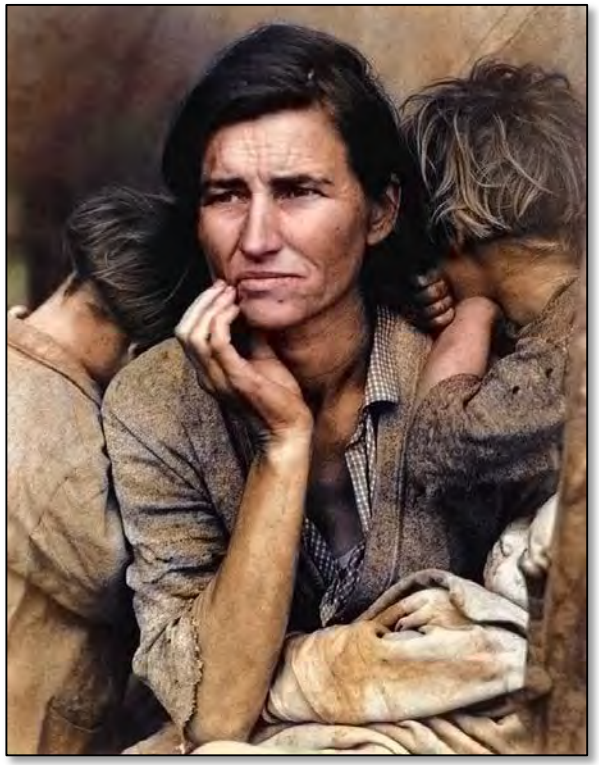
Migrant Mother
Dorothea Lange, 1936.

Library of Congress, Prints & Photographs Division, FSA/OWI Collection, reproduction number: LC-USF34-9058-C

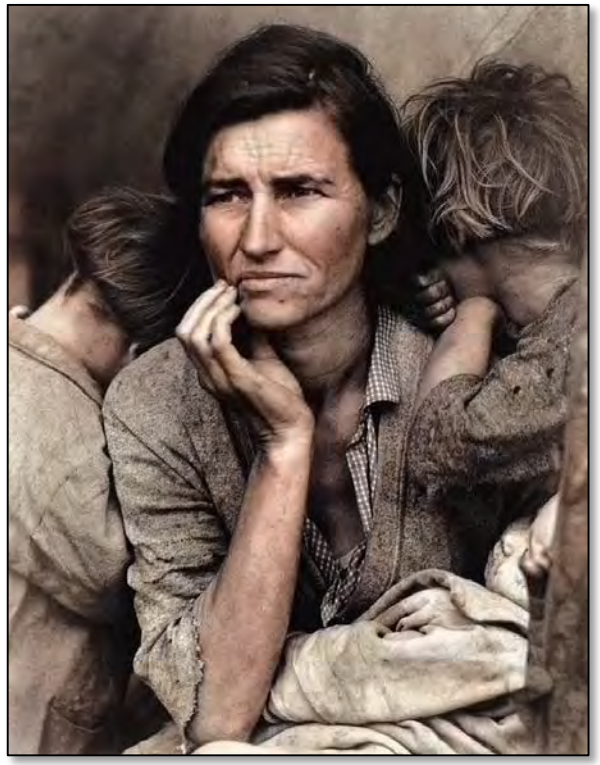
Automatic Results with Deep Networks



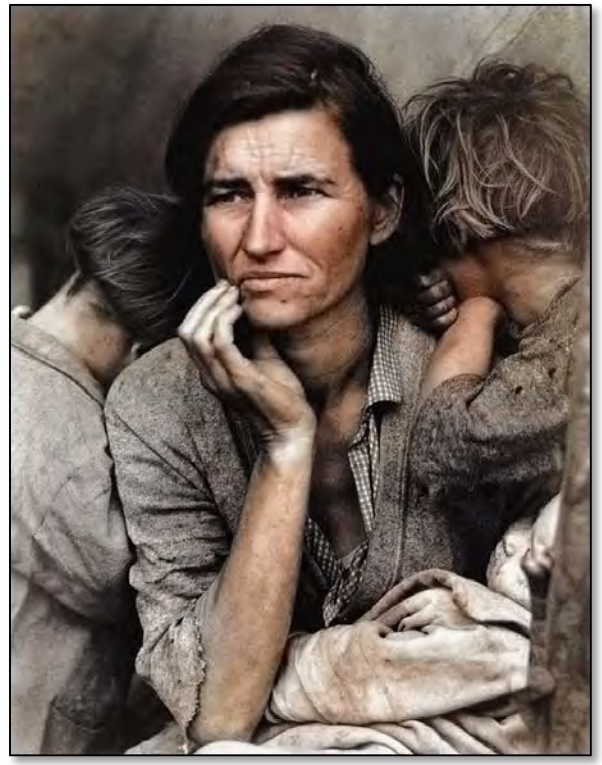
Grayscale Input



Zhang, Isola, Efros.
ECCV 2016.



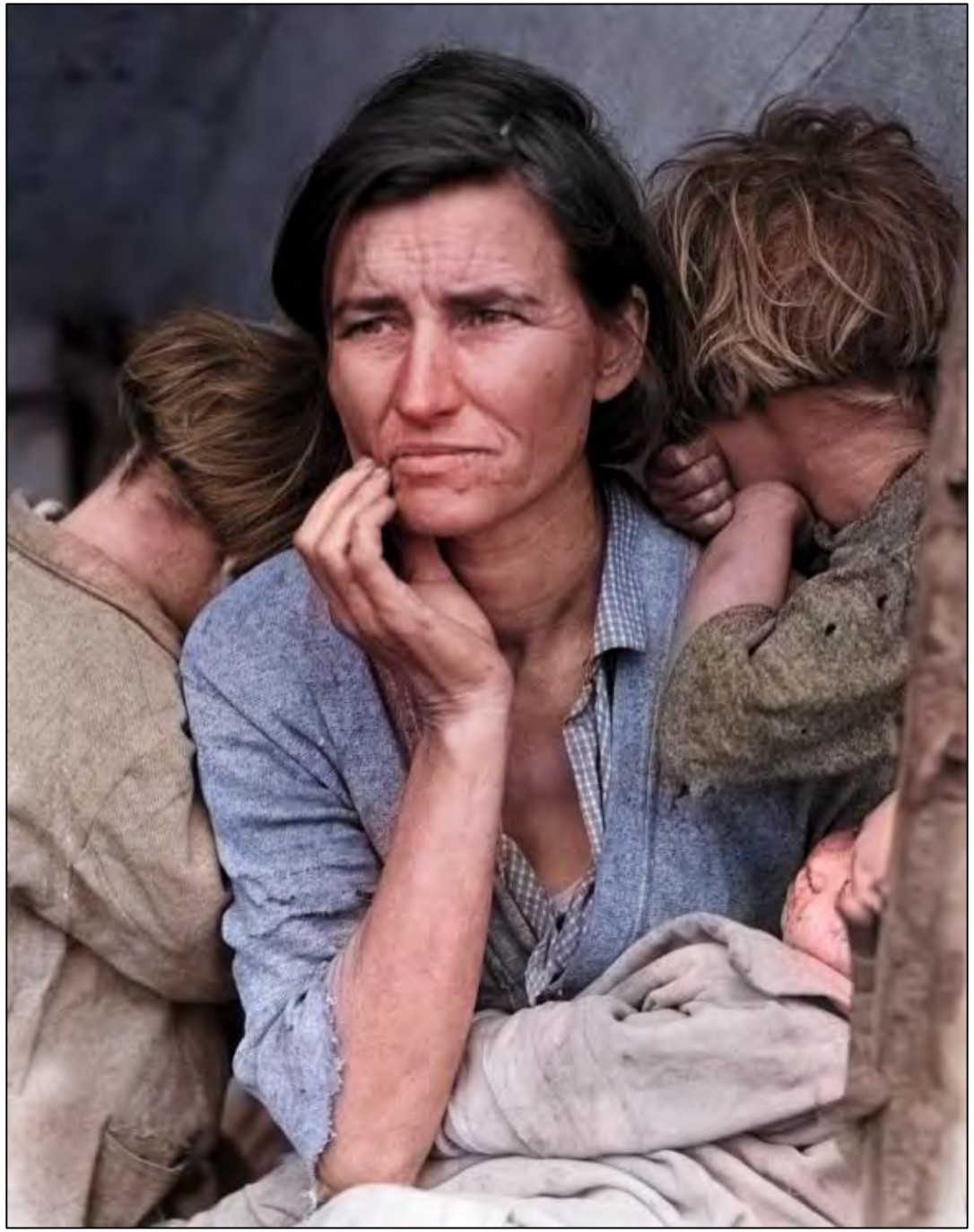
Larsson *et al.*
ECCV 2016.



Iizuka *et al.*
SIGGRAPH 2016.

Migrant Mother. Dorothea Lange, 1936.

Library of Congress, Prints & Photographs Division, FSA/OWI Collection, reproduction number: LC-USF34-9058-C

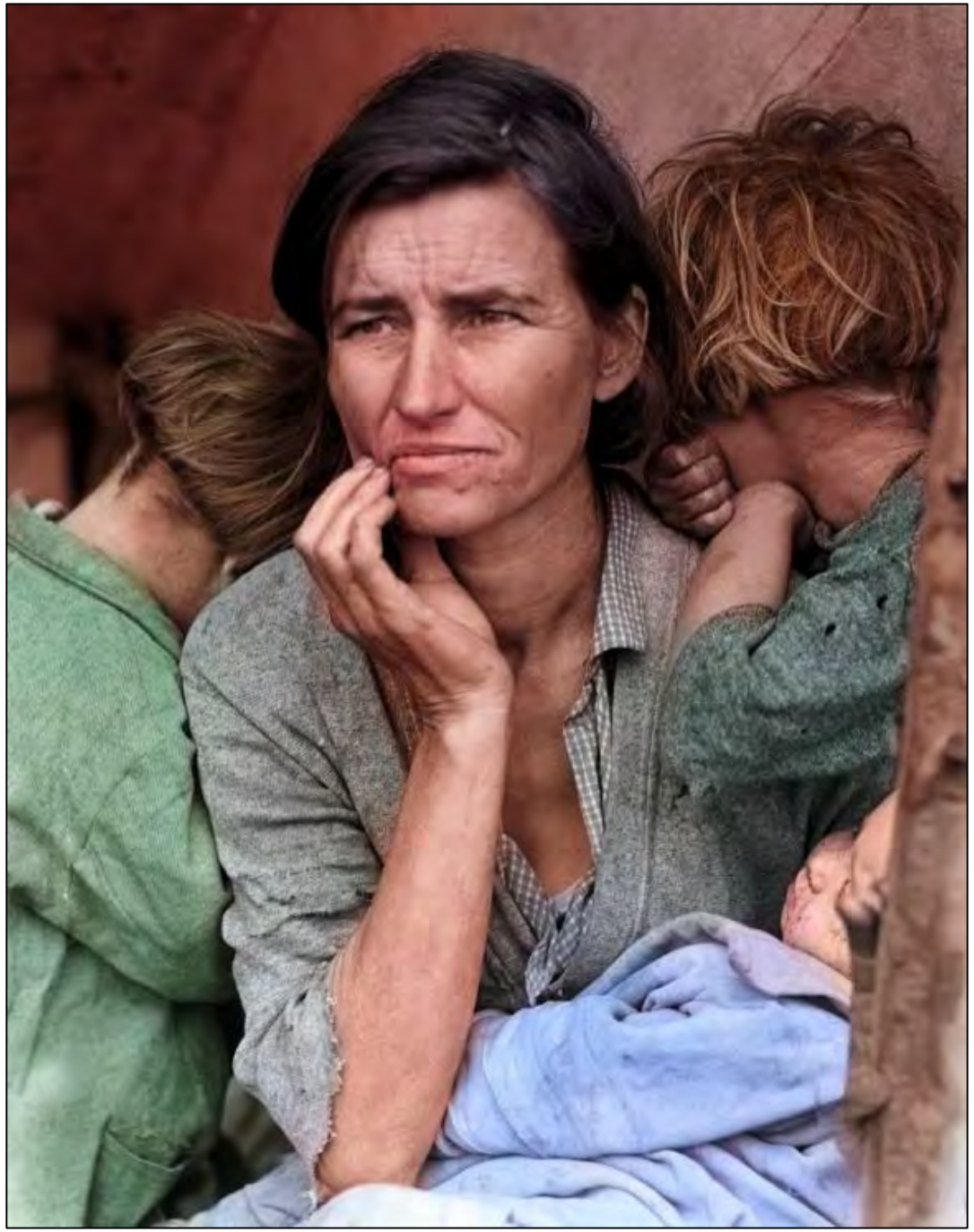


Dorothea Lange
Migrant Mother, 1936.

Library of Congress, Prints & Photographs Division, FSA/OWI Collection, reproduction number: LC-USF34-9058-C

Dorothea Lange
Migrant Mother, 1936.

Library of Congress, Prints & Photographs Division, FSA/OWI Collection, reproduction number: LC-USF34-9058-C

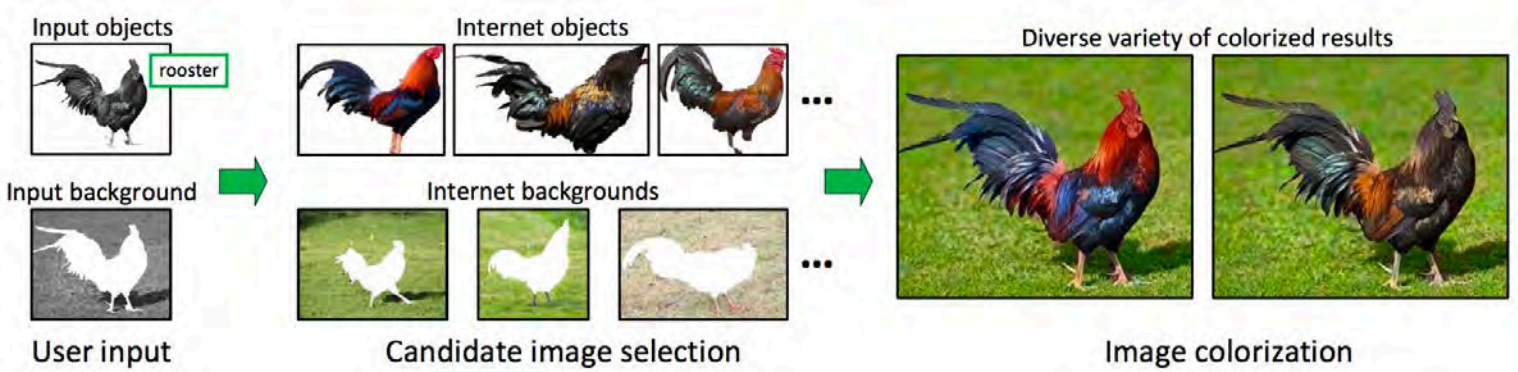


Dorothea Lange
Migrant Mother, 1936.

Library of Congress, Prints & Photographs Division, FSA/OWI Collection, reproduction number: LC-USF34-9058-C

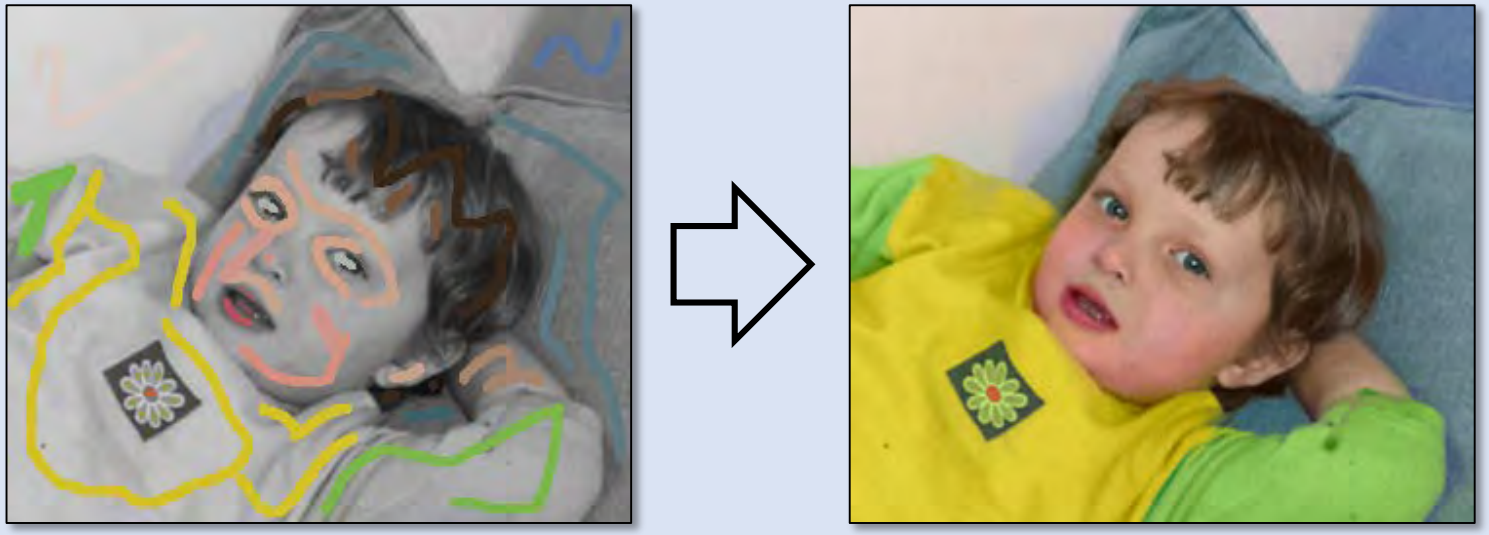
User-Guided Approaches

Global-Based



Hertzmann et al. SIGGRAPH 2001.
Welsh et al. TOG 2002.
Irony et al. Eurographics 2005.
Liu et al. TOG 2008.
Wang et al. TOG 2010.
Chia et al. TOG 2011.
Gupta et al. MM 2012.
Li et al. CVM 2015.
Chang et al. TOG 2015.

Stroke-Based

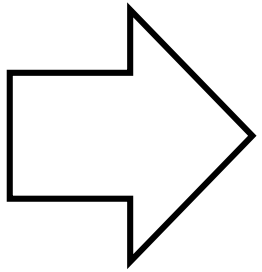


Grayscale image + user strokes

Colorization

Levin et al. SIGGRAPH 2004.
Huang et al. MM 2005.
Qu et al. TOG 2006.
Luan et al. Eurographics 2007.
An and Pellacini. SIGGRAPH 2008.
Li et al. CVM 2008.
Xu et al. TOG 2009.
Chen et al. TOG 2012.
Xu et al. TOG 2013.
Barron and Poole. ECCV 2016.
Endo et al. Eurographics 2016.

User-Guided Colorization



Grayscale image + user strokes

Output Colorization

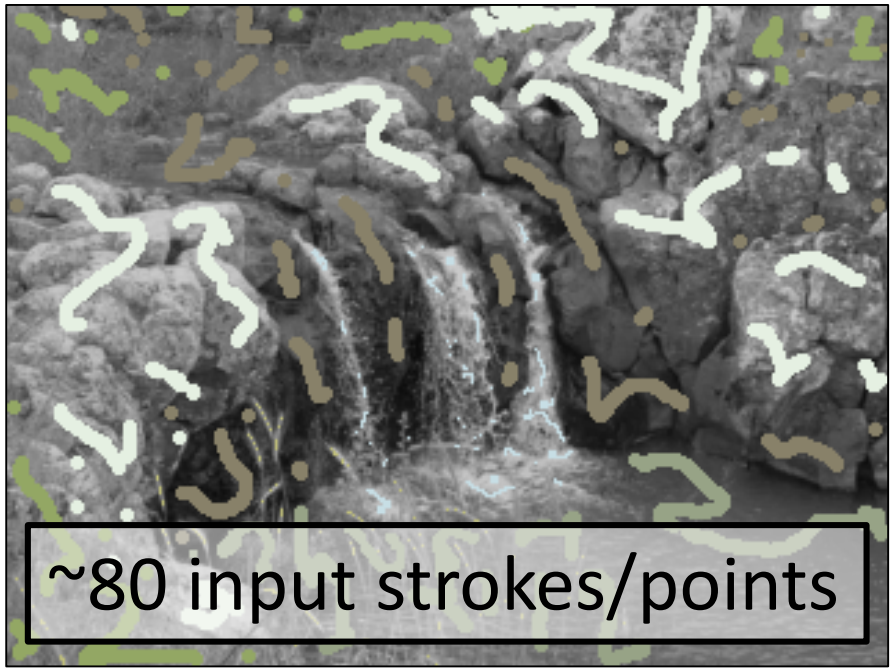
Known pixel colors: provided by user input

Unknown pixel colors: linear combination of neighboring pixel colors, weighted by low-level similarity

+ Propagates user edits + Robust

Levin et al. SIGGRAPH 2004

User-Guided Colorization

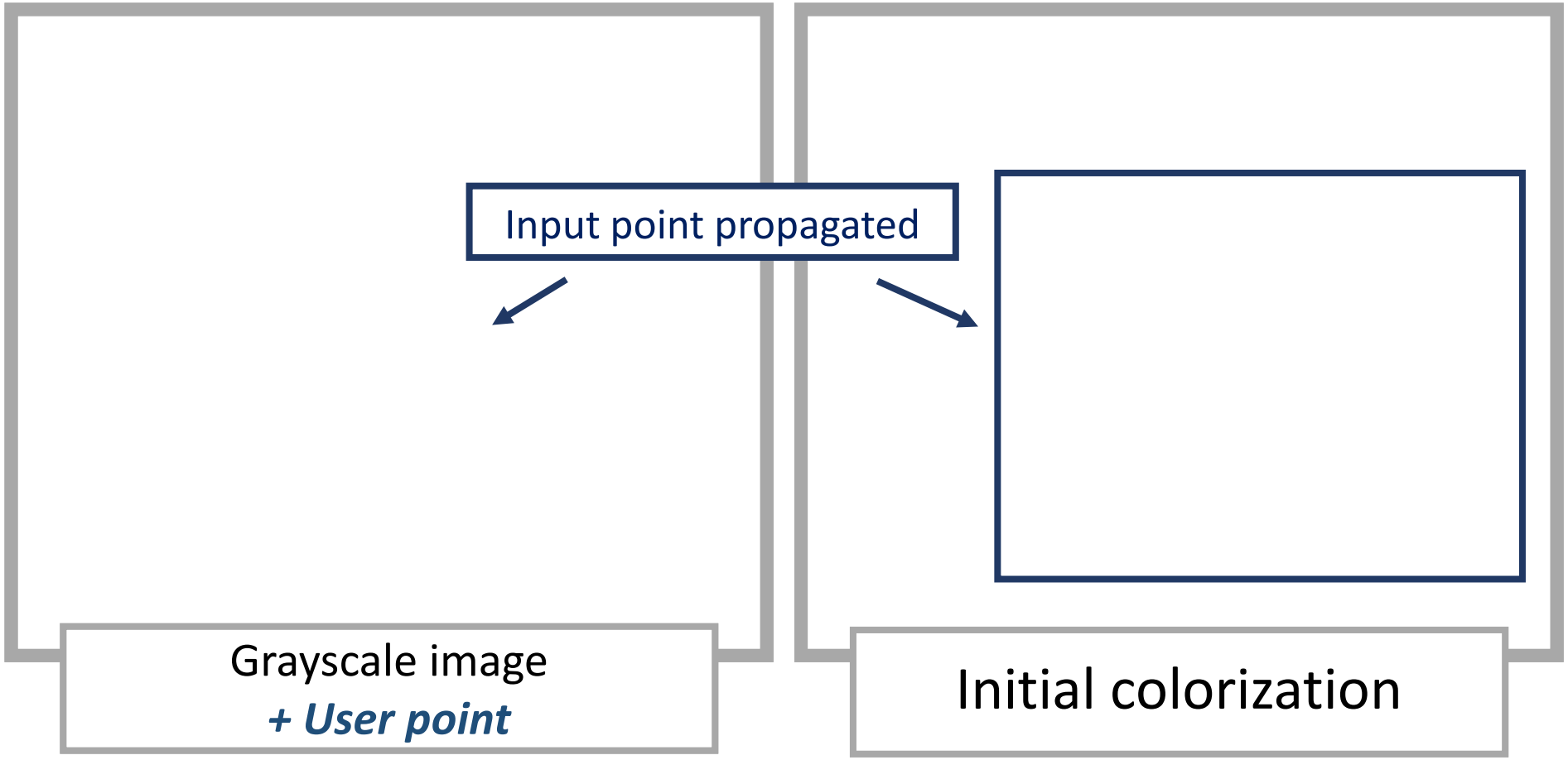


– many user strokes often needed

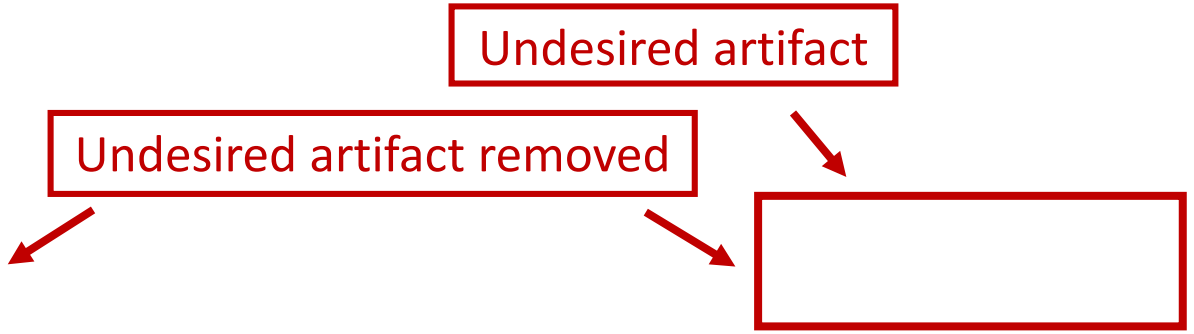
Desired: Learn ***natural image priors*** and ***edit propagation*** from large-scale data

Levin et al. SIGGRAPH 2004

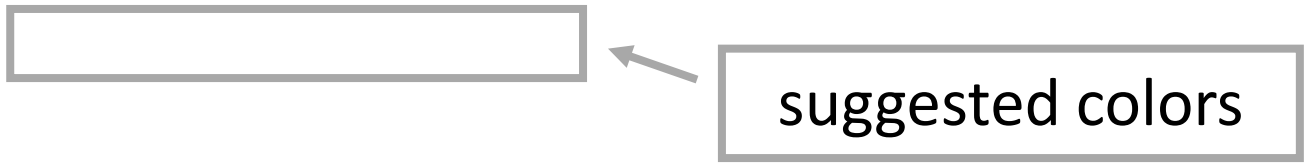
System Demo


Input point propagated
Grayscale image
+ User point
Initial colorization

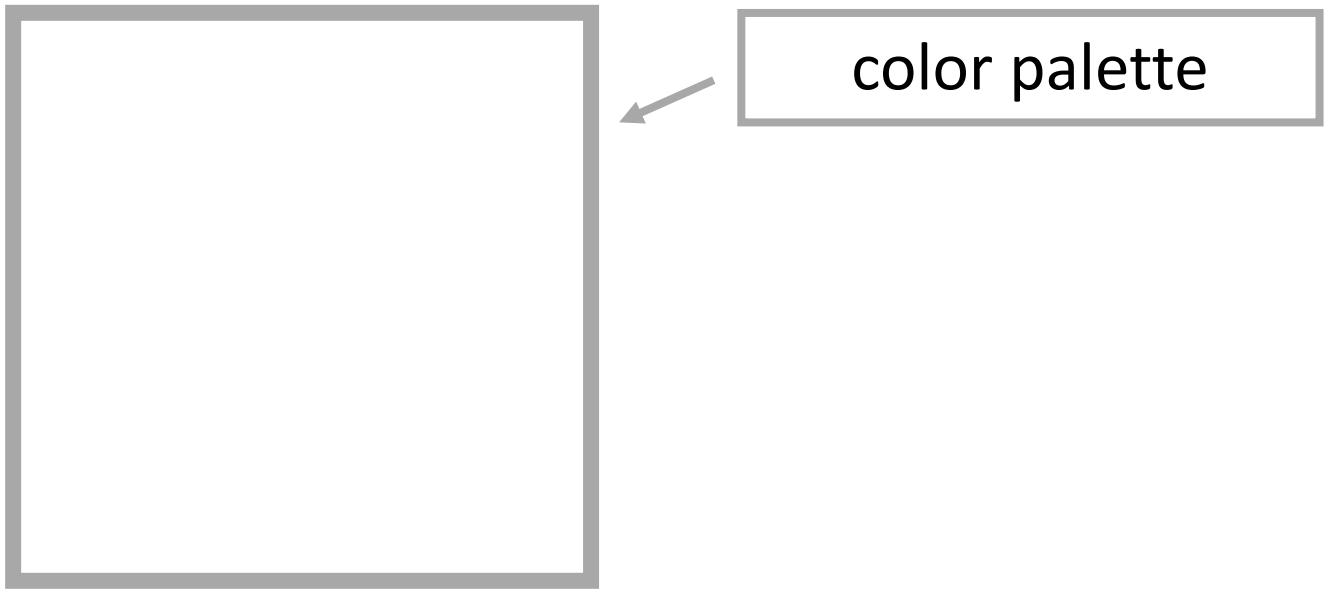
System Demo



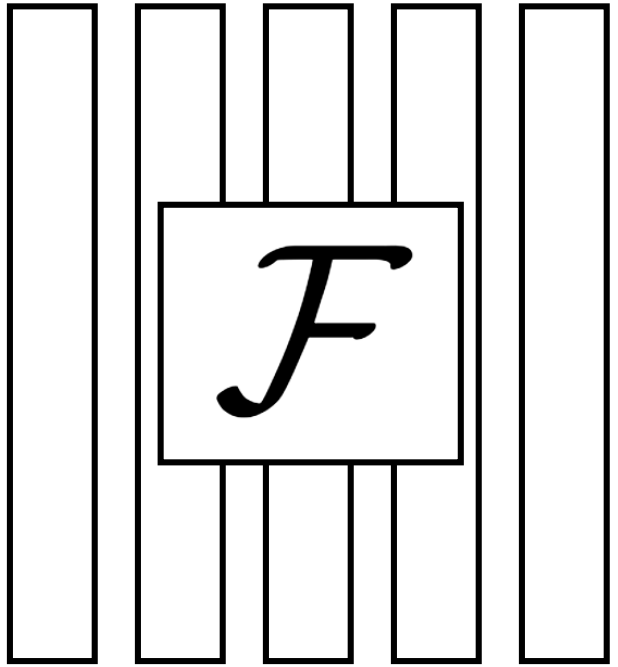
System Demo



System Demo



Colorization with user inputs



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

User points $\mathbf{U} \in \mathbb{R}^{H \times W \times 3}$

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

Training the System

Automatic

$$\arg\min_{\mathcal{F}} \ell(\mathcal{F}(\mathbf{X}), \mathbf{Y})$$

User-guided

$$\arg\min_{\mathcal{F}} \ell(\mathcal{F}(\mathbf{X}, \mathbf{U}), \mathbf{Y})$$

Randomly Simulated User Interactions

$$\arg\min_{\mathcal{F}} \ell(\mathcal{F}(\mathbf{X}, RandR$$

"Hint

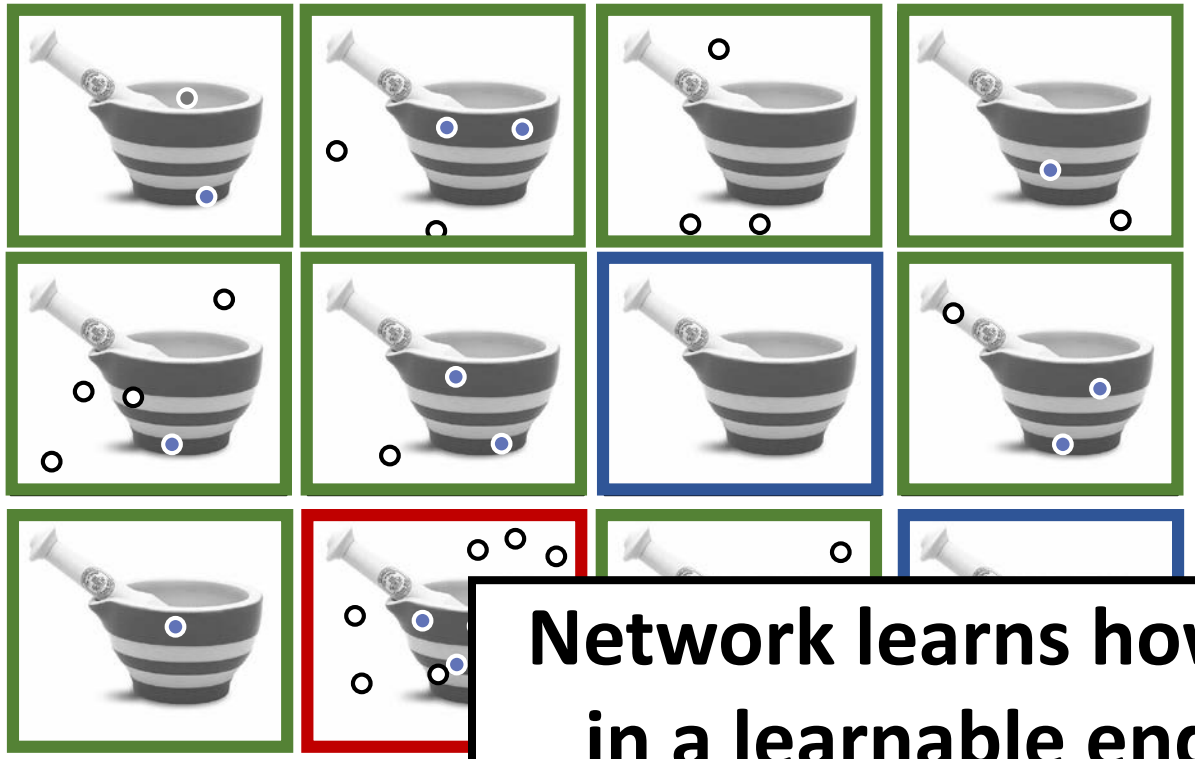
Training data

– User data not easily obtainable

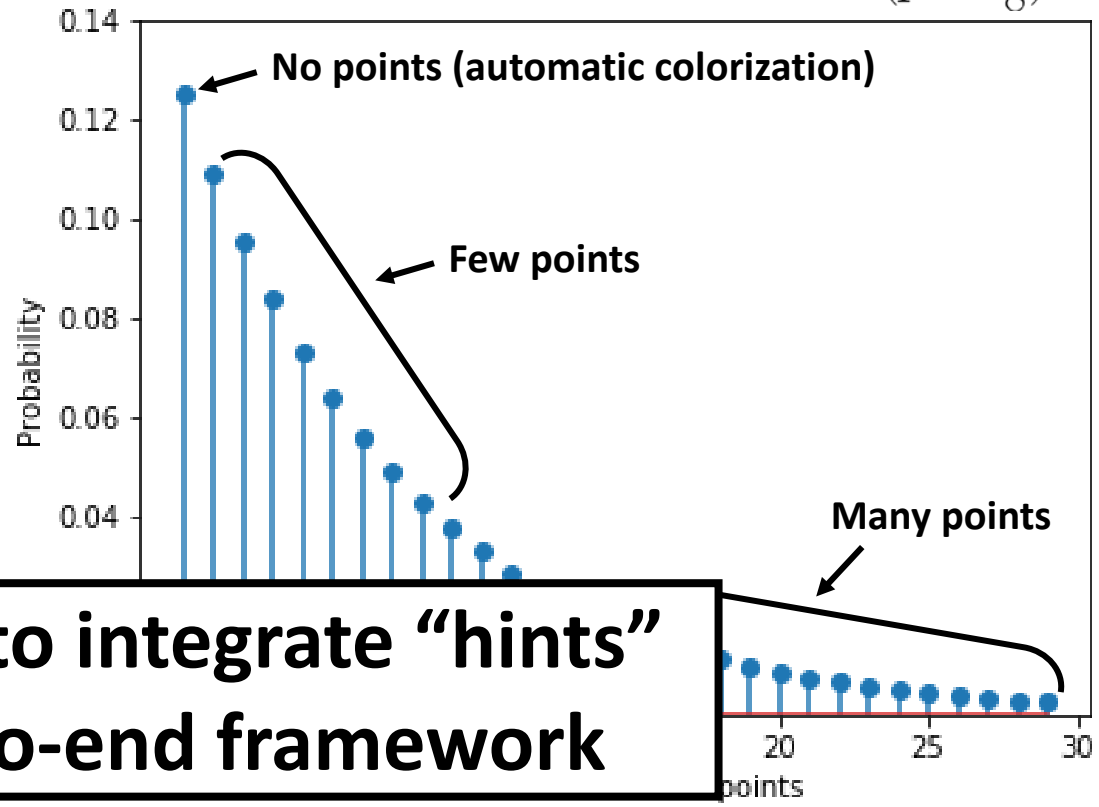
Randomly Simulated User Points

$$\arg\min_{\mathcal{F}} \ell(\mathcal{F}(\mathbf{X}, RandReveal(\mathbf{Y})), \mathbf{Y})$$

Example Simulated User Points



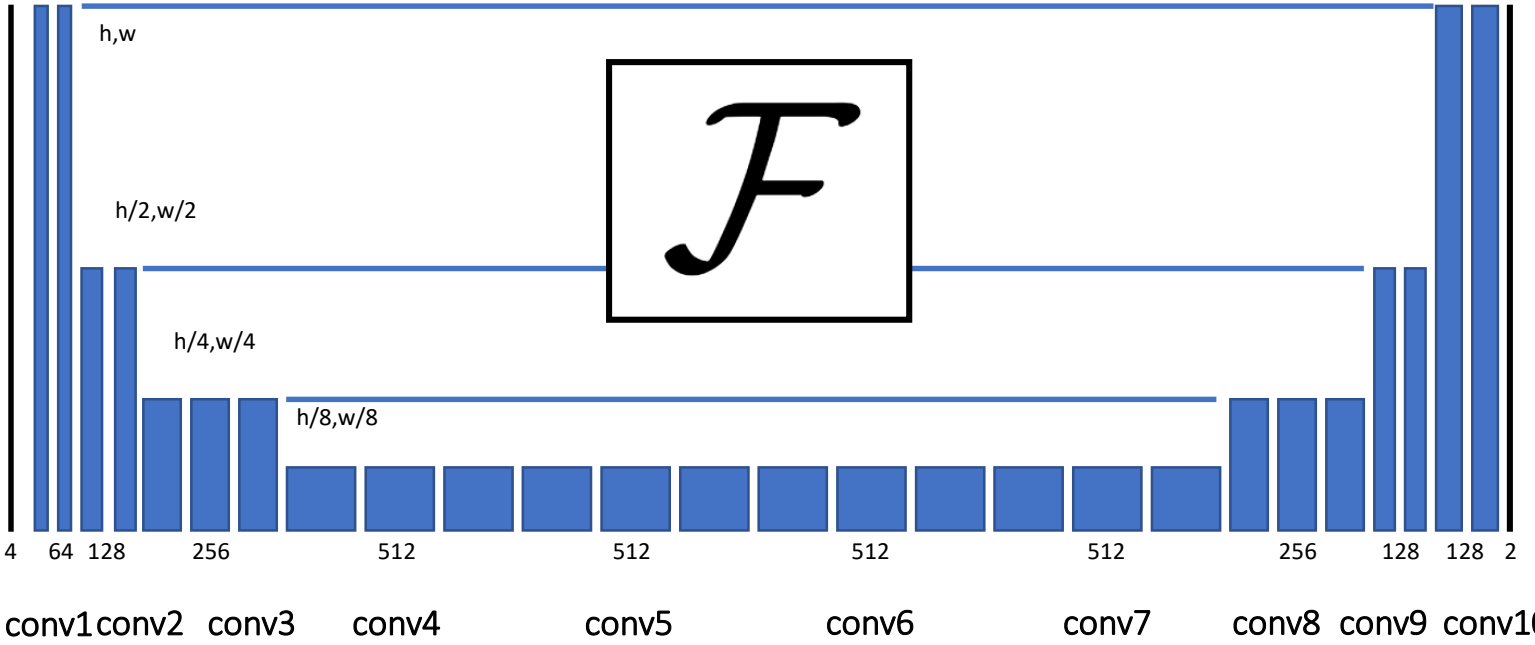
Number of revealed points $\sim G(p = \frac{1}{8})$



Network learns how to integrate "hints" in a learnable end-to-end framework

Network Architecture

Grayscale **X**
User **U**



Output **Y**

ImageNet database (1.3M train/10k val/10k test)
Finetune from Zhang et al. 2016

Ronneberger et al. MICCAI, 2015.
Isola et al. CVPR, 2017.

Colorizing Legacy Photos

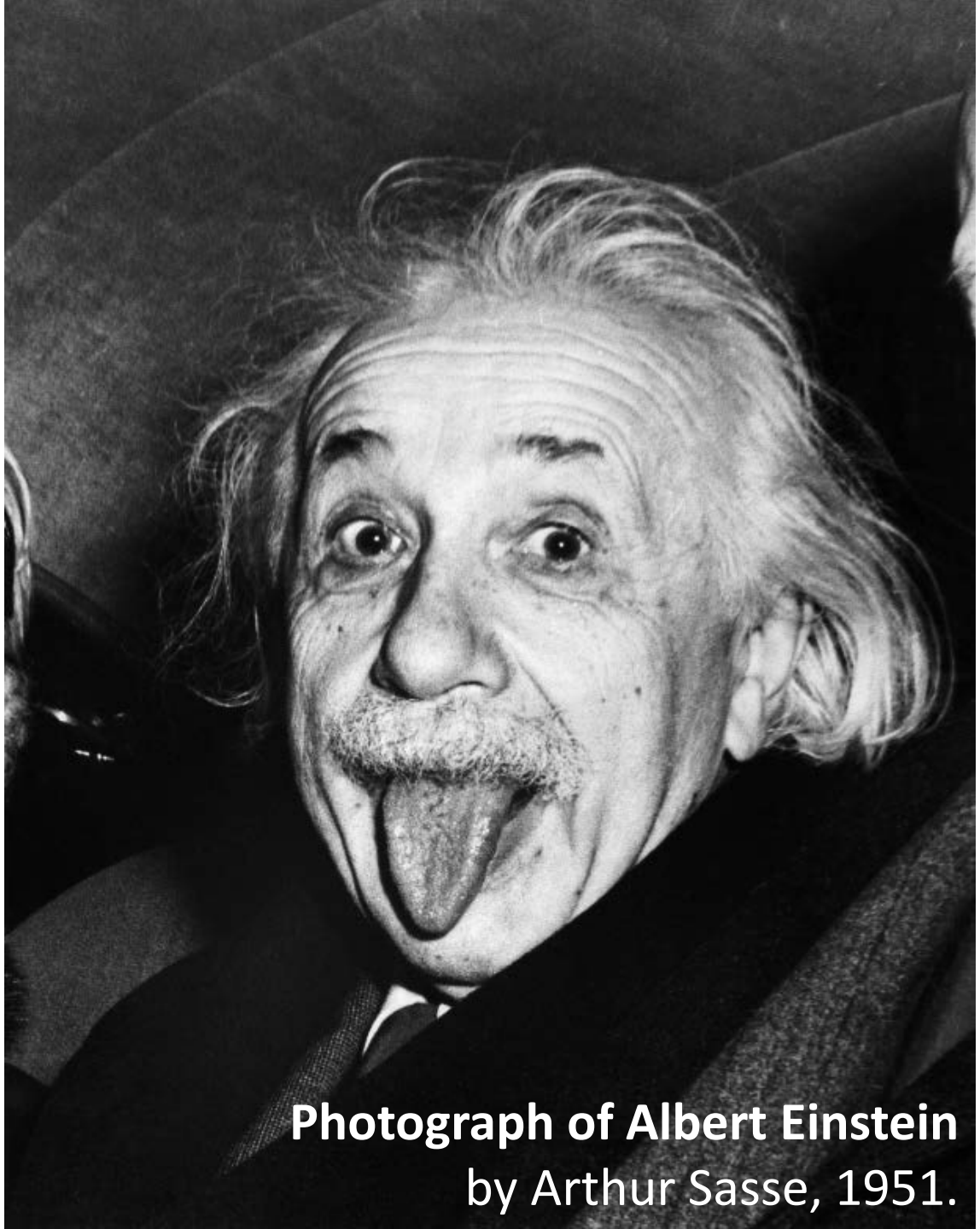
Amateur family photograph, 194

Amateur family photograph, 1949.

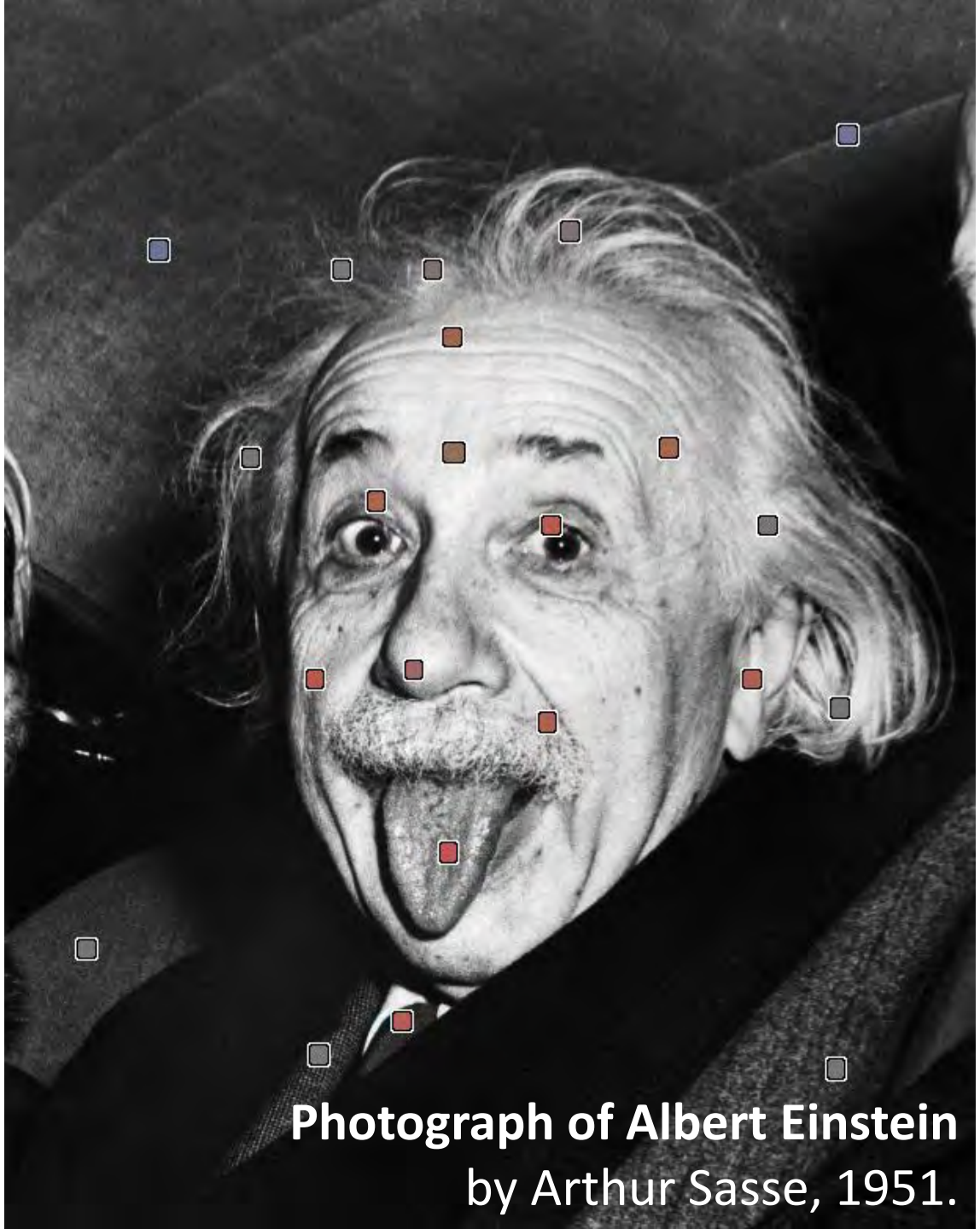
Amateur family photograph, 1949.

Amateur family photograph, 1949.

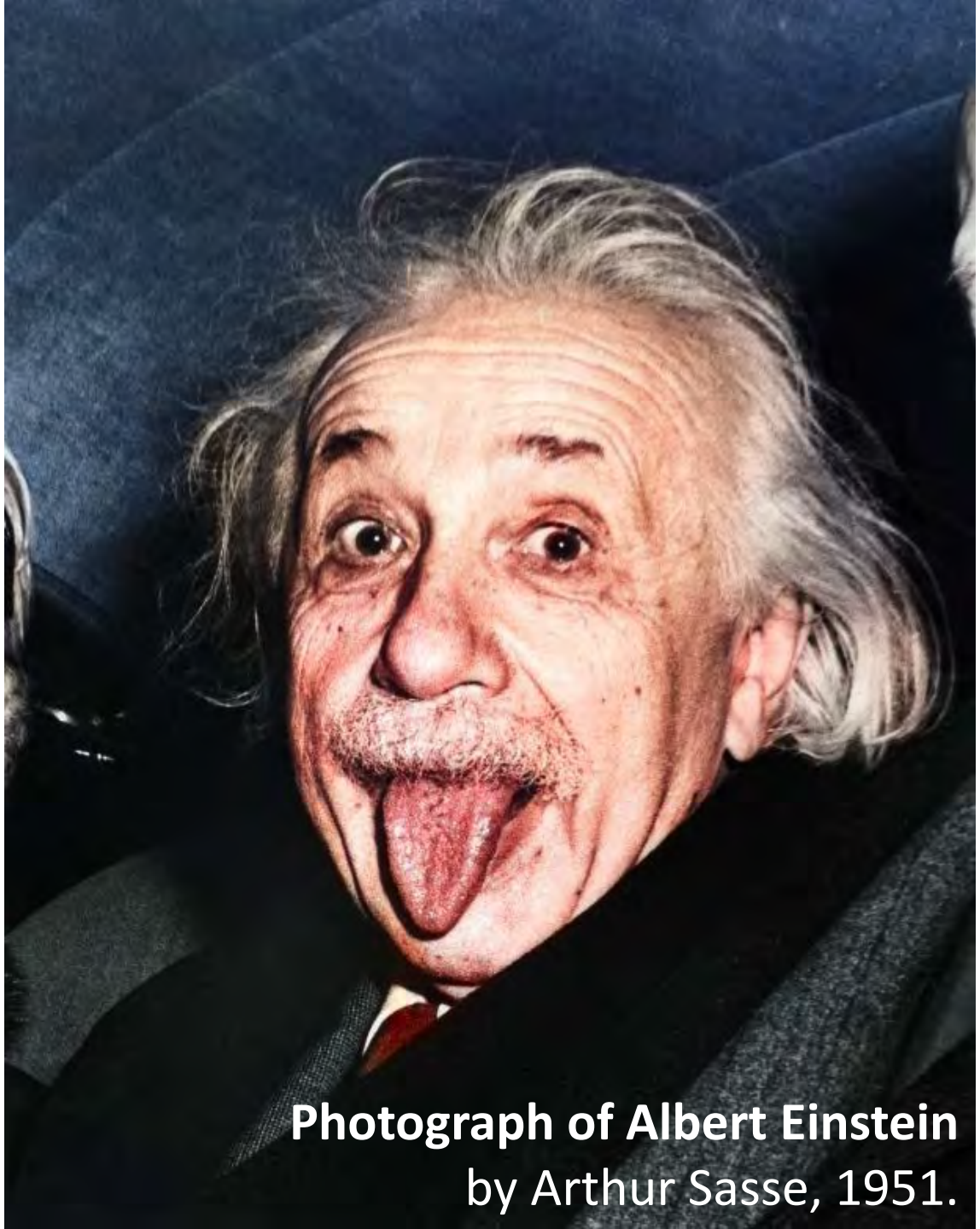
Photograph of Albert Einstein
by Arthur Sasse, 1951.



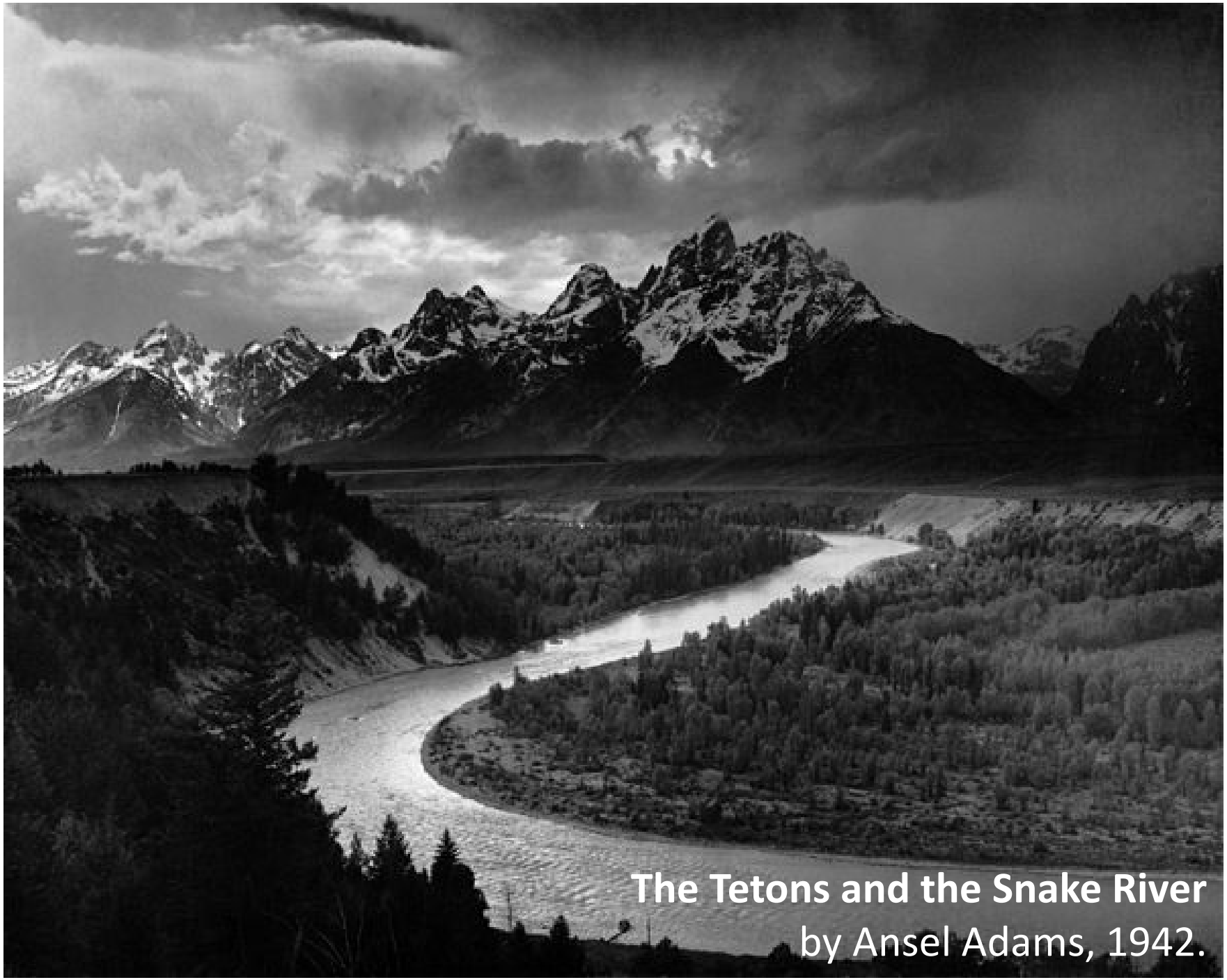
Photograph of Albert Einstein
by Arthur Sasse, 1951.



Photograph of Albert Einstein
by Arthur Sasse, 1951.



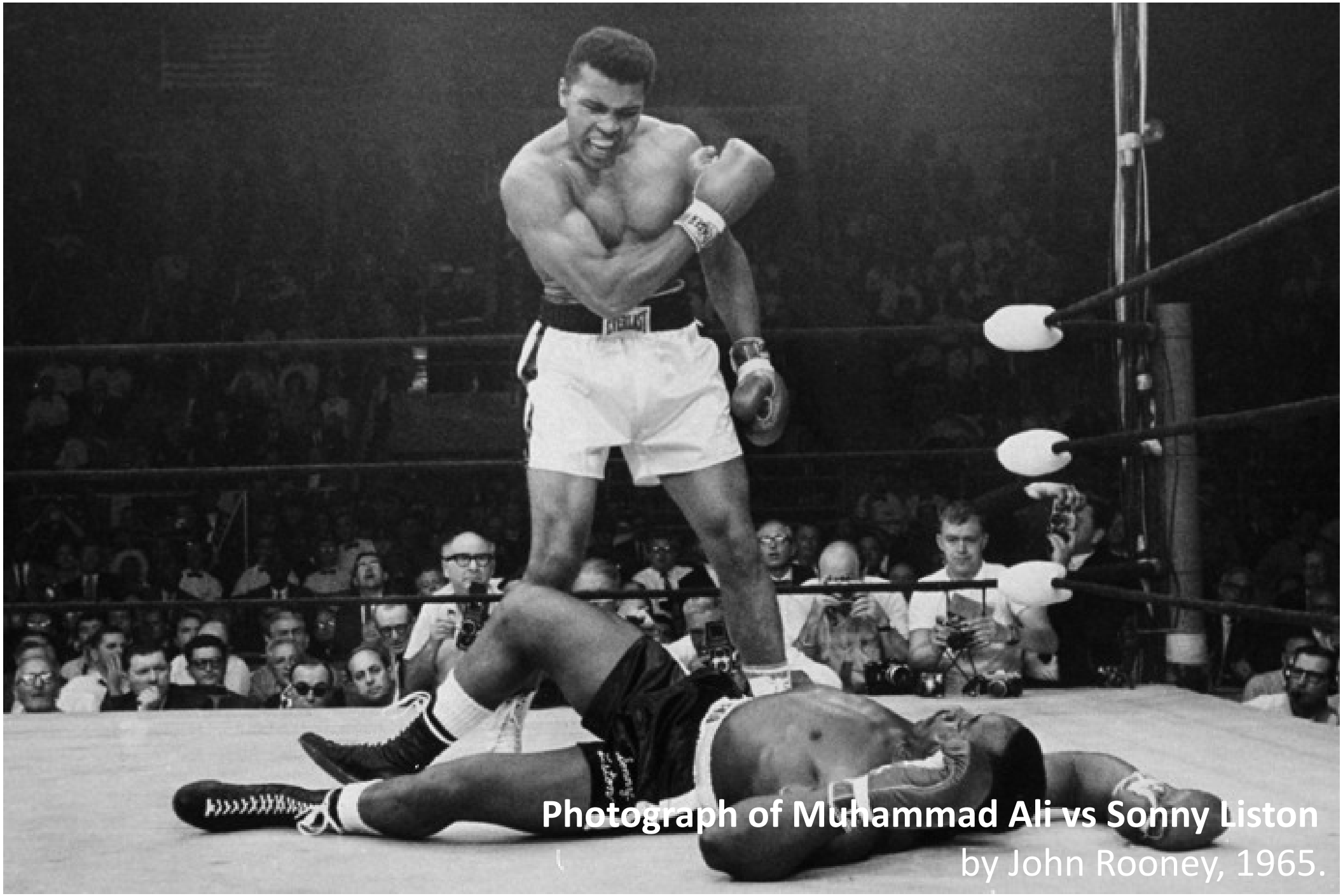
Photograph of Albert Einstein
by Arthur Sasse, 1951.



The Tetons and the Snake River
by Ansel Adams, 1942.

The Tetons and the Snake River
by Ansel Adams, 1942.

The Tetons and the Snake River
by Ansel Adams, 1942.

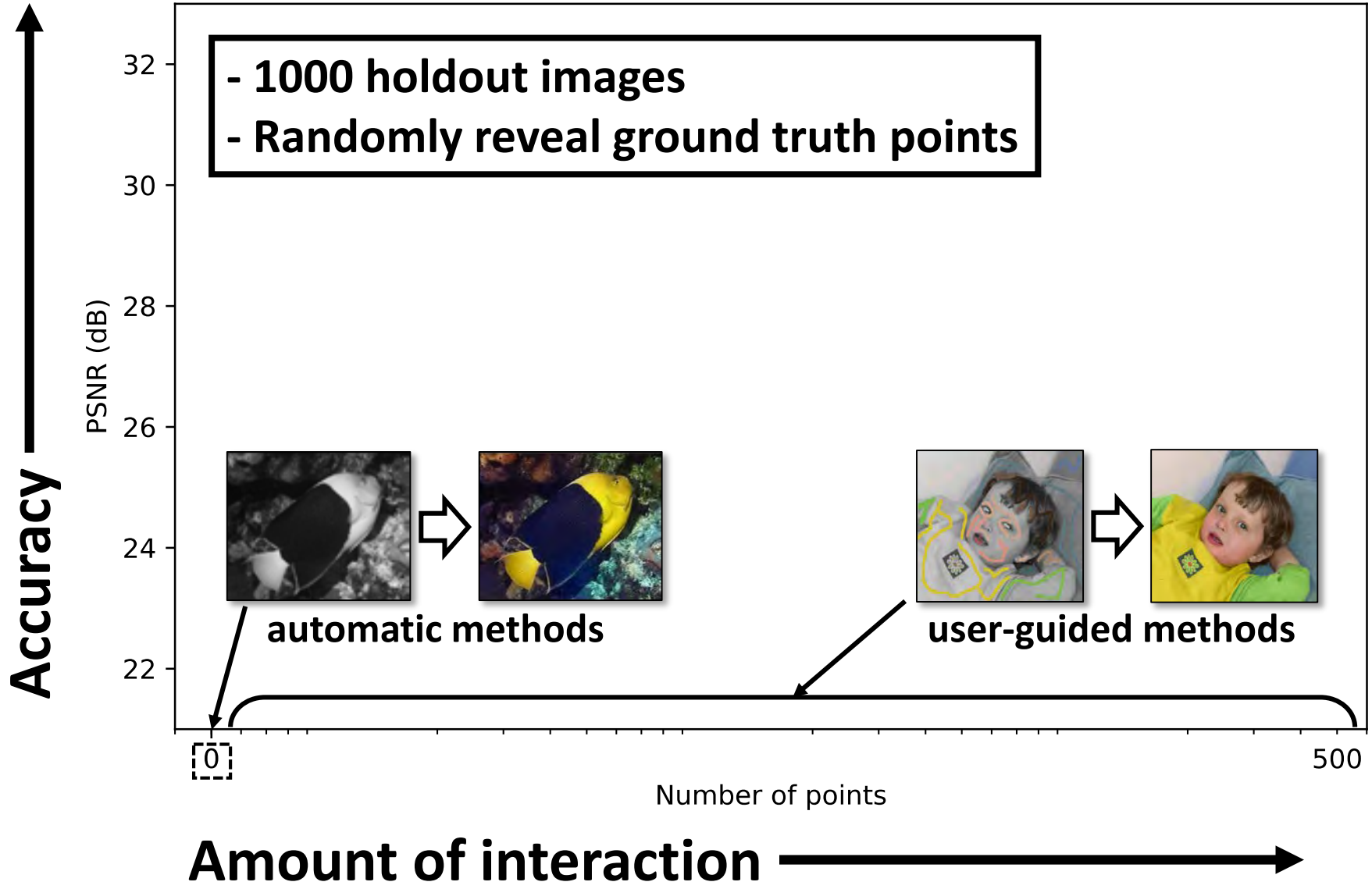


Photograph of Muhammad Ali vs Sonny Liston
by John Rooney, 1965.

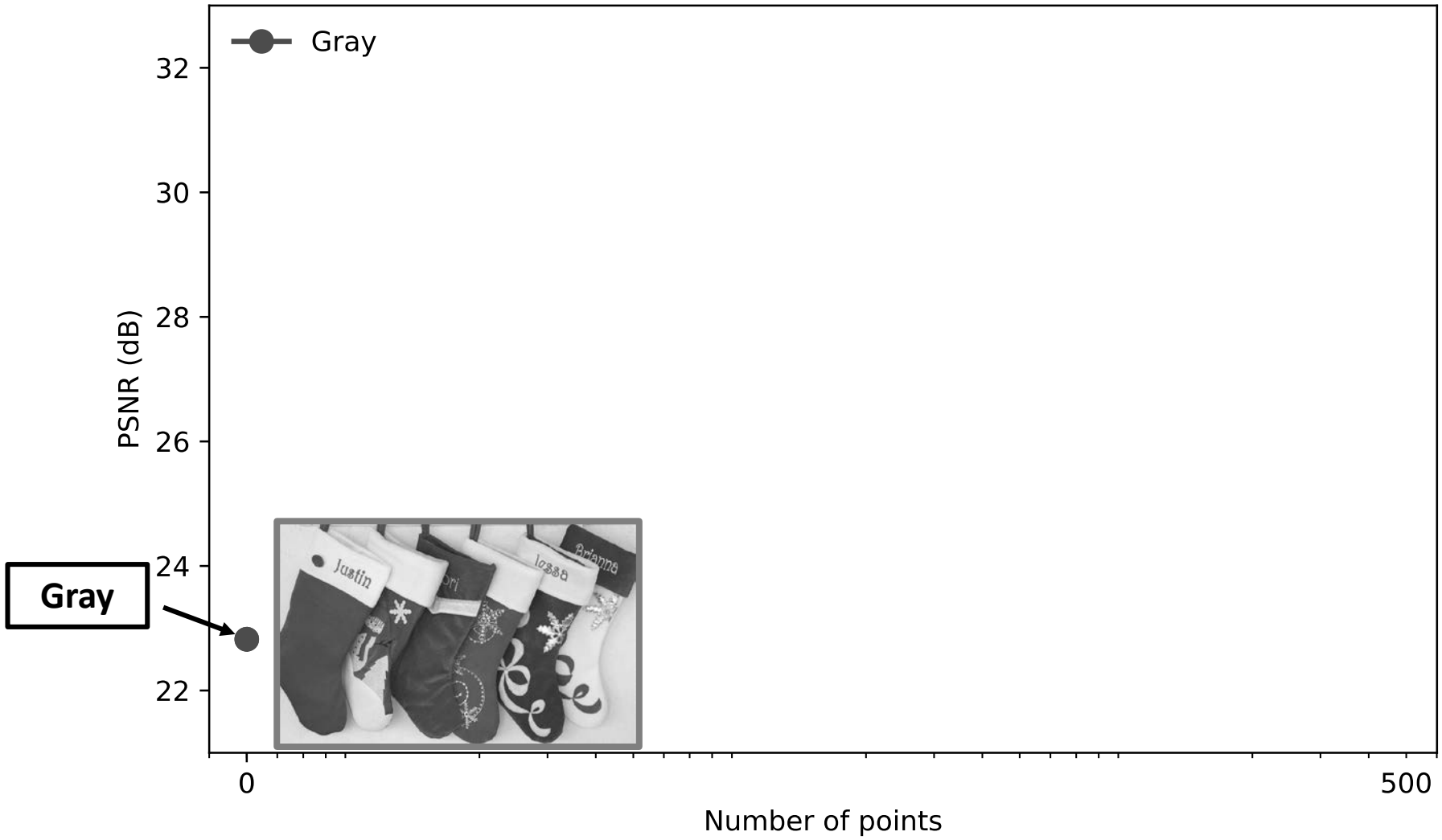
Photograph of Muhammad Ali vs Sonny Liston
by John Rooney, 1965.

Photograph of Muhammad Ali vs Sonny Liston
by John Rooney, 1965.

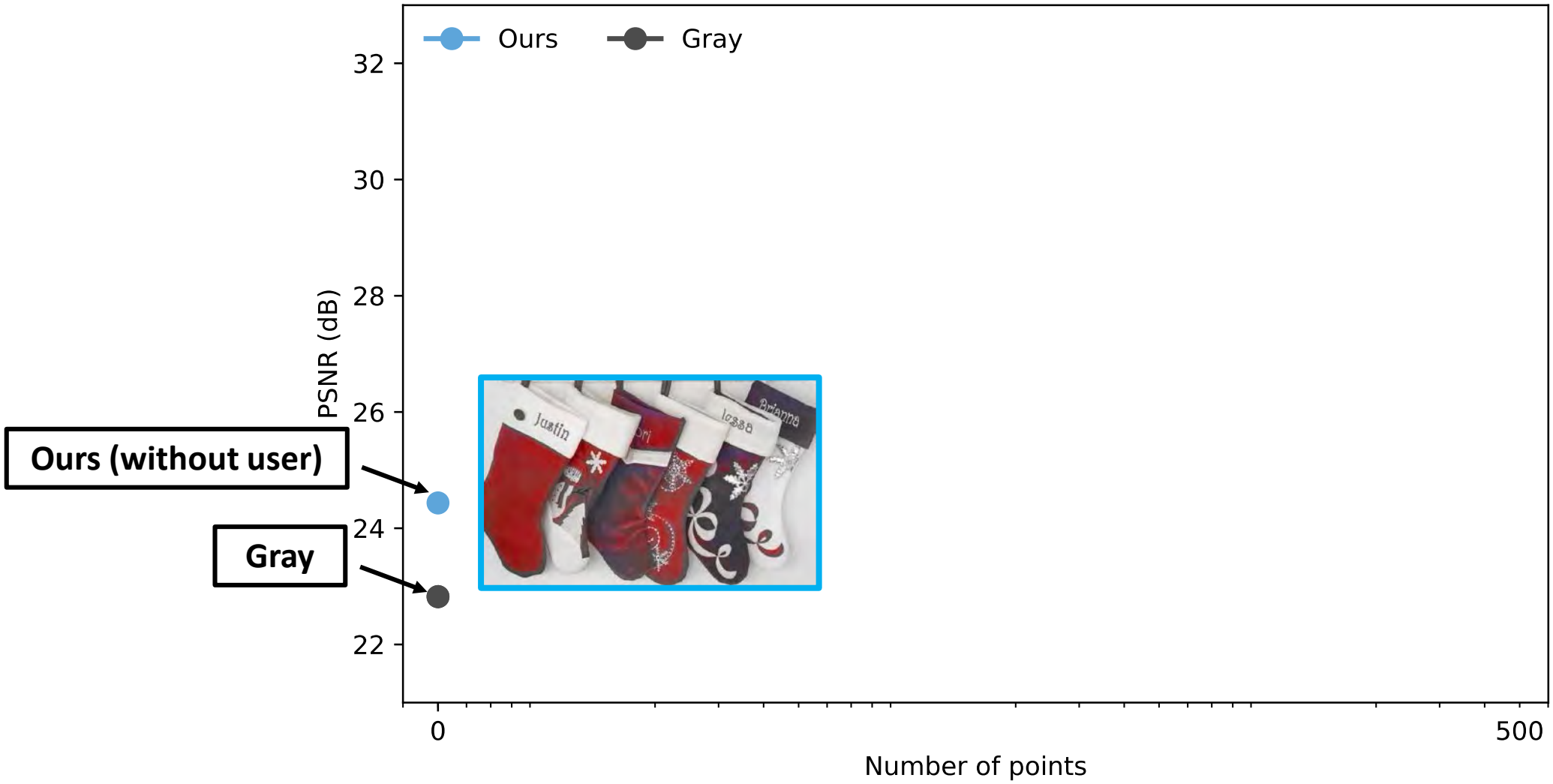
Evaluating the benefit of user hints



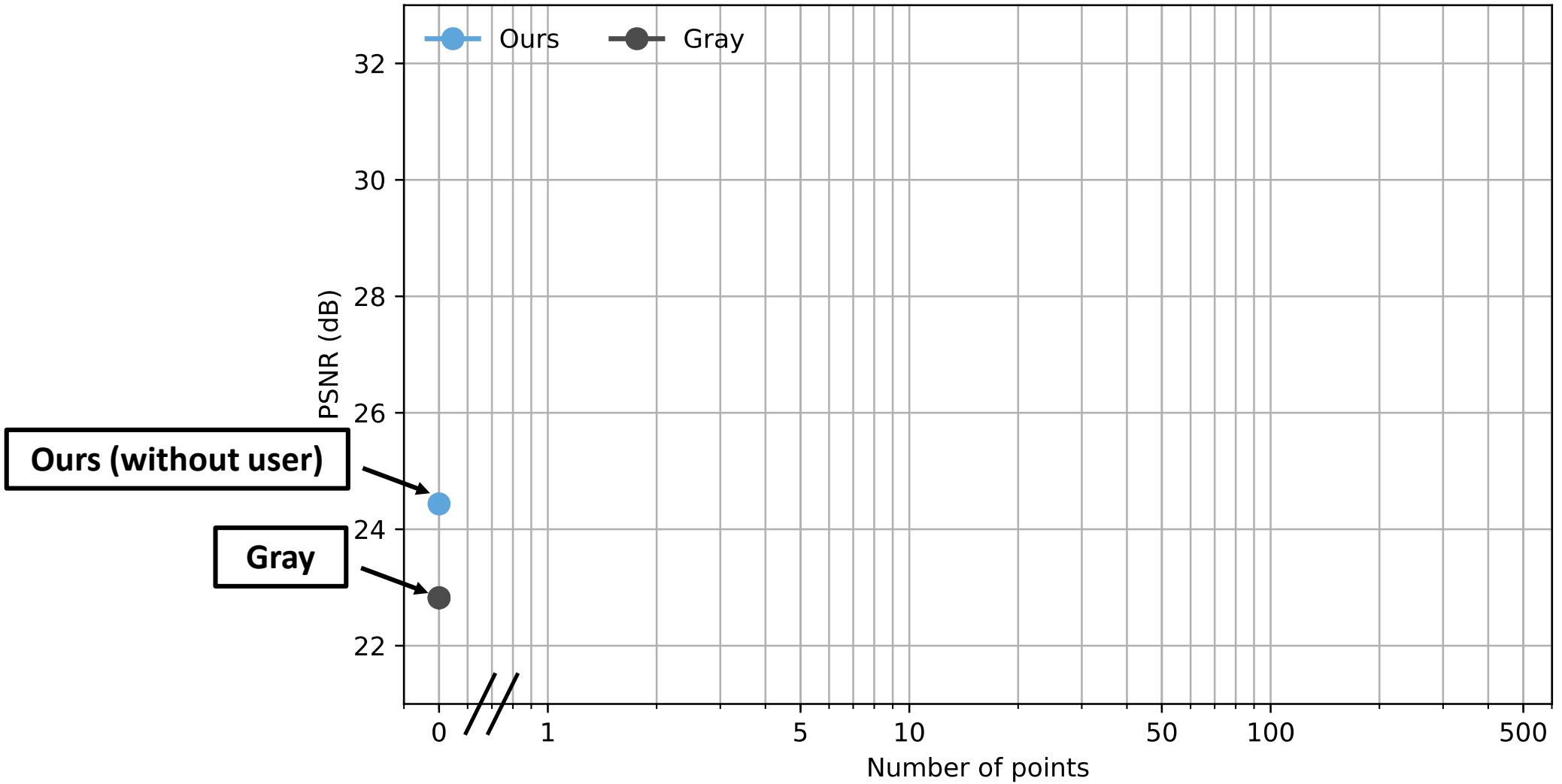
Evaluating the benefit of user hints



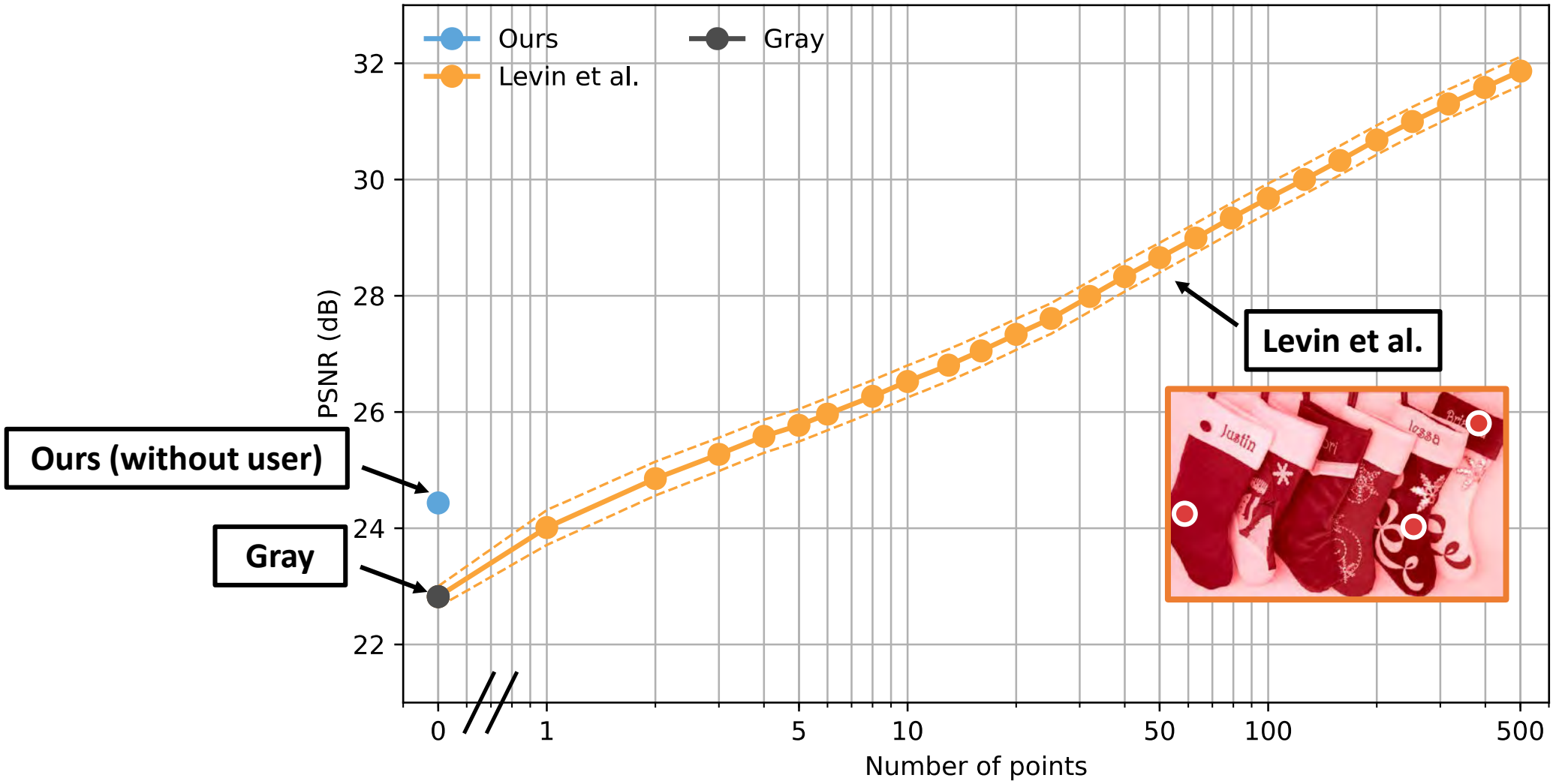
Evaluating the benefit of user hints



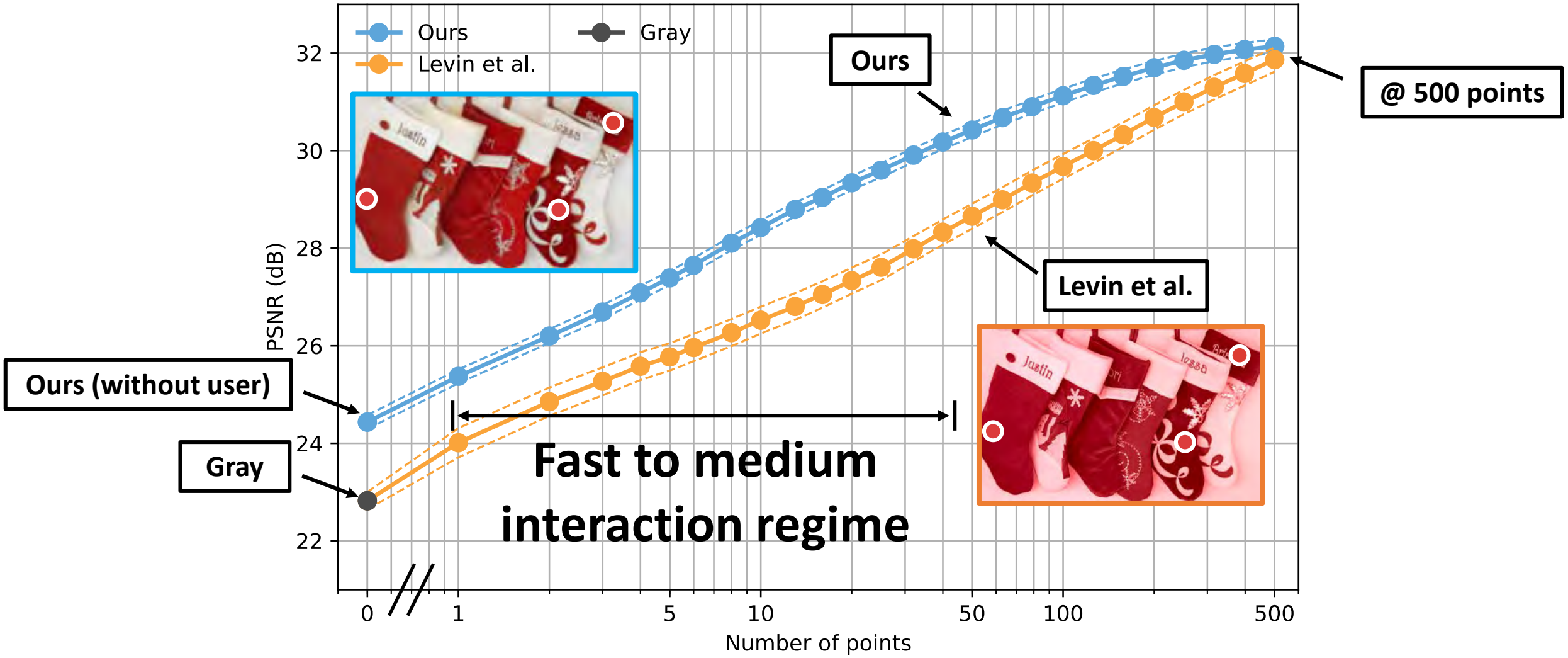
Evaluating the benefit of user hints



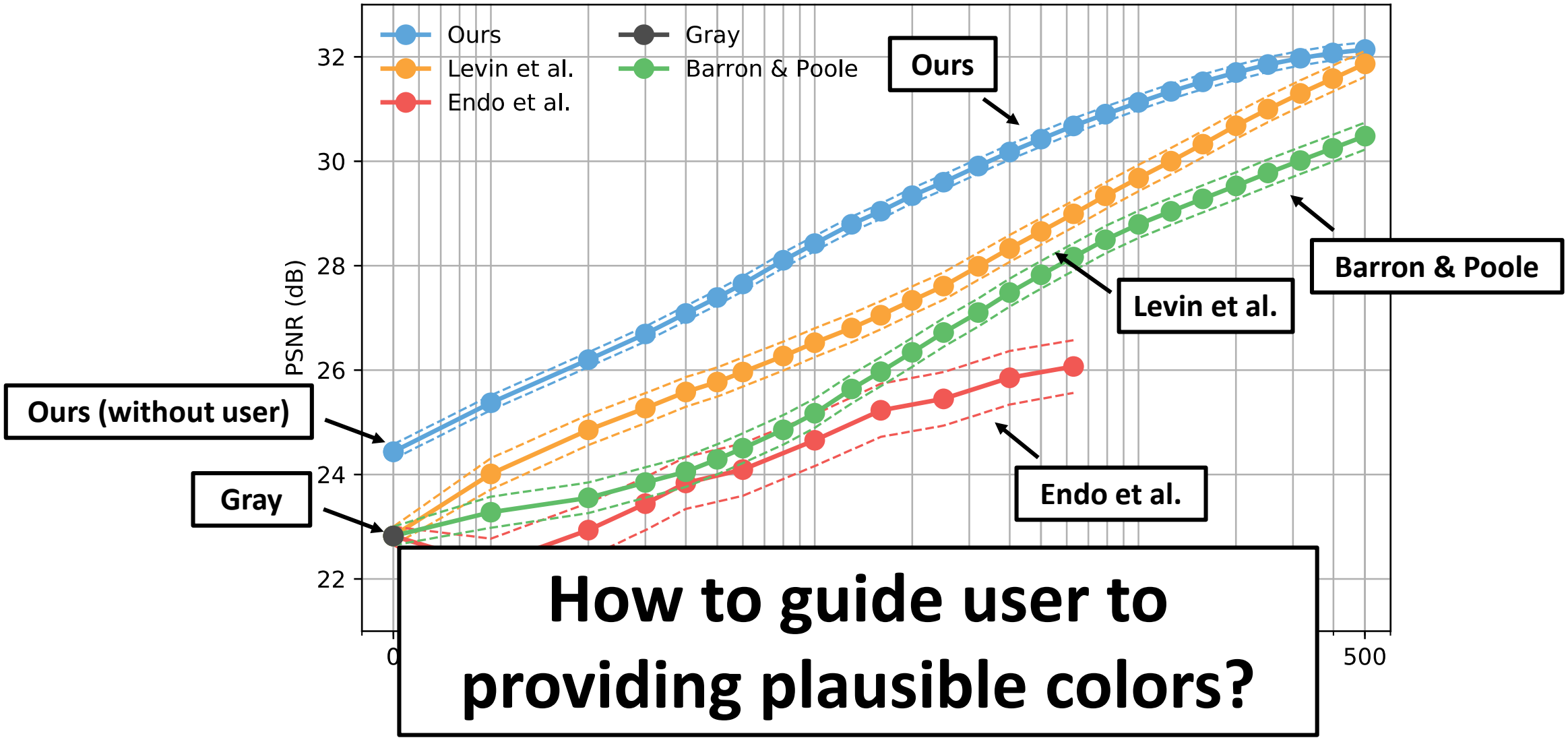
Evaluating the benefit of user hints



Evaluating the benefit of user hints

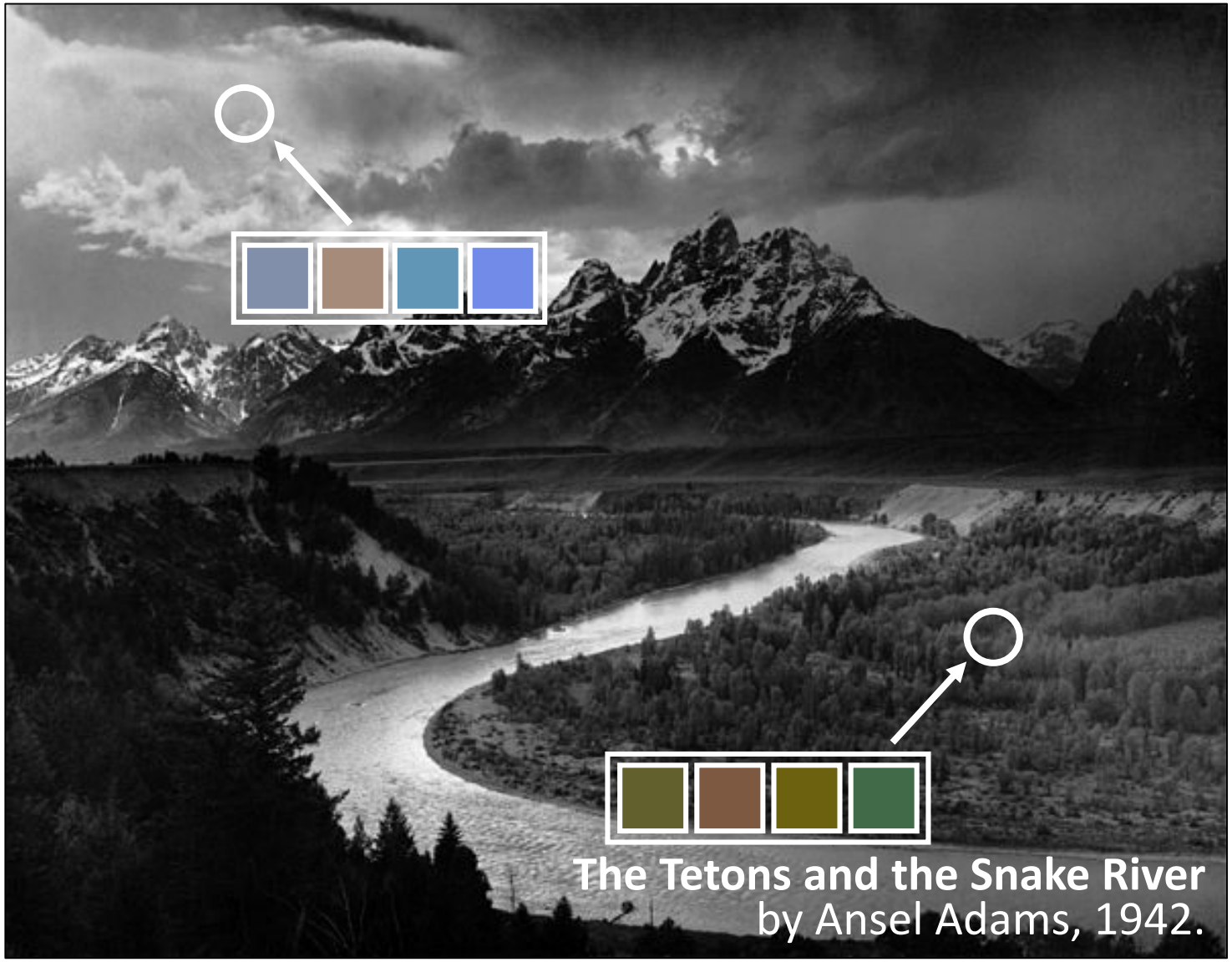


Evaluating the benefit of user hints



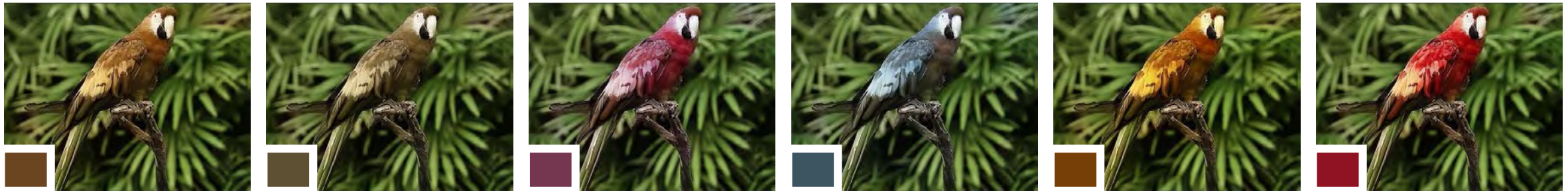
How to guide user to providing plausible colors?

Data-Driven Recommendations

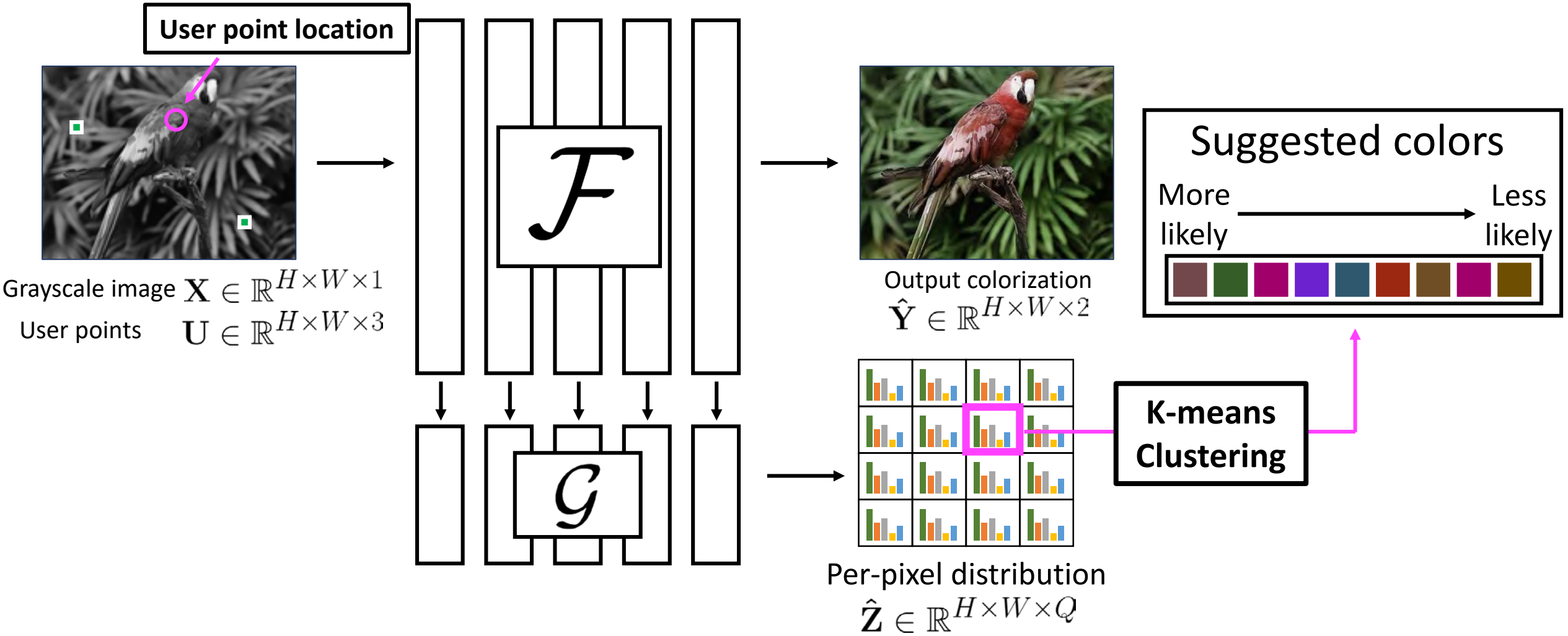


User point

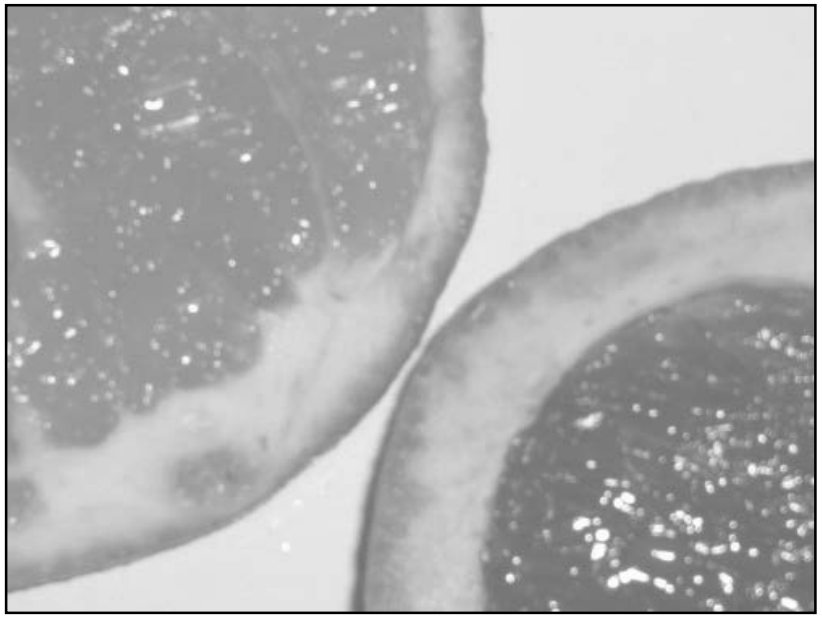
Suggested colors



Data-Driven Recommendations



Example User Result

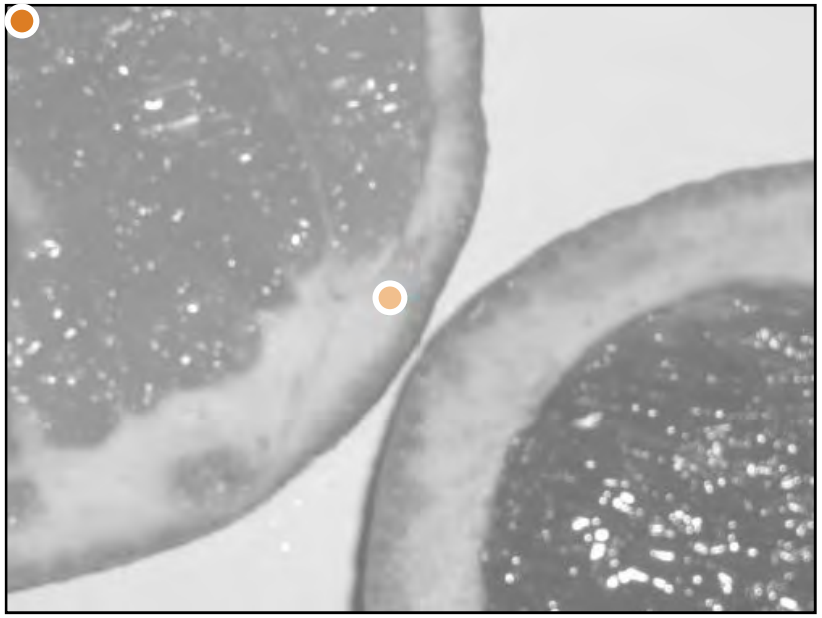


Grayscale image

Automatic result

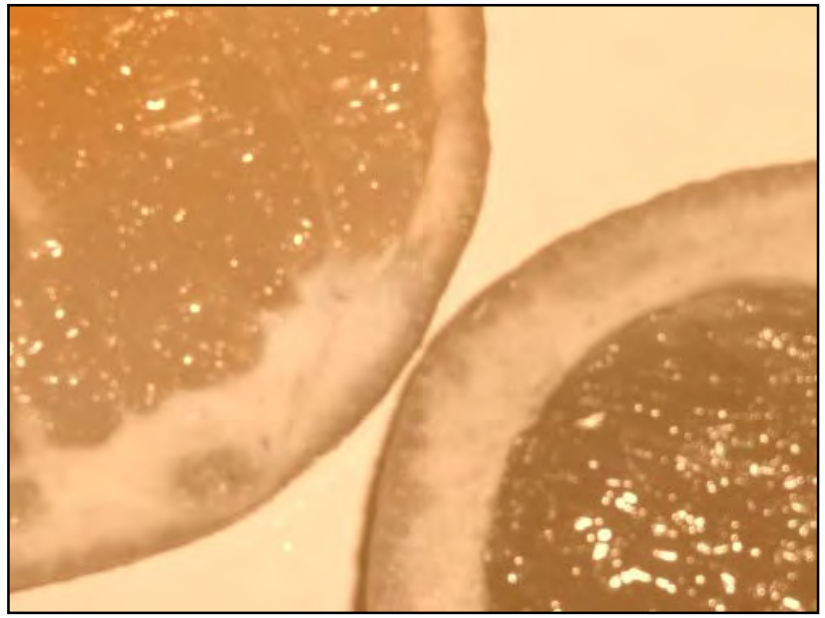
From novice user with < 1 minute of use

Example User Result



Grayscale + User

Interactive result



Levin et al.

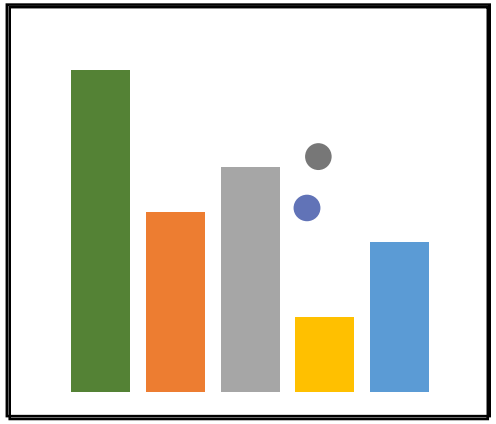
From novice user with < 1 minute of use

Unusual colorizations

Mark Ruffalo © wikipedia

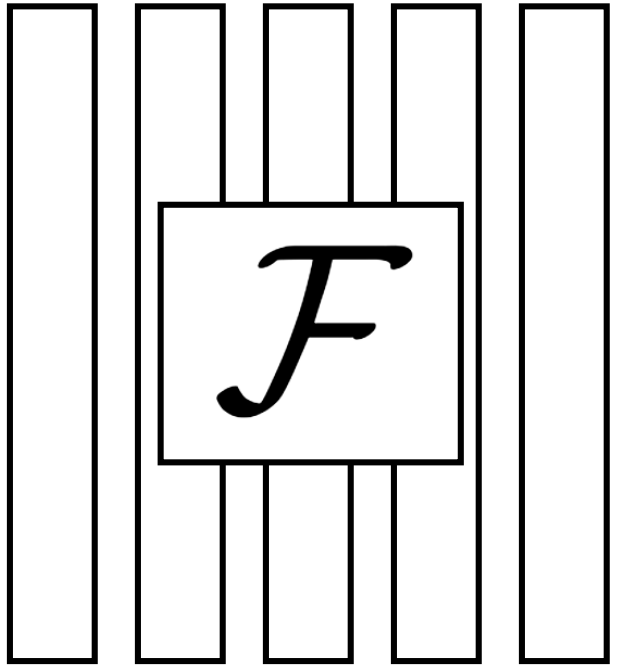
Colorization with Hints

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



User-given histogram

Global histogram $\mathbf{Z} \in \Delta^{1 \times 1 \times Q}$



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

Grayscale image

Resulting colorizations given global histograms

See also Deep Photo Style Transfer [Luan et al. CVPR 17]

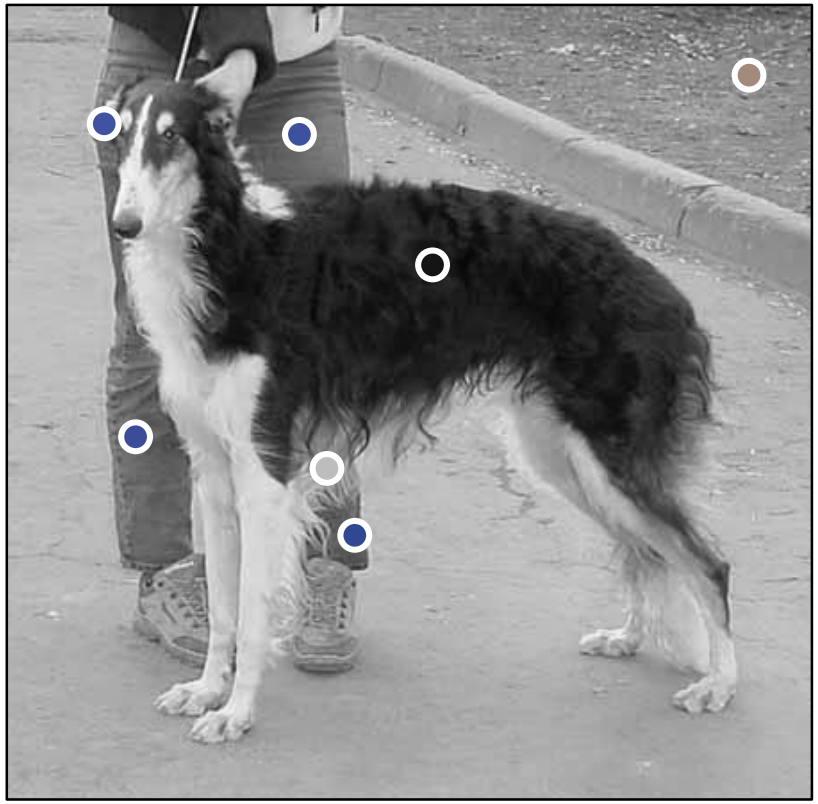
Grayscale image

Resulting colorizations given global histograms

See also Deep Photo Style Transfer [Luan et al. CVPR 17]

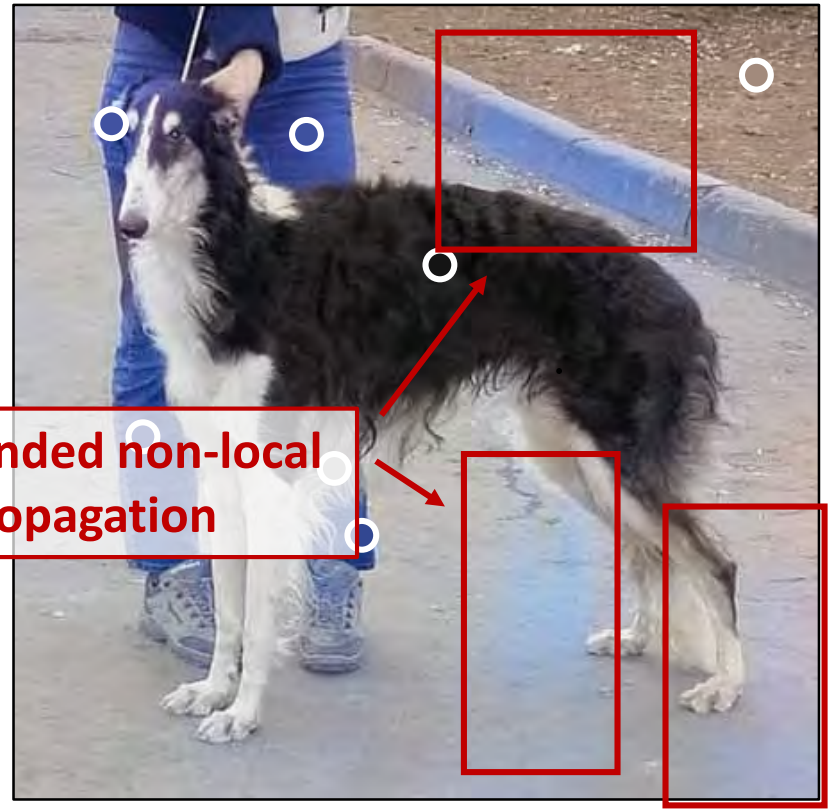
Failure Cases

Failure Case: Non-local effects



Grayscale

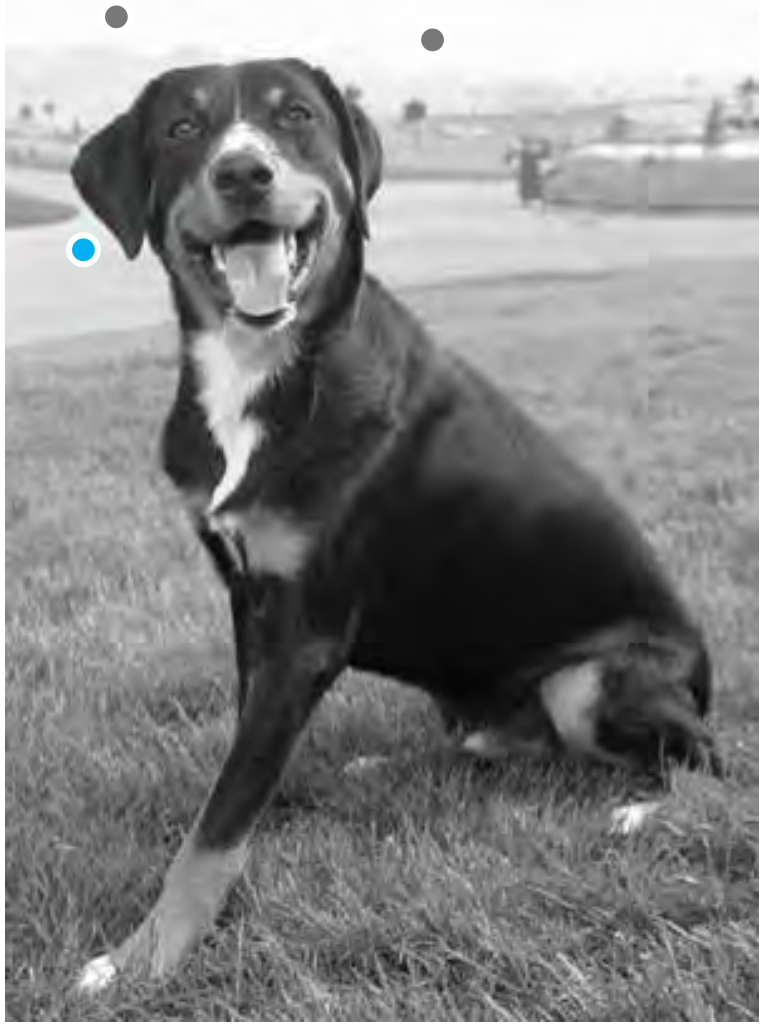
Automatic result



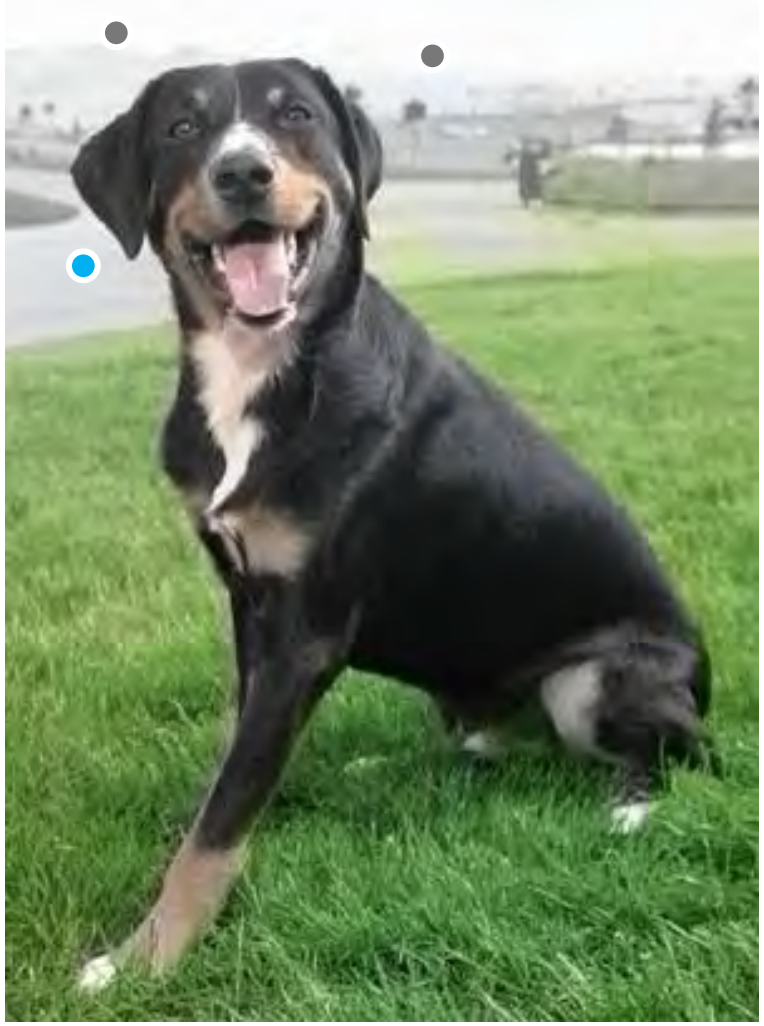
Interactive Result

From novice user with < 1 minute of use

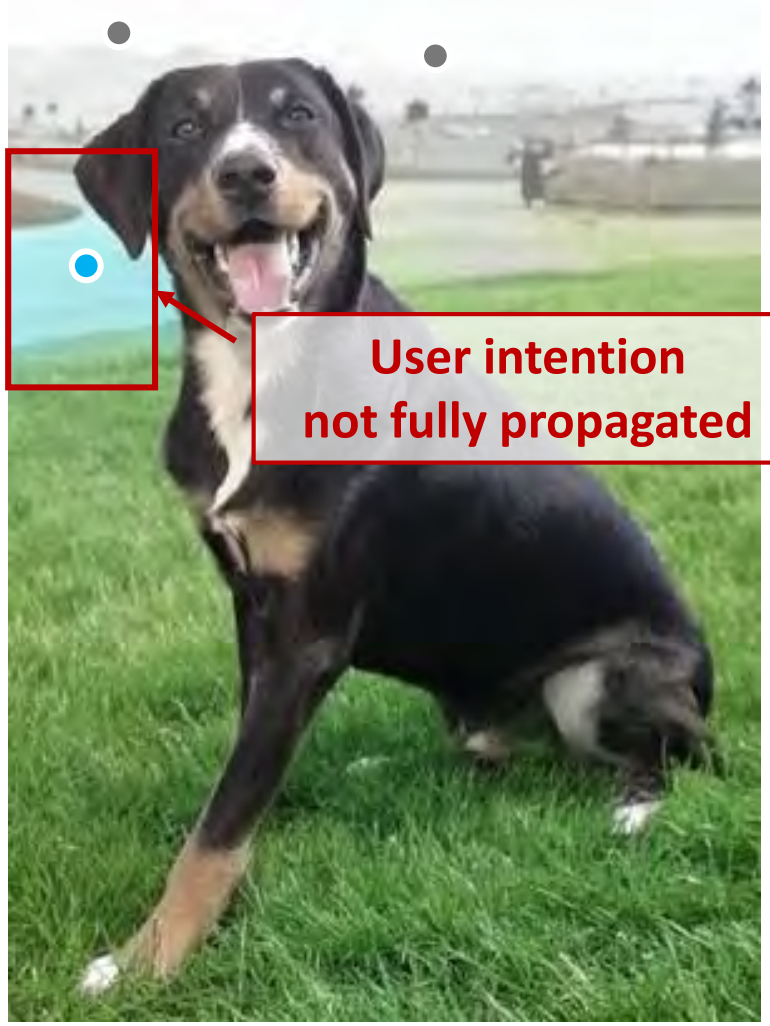
Failure Case: Incomplete Propagation



Grayscale



Automatic result



Interactive result

From novice user with < 1 minute of use

Failure Case: incompatible color style




Deep Learning for User Interaction

Proposed

Grayscale image + user hints

Colorized Image

Deep Interactive Object Selection

Xu *et al.* CVPR 2016.

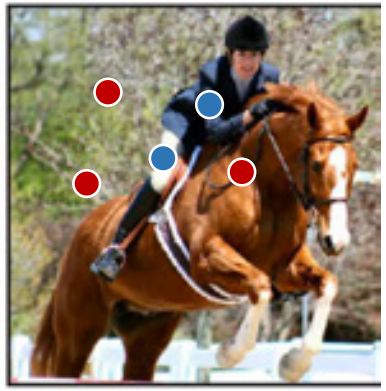


Image + pos/neg seeds

Figure/ground segmentation

Pix2pix, Scribbler

Isola *et al.*, Sangkloy *et al.* CVPR 2017.

Sketch

Image

Code: https://github.com/junyanz/interactive-deep-colorization

Website: http://richzhang.github.io/ideepcolor

More Information

- R. Zhang, P. Isola, A. A. Efros. Colorful Image Colorization. In *ECCV*, 2016.
  - Website: https://richzhang.github.io/colorization
  - Demo: https://demos.algorithmia.com/colorize-photos/
  - Code: https://www.github.com/richzhang/colorization
- R. Zhang, P. Isola, A. A. Efros. Split-Brain Autoencoders: Cross-Channel Prediction by Cross-Channel Prediction. In *CVPR*, 2017.
- R. Zhang*, J. Zhu*, P. Isola, X. Geng, A. S. Lin, T. Yu, A. A. Efros. Real-Time User-Guided Image Colorization with Learned Deep Priors. In *SIGGRAPH*, 2017.
  - Website: https://richzhang.github.io/ideepcolor
  - Code: https://github.com/junyanz/interactive-deep-colorization

Lukas Graham – 7 Years

For the full paper, code, and live demo:
Automatic: richzhang.github.io/colorization
Interactive: richzhang.github.io/ideepcolor