Automatic Colorization

Gustav Larsson

TTI Chicago / University of Chicago

Joint work with Michael Maire and Greg Shakhnarovich

NVIDIA @ SIGGRAPH 2016
Let us first define “colorization”
Colorization

Definition 1: The inverse of desaturation.
Colorization

Definition 1: The inverse of desaturation.

Original  \[\text{Desaturate}\]  Grayscale
Definition 1: The inverse of desaturation.
Definition 1: The inverse of desaturation.
Colorization

Definition 1: The inverse of desaturation. (**Note: Impossible!**)
Colorization

Definition 2: An inverse of desaturation, that...
Colorization

Definition 2: An inverse of desaturation, that...

Our Method

... is plausible and pleasing to a human observer.
Definition 2: An inverse of desaturation, that...

... is plausible and pleasing to a human observer.

- Def. 1: Training + Quantitative Evaluation
- Def. 2: Qualitative Evaluation
I thought I would give it a quick try...
Manual colorization

Grass is green
(low-level: grass texture / mid-level: tree recognition / high-level: scene understanding)
Manual colorization

Sky is blue
Manual colorization

Mountains are... brown?
Manual colorization

Use the original luminosity
Manual colorization

Manual (≈ 15 s)
Manual colorization

Manual (≈ 15 s)

Manual (≈ 3 min)
Manual colorization

Manual (≈ 15 s)

Manual (≈ 3 min)

Automatic (< 1 s)
A brief history

The history of computer-aided colorization in 3 slides.
Method 1: Scribbles

User-defined scribbles define colors. Algorithm fills it in.

Input

Output
Levin et al. (2004)

→ Levin et al. (2004); Huang et al. (2005); Qu et al. (2006); Luan et al. (2007)
Method 2: Transfer

Reference image(s) is provided. Scribbles are automatically created from correspondences.

→ Welsh et al. (2002); Irony et al. (2005); Charpiat et al. (2008); Morimoto et al. (2009); Chia et al. (2011)
Method 2: Transfer

Reference image(s) is provided. Scribbles are automatically created from correspondences.

Reference

Input

Output

Charpiat et al. (2008)

→ Welsh et al. (2002); Irony et al. (2005); Charpiat et al. (2008); Morimoto et al. (2009); Chia et al. (2011)
Method 3: Prediction

Fully parametric prediction.

Automatic colorization is gaining interest recently:

→ Deshpande et al., Cheng et al.; Iizuka & Simo-Serra et al.; Zhang et al., Larsson et al.

**ICCV 2015**
**SIGGRAPH 2016 (2pm, Ballroom E)**
**ECCV 2016**
Method 3: Prediction

Fully parametric prediction.

Automatic colorization is gaining interest recently:

- Deshpande et al., Cheng et al.; Iizuka & Simo-Serra et al.; Zhang et al., Larsson et al.
  - ICCV 2015
  - SIGGRAPH 2016 (2pm, Ballroom E)
  - ECCV 2016
Model

Design principles:
  • Semantic knowledge
Model

Design principles:

- Semantic knowledge $\rightarrow$ Leverage ImageNet-based classifier
Design principles:

- Semantic knowledge → Leverage ImageNet-based classifier
- Low-level/high-level features
Model

Design principles:

- Semantic knowledge → Leverage ImageNet-based classifier
- Low-level/high-level features → Zoom-out/Hypercolumn architecture

Input: Grayscale Image
Output: Color Image
Design principles:

- Semantic knowledge \(\rightarrow\) Leverage ImageNet-based classifier
- Low-level/high-level features \(\rightarrow\) Zoom-out/Hypercolumn architecture
- Colorization not unique

Input: Grayscale Image

Output: Color Image
Model

Design principles:

- Semantic knowledge → Leverage ImageNet-based classifier
- Low-level/high-level features → Zoom-out/Hypercolumn architecture
- Colorization not unique → Predict histograms

![Diagram of the model](image-url)

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

- p
- h_fc1
Going from histogram prediction to RGB:
- Sample
Instantiation

Going from histogram prediction to RGB:

- Sample
- Mode
Going from histogram prediction to RGB:

- Sample
- Mode
- Median
Going from histogram prediction to RGB:

- Sample
- Mode
- Median
- Expectation
Going from histogram prediction to RGB:

- Sample
- Mode
- Median ← Chroma
- Expectation ← Hue
Going from histogram prediction to RGB:

- Sample
- Mode
- Median ← Chroma
- Expectation ← Hue

The histogram representation is rich and flexible:
Instantiation

Going from histogram prediction to RGB:

- Sample
- Mode
- Median $\leftarrow$ Chroma
- Expectation $\leftarrow$ Hue

The histogram representation is rich and flexible:
Results

Significant improvement over state-of-the-art:

Cheng et al. (2015)

Deshpande et al. (2015)
<table>
<thead>
<tr>
<th>Model</th>
<th>AuC CMF</th>
<th>VGG Top-1 Classification Accuracy (%)</th>
<th>Turk Labeled Real (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>non-rebal (%)</td>
<td>rebal (%)</td>
<td>mean</td>
</tr>
<tr>
<td>Ground Truth</td>
<td>100.00</td>
<td>100.00</td>
<td>68.32</td>
</tr>
<tr>
<td>Gray</td>
<td>89.14</td>
<td>58.01</td>
<td>52.69</td>
</tr>
<tr>
<td>Random</td>
<td>84.17</td>
<td>57.34</td>
<td>41.03</td>
</tr>
<tr>
<td>Dahl</td>
<td>90.42</td>
<td>58.92</td>
<td>48.72</td>
</tr>
<tr>
<td>Zhang et al.</td>
<td>91.57</td>
<td>65.12</td>
<td>56.56</td>
</tr>
<tr>
<td>Zhang et al. (rebal)</td>
<td>89.50</td>
<td><strong>67.29</strong></td>
<td>56.01</td>
</tr>
<tr>
<td>Ours</td>
<td><strong>91.70</strong></td>
<td>65.93</td>
<td><strong>59.36</strong></td>
</tr>
</tbody>
</table>

Table: Source: Zhang et al. (2016)
Figure: Failure modes.

Figure: B&W photographs.
Self-supervision (ongoing work)

Colorization as a means to learn visual representations:

1. Train colorization from scratch
Colorization as a means to learn visual representations:

1. Train colorization from scratch
2. Use network for segmentation, detection, style transfer, texture generation, etc.
Self-supervision (ongoing work)

Colorization as a means to learn visual representations:
1. Train colorization from scratch
2. Use network for segmentation, detection, style transfer, texture generation, etc.

<table>
<thead>
<tr>
<th>Initialization</th>
<th>Architecture</th>
<th>$X_{\text{ImageNet}}$</th>
<th>$Y_{\text{ImageNet}}$</th>
<th>Color</th>
<th>mIU (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classifier (ours)</td>
<td>VGG-16</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>64.0</td>
</tr>
<tr>
<td>Random</td>
<td>VGG-16</td>
<td></td>
<td></td>
<td></td>
<td>32.5</td>
</tr>
<tr>
<td>Classifier</td>
<td>AlexNet</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>48.0</td>
</tr>
<tr>
<td>Random</td>
<td>AlexNet</td>
<td></td>
<td></td>
<td>✓</td>
<td>19.8</td>
</tr>
</tbody>
</table>

Table: VOC 2012 segmentation validation set.
Self-supervision (ongoing work)

Colorization as a means to learn visual representations:
1. Train colorization from scratch
2. Use network for segmentation, detection, style transfer, texture generation, etc.

<table>
<thead>
<tr>
<th>Initialization</th>
<th>Architecture</th>
<th>$X_{\text{ImageNet}}$</th>
<th>$Y_{\text{ImageNet}}$</th>
<th>Color</th>
<th>mIU (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classifier (ours)</td>
<td>VGG-16</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>64.0</td>
</tr>
<tr>
<td>Random</td>
<td>VGG-16</td>
<td></td>
<td></td>
<td></td>
<td>32.5</td>
</tr>
<tr>
<td>Classifier</td>
<td>AlexNet</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>48.0</td>
</tr>
<tr>
<td>BiGAN (Donahue et al.)</td>
<td>AlexNet</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>34.9</td>
</tr>
<tr>
<td>Inpainter (Deepak et al.)</td>
<td>AlexNet</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>29.7</td>
</tr>
<tr>
<td>Random</td>
<td>AlexNet</td>
<td></td>
<td></td>
<td>✓</td>
<td>19.8</td>
</tr>
</tbody>
</table>

Table: VOC 2012 segmentation validation set.
Self-supervision (ongoing work)

Colorization as a means to learn visual representations:
1. Train colorization from scratch
2. Use network for segmentation, detection, style transfer, texture generation, etc.

<table>
<thead>
<tr>
<th>Initialization</th>
<th>Architecture</th>
<th>$X_{\text{ImageNet}}$</th>
<th>$Y_{\text{ImageNet}}$</th>
<th>Color</th>
<th>mIU (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classifier (ours)</td>
<td>VGG-16</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>64.0</td>
</tr>
<tr>
<td><strong>Colorizer</strong></td>
<td>VGG-16</td>
<td>✓</td>
<td></td>
<td></td>
<td><strong>50.2</strong></td>
</tr>
<tr>
<td>Random</td>
<td>VGG-16</td>
<td></td>
<td></td>
<td></td>
<td>32.5</td>
</tr>
<tr>
<td>Classifier</td>
<td>AlexNet</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>48.0</td>
</tr>
<tr>
<td>BiGAN (Donahue et al.)</td>
<td>AlexNet</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>34.9</td>
</tr>
<tr>
<td>Inpainter (Deepak et al.)</td>
<td>AlexNet</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>29.7</td>
</tr>
<tr>
<td>Random</td>
<td>AlexNet</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>19.8</td>
</tr>
</tbody>
</table>

Table: VOC 2012 segmentation validation set.
Questions?

Try it out yourself:
http://colorize.ttic.edu
References


