Learning Representations for Automatic Colorization

Gustav Larsson, Michael Maire, Gregory Shakhnarovich

ECCV 2016

Background

- Requires high level understanding
  - Object identification
  - Segmentation
- Proxy for visual understanding

Contributions

- New ImageNet colorization benchmark
- Best results against all metrics
- Learning segmentation from colorization

Prior Methods

- Scribble [Levin et al. SIGGRAPH 2004]
    - Interactive
    - Color consistency assumption

- Transfer [Charpiat et al. ECCV 2008]
    - Reference image repo

- Fully Automatic [Deshpande et al. ICCY 2015]
    - Works on few scene classes
    - Best results with known scene type

Overview

- Self supervised on grayscale-converted color images
- Train CNN to predict hue/chroma given lightness
  - Histogram of potential colors
- Optional manually specified color biases

Author’s Approach

- Baked-in semantic information
    - CNN trained on ImageNet
- Localize and recognize objects
- No hand-crafted features

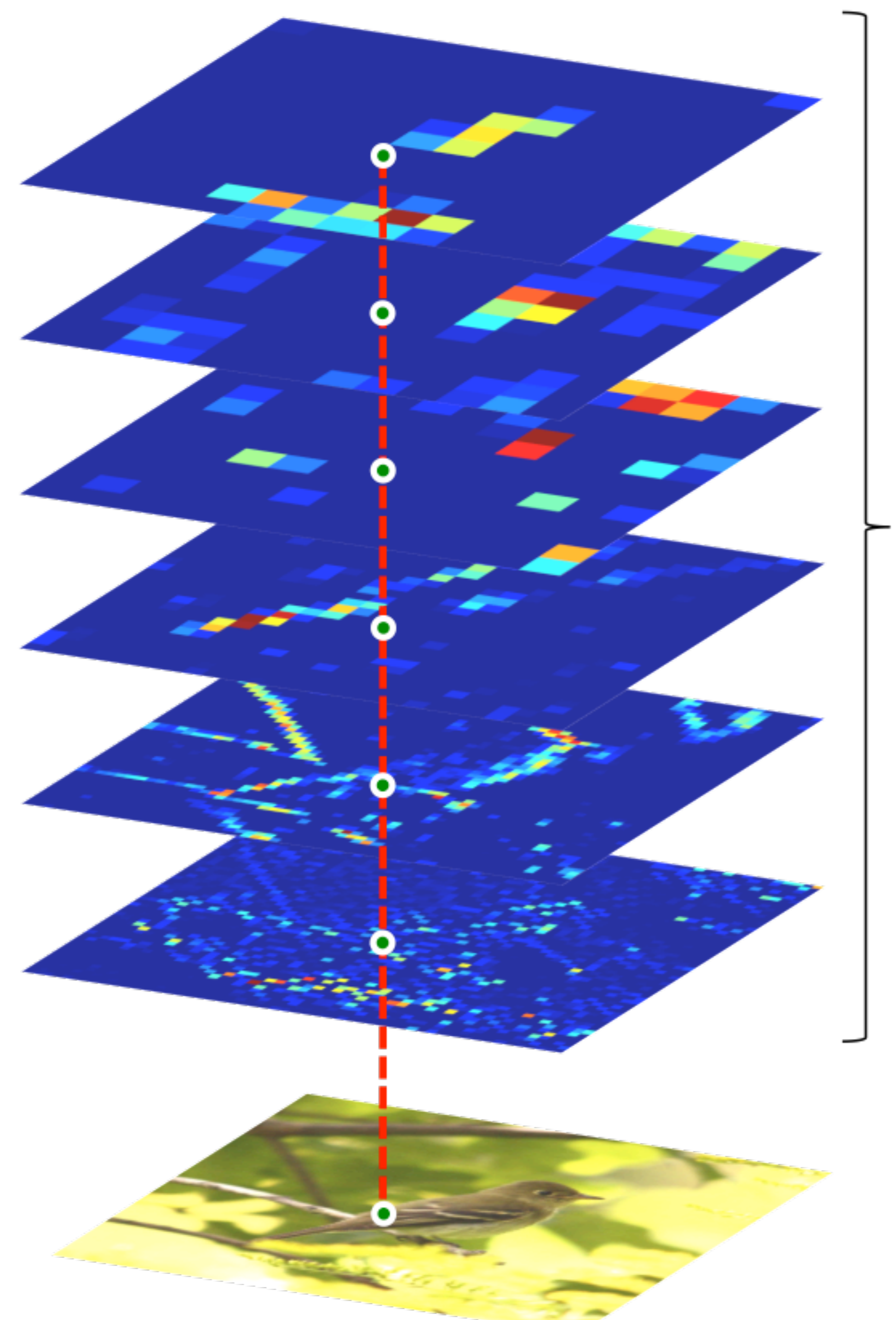
Author’s Approach

- Predict color histogram
  - Not single color
  - e.g. shirts can be many different colors

Fig. 9: **Sampling colorizations.** *Left:* Image & 3 samples; *Right:* Uncertainty map.

Author’s Approach

- Hypercolumns
  - 16 vs 2
- Pixel output based on local patch
- Option to bias color towards reference image



Color Parameterization

- Images converted to grayscale $$L = \frac{R+G+B}{3}$$
- RGB overdetermined
    - Intensity is a given
- HSV/HSL
- L*a*b (Lab)

Chrominance

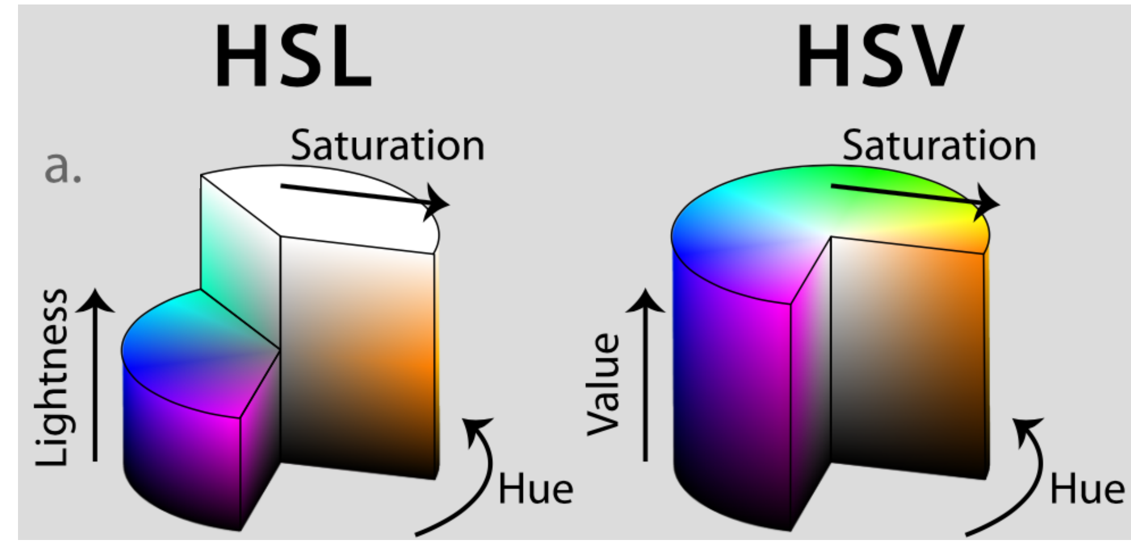
Luminance

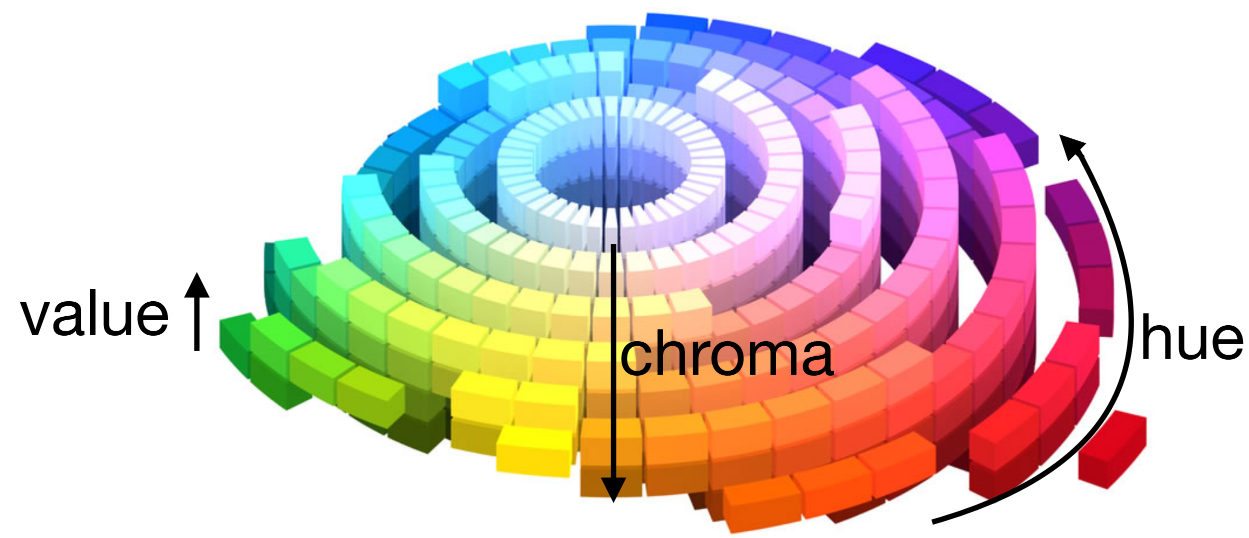
Chrominance

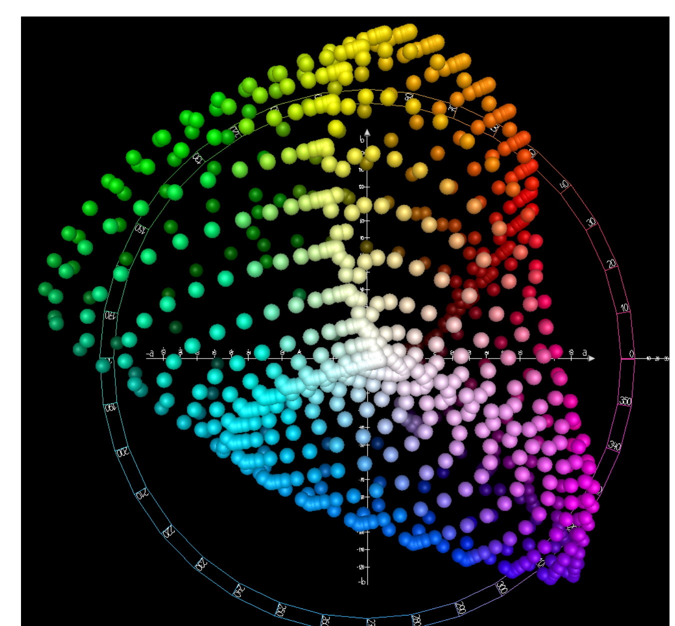
Both

Color Spaces

- HSL/HSV
- HVC
- Lab
  - Perceptually linear
  - L -> intensity
  - a -> green/red
  - b -> blue/yellow

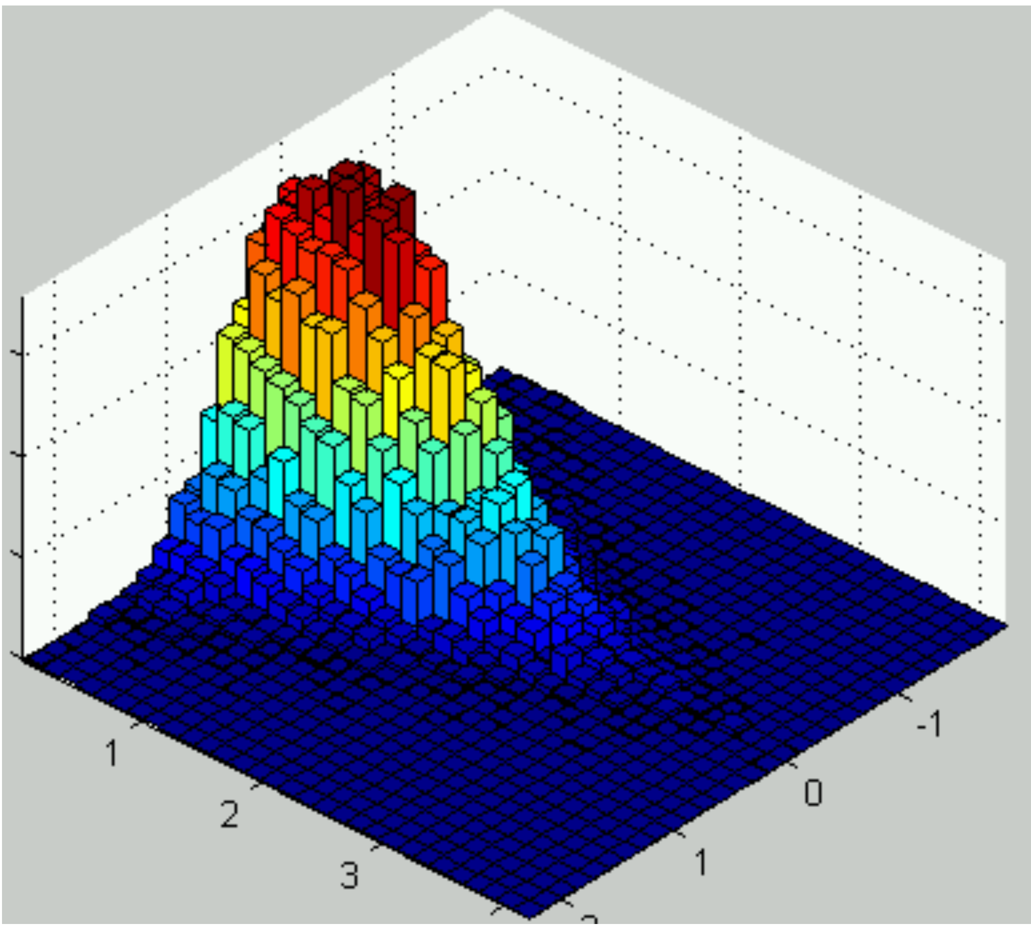






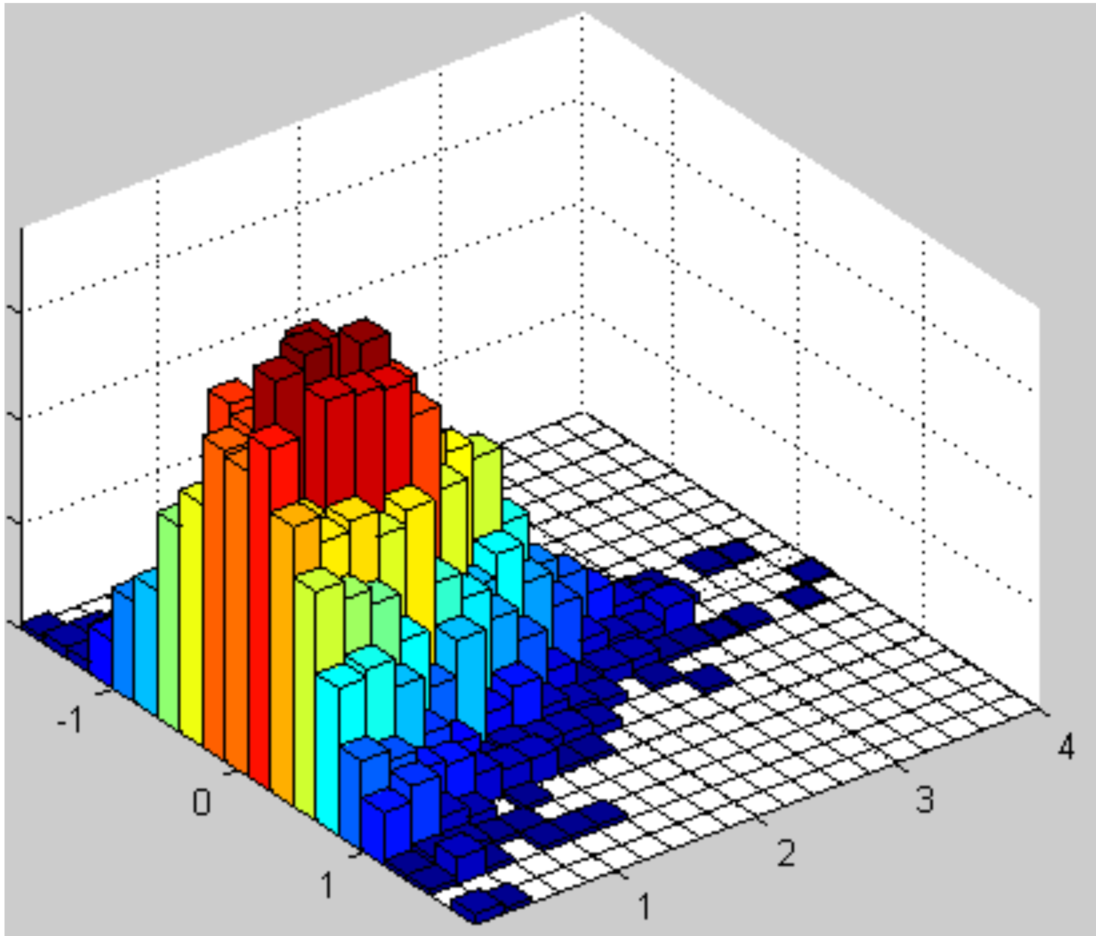
Loss Function

- Single color prediction
  - $L_{reg}(x, y) = ||f(x) - y||^2$
- Histogram prediction
  - $L_{hist}(x, y) = D_{KL}(y \| f(x))$
  - KL-Divergence



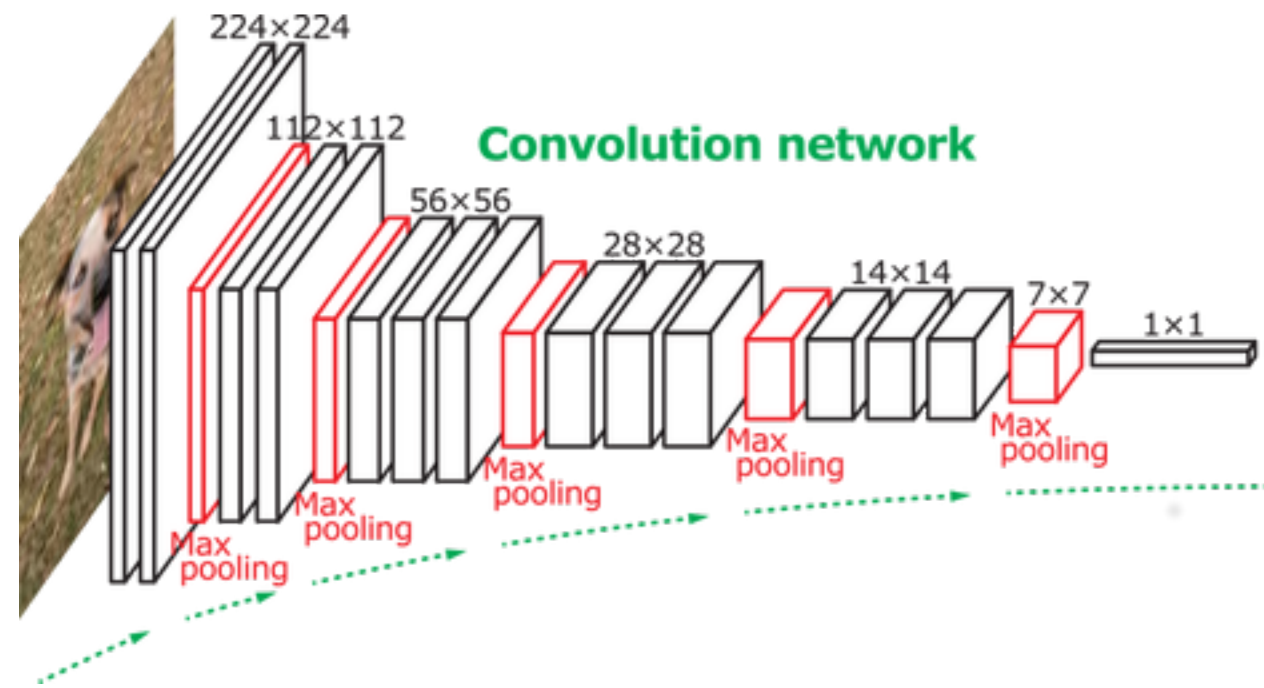
Inferring Final Color

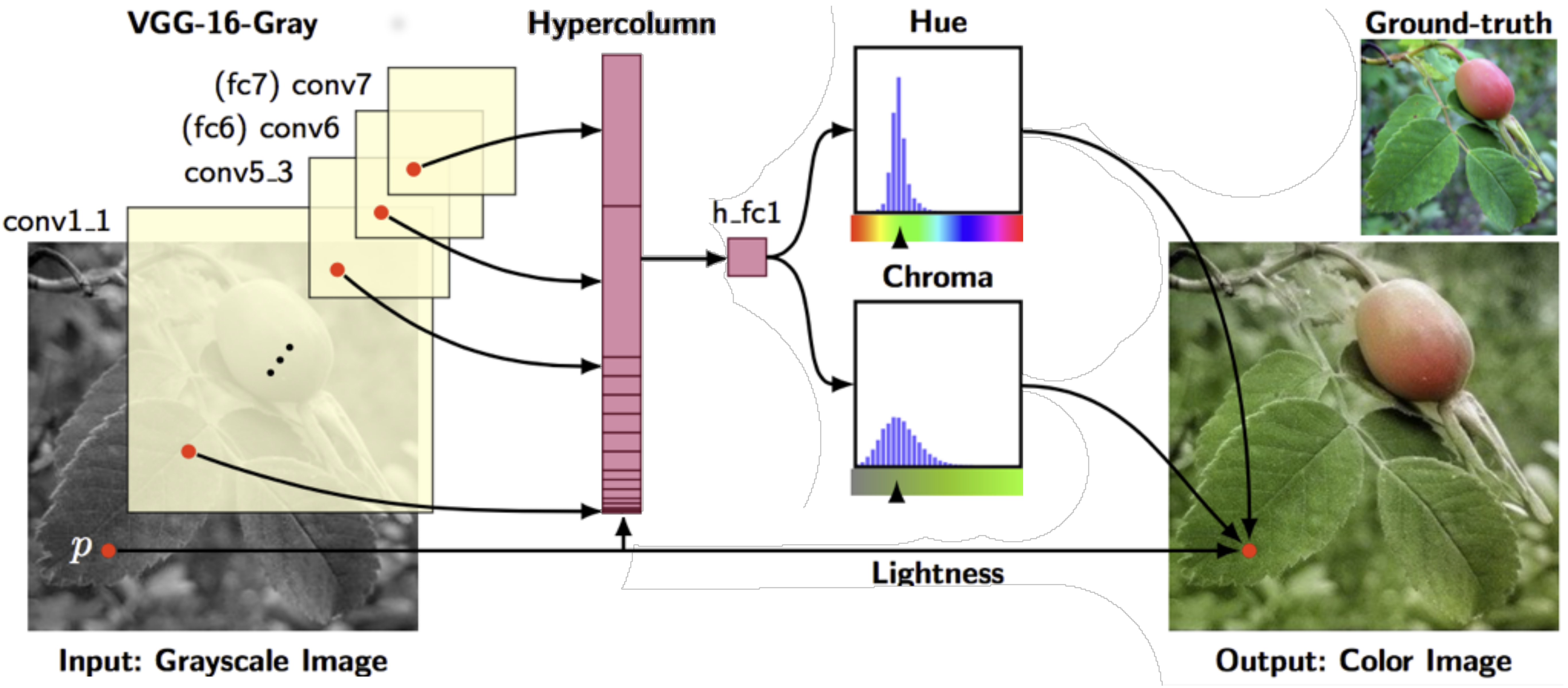
- 4 Methods
  - Sample
  - Mode
  - Median
  - Expectation



Architecture

- Modified VGG-16
- Hypercolumns fed to fully connected layer
  - approximation
- Pre-trained on ImageNet




VGG-16-Gray
Hypercolumn
Hue
Ground-truth
(fc7) conv7
(fc6) conv6
conv5_3
conv1_1
h_fc1
Chroma
p
Lightness
Input: Grayscale Image
Output: Color Image

Results

Input

Author’s Method

Ground Truth

Input

Author’s Method

Ground Truth

Input

Author's Method

Ground Truth

Input

Author’s Method

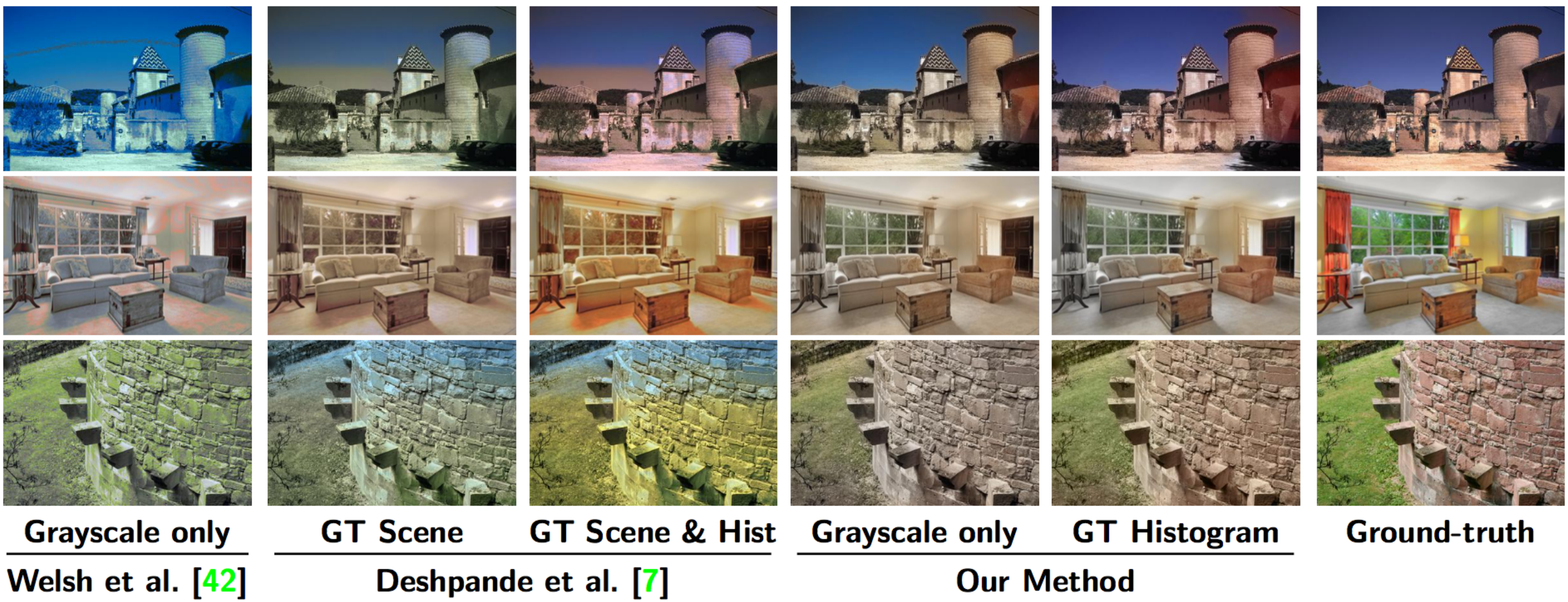
Ground Truth

Failures

Failures

Failures

Comparison Between Methods



Comparison

Method	RMSE
Grayscale (no colorization)	0.285
Welsh *et al.* [42]	0.353
Deshpande *et al.* [7]	0.262
+ GT Scene	0.254
Our Method	**0.211**

Table 3: **SUN-6.** Comparison with competing methods.

Colorization for Image Segmentation

Initialization	Architecture	X	Y	C	mIU (%)
Classifier	VGG-16	✓	✓		64.0
Colorizer	VGG-16	✓			50.2
Random	VGG-16				32.5
Classifier [9, 30]	AlexNet	✓	✓	✓	48.0
BiGAN [9]	AlexNet	✓		✓	34.9
Inpainter [30]	AlexNet	✓		✓	29.7
Random [30]	AlexNet			✓	19.8

Table 6: **VOC 2012 segmentation validation set.** Pretraining uses ImageNet images (X), labels (Y). VOC 2012 images are in color (C).

Strengths and Weaknesses

- Strengths
    - Best results: qualitative and quantitative
    - Fully automated
        - Optional human interaction
    - No hand-crafted features
- Weaknesses
    - Often undersaturated images
    - Trouble predicting background colors

Discussion Points

- Fix undersaturation
- Extend to video colorization
- Post-processing to remove artifacts
  - Edge-detector
  - Texture information