Let there be Color!: Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification

Satoshi Iizuka*  Edgar Simo-Serra*  Hiroshi Ishikawa
Waseda University
(*equal contribution)
Colorization of Black-and-white Pictures
Our Goal: Fully-automatic colorization
Colorization of Old Films

Night of the Living Dead (1968)
Related Work

- **Scribble-based** [Levin+ 2004; Yatziv+ 2004; An+ 2009; Xu+ 2013; Endo+ 2016]
  - Specify colors with scribbles
  - Require manual inputs

- **Reference image-based** [Chia+ 2011; Gupta+ 2012]
  - Transfer colors of reference images
  - Require very similar images
Related Work

- Automatic colorization with hand-crafted features [Cheng+ 2015]
  - Uses existing multiple image features
  - Computes chrominance via a shallow neural network
  - Depends on the performance of semantic segmentation
  - Only handles simple outdoor scenes
Contributions

• Novel end-to-end network that jointly learns **global and local features** for automatic image colorization
  • New fusion layer that elegantly merges the global and local features
  • Exploit classification labels for learning
Layers of Our Model

- Fully-connected layer
  - All neurons are connected between layers
- Convolutional layer
  - Takes into account underlying spatial structure
Our Model

- Two branches: local features and global features
- Composed of four networks
Low-Level Features Network

- Extract low-level features such as edges and corners
- Lower resolution for efficient processing
Global Features Network

- Compute a **global** 256-dimensional vector representation of the image
- Extract mid-level features such as texture
Fusion Layer

- Low-Level Features Network
- Global Features Network
- Mid-Level Features Network
- Colorization Network
- Upsampling

Sharing weights and scaling in the Fusion Layer.
Fusion Layer

- Combine the global features with the mid-level features
- The resulting features are independent of any resolution

\[
y_{u,v}^{\text{fusion}} = \sigma \left( b + W \begin{bmatrix} y_{u,v}^{\text{global}} \\ y_{u,v}^{\text{mid}} \end{bmatrix} \right)
\]
Colorization Network

- Compute chrominance from the fused features
- Restore the image to the input resolution
Training of Colors

- Mean Squared Error (MSE) as loss function
- Optimization using ADADELTA [Zeiler 2012]
  - Adaptively sets a learning rate
Joint Training

- Training for classification jointly with the colorization
- Classification network connected to the global features
Dataset

- MIT Places Scene Dataset [Zhou+ 2014]
- 2.3 million training images with 205 scene labels
  - 256 \times 256 pixels

Abbey, Airport terminal, Aquarium, Baseball field

Dining room, Forest road, Gas station, Gift shop

...
Results
Computational Time

- Colorize within a few seconds

<table>
<thead>
<tr>
<th>Image Size</th>
<th>Pixels</th>
<th>CPU (s)</th>
<th>GPU (s)</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>$224 \times 224^\dagger$</td>
<td>50,176</td>
<td>0.399</td>
<td>0.080</td>
<td>5.0×</td>
</tr>
<tr>
<td>$512 \times 512$</td>
<td>262,144</td>
<td>1.676</td>
<td>0.339</td>
<td>4.9×</td>
</tr>
<tr>
<td>$1024 \times 1024$</td>
<td>1,048,576</td>
<td>5.629</td>
<td>1.084</td>
<td>5.2×</td>
</tr>
<tr>
<td>$2048 \times 2048$</td>
<td>4,194,304</td>
<td>20.116</td>
<td>4.218</td>
<td>4.8×</td>
</tr>
</tbody>
</table>
Colorization of MIT Places Dataset
Comparisons

Input

[Cheng+ 2015]

Ours (w/o global features)

Ours (w/ global features)
Effectiveness of Global Features

Input  w/o global features  w/ global features
User Study

- 10 users participated
- We show 500 images of each type: total 1,500 images per user
- 90% of our results are considered “natural”
Colorization of Historical Photographs

Mount Moran, 1941  
Scott's Run, 1937  
Youngsters, 1912  
Burns Basement, 1910
Style Transfer

Low-Level Features Network
Style Transfer

Low-Level Features Network
Style Transfer

- Adapting the colorization of one image to the style of another

Inputs

Output
Limitations

- Difficult to output colorful images

- Cannot restore exact colors
Conclusion

- Novel approach for image colorization by fusing *global and local information*
  - Fusion layer
  - Joint training of colorization and classification
  - Style transfer
Thank you!

- Project Page  

- Code on GitHub! 
  [https://github.com/satoshiizuka/siggraph2016_colorization](https://github.com/satoshiizuka/siggraph2016_colorization)

Community Center, 1936  
North Dome, 1936  
Norris Dam, 1933  
Miner, 1937

The Lost World (1925)