A convolutional classification approach to colorization

Vincent Billaut
Department of Statistics
Stanford University
vbillaut@stanford.edu

Matthieu de Rochemonteix
Department of Statistics
Stanford University
mderoche@stanford.edu

Marc Thibault
ICME
Stanford University
marcthib@stanford.edu

Abstract

Insert abstract here

Introduction

The problem of colorization is one that comes quickly to mind when thinking about interesting challenges involving pictural data. Namely, the goal is to build a model that takes the greyscale version of an image (or even an actual "black and white" picture) and outputs its colorized version, as close to the original as possible (or at least realistic, if the original is not in colors). This problem is complex and interesting for several reasons, as the final output needs to be an image of the same dimension as the input image. We want to train a model that is able to recognize shapes that are typical of a category of items and apply the appropriate colorization.

One clear upside to this challenge is that any computer vision dataset, and even any image bank really, is a proper dataset for the colorization problem (the image itself is the model's expected output, and its greyscale version is the input to the model). The input to our algorithm is simply an image in grayscale, and the output is the same image, colorized. A conversion of the images to the YUV format allows an easy formulation of the problem in terms of a reconstitution of the U and V channels.

We formulate the colorization problem as a classification problem to gain flexibility in the prediction process and output more colorful images. We aim at reproducing state of the art results that give vivid, visually appealing results, with a much smaller network.

1. Related Work

Classical approaches to this task, *e.g.* [1] and [2], aim at predicting an image as close as possible to the ground truth, and notably make use of a simple L_2 loss, which penalizes predictions that fall overall too far from the ground truth. As a consequence, the models trained following such methods usually tend to be very conservative, and to give desaturated, pale results.

On the contrary, authors of [4] take another angle and set their objective to be "*plausible* colorization" (and not necessarily *accurate*), which they validate with a large-scale human trial.

Their approach is the most appealing as they have colorful results and we found the classification formulation to be interesting. However, the network architecture they use is very heavy (taking more than 20 GB of memory), and the scale of their training set (several millions of images). The reason behind this is that to encode meaningful fetures that help to colorize the image, one needs to have a large spatial receptive field. THe approach of the article is to downsample the image a lot in the layers in the middle and then upsample using Convolutional Transpose layers. To keep a lot of information in the intermediate layers, the number of filters in the model of [4] is very large, resulting in a model that is large and expensive to train.

Authors of [3] have shown that conenctions between hidden layers of a bottleneck neural network could enhance the performance gratly, by helping the upsampling process and improving the gradient flow.

We hope that applying this method will allow us to train a colorizing model more quickly and more efficently, with less parameters.

2. Methods

3. Dataset and Features

4. Results and discussion

5. Conclusion and perspectives

Contributions and acknowledgements

References

[1] Z. Cheng, Q. Yang, and B. Sheng. Deep colorization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 415–423, 2015.

[2] R. Dahl. Automatic colorization. `http://tinyclouds.org/colorize/`, 2016.

[3] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.

[4] R. Zhang, P. Isola, and A. A. Efros. Colorful image colorization. In *European Conference on Computer Vision*, pages 649–666. Springer, 2016.