Review on “Real-Time User-Guided Image Colorization with Learned Deep Priors”

Authors: R. Zhang, J. Zhu et Al.

# Short Summary

This paper formulates an end-to-end combination of deep-learning and user-guidance for image colorization. The network is a CNN that transforms a grayscale image along with user-provided color hints to a colorized output. It is expected that the CNN architecture provides high-level color semantics while user hints provide lower-level cues. During training, the user inputs are simulated and thus the network can colorize with an unusual palette.

In the main colorization network, a U-Net architecture is employed with dilated convolutions. The authors propose two variants of their model: the Local Hints Network (leveraging sparse user points) and the Global Hints Network (leveraging global statistics). In the Local Hints Network, the user interactions are concatenated with the grayscale input. It is found that random sampling of small patches is sufficient for training an effective system. In the Global Hints Network, information is integrated in the middle of the main colorization network. Global statistics for color (via histogram) and saturation (HSV colorspace) are learned.

While experimenting, the authors found that even with minimal training and a minute to work with an image, novice users can quickly create realistic colorizations. Both the global and local networks achieve a higher PSNR (27.85 and 37.70) than the baselines at automatic colorizations even with no user input. The full model is able to score 30.04% on the Amazon Mechanical Turk test meaning that it performs well at fooling humans. Furthermore, it is shown that the model can also generalize to unusual coloring schemes with minimal user input.

# Main Contributions

1. End-to-end model for colorizing an image that uses both global learned data and user input
2. Provides an interactive tool for users to define their own colorizations which generalizes well
3. Image colorizations can be learned by the model within one minute, even with minimal training, by a novice user
4. Model can be trained with any statistic of the output (color dist. or saturation)

# High-Level Evaluation of Paper

The paper does an excellent job setting up the problem and then describing the various historical solutions. It also provides a strong intuition as to why their approach, using both deep-learning and user cues, would outperform older methods. I didn’t, however, get a good sense of how the architectures of the global and local networks were selected. Interestingly, in the results section, the authors structured the sub-section headers as questions that they were going to answer. This made it easy to find relevant information and get a sense of how well the model realistically performs. The visualizations were also helpful in following the paper; for example, coloring Mark Ruffalo green to showcase unusual colorings.

# Discussion on Evaluation Methodology

The paper makes several distinct evaluations: PSNR on automatic colorization compared to baselines, AMT rate of model, and PSNR on the Global Hints Network at incorporating the global truth. In all cases, the proposed model seems to exceed previous state-of-the-art. The evaluation was very thorough though the paper didn’t really describe how one should interpret PSNR and what it means in the context of the task.