

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
marcth@stanford.edu

Abstract

Insert abstract here

Introduction

The problem of colorization is one that comes quickly to mind when thinking about interesting challenges involving pictorial 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

Historically, older approaches of the colorization problem use an additional input, a *seed scribble* from the user to propagate colors, as does [8]. It may also be seen as a corollary of color style transfer algorithms using a similar image as a “seed” as in [5]. Here, we are interested in fully

automatic recolorization

Classical approaches to this task, *e.g.* [2] and [3], 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. Usually, the images need some postprocessing adjustments as in [4] to have a realistic aspect.

On the contrary, authors of [10] take another angle and set their objective to be “*plausible* colorization” (and not necessarily *accurate*), which they validate with a large-scale human trial. To achieve such results, they formulate the colorization task as a classification task using color bins, as suggested already in [1].

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 features that help to colorize the image, one needs to have a large spatial receptive field. The approach of the article is to down-sample 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 [10] is very large, resulting in a model that is large and expensive to train.

Authors of [9] have shown that connections between hidden layers of a bottleneck neural network could enhance the performance greatly, 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 efficiently, with less parameters.

Part of the challenges that are interesting but not yet tackled in the literature involve videos. General information propagation frameworks in a video involving bilateral networks as discussed in [6] could be seen as a potential way

to implement consistent colorization of video sequences. The work realized in [11] is also interesting since it tackles video stylization by grouping the frames, choose a representative frame for each group and use the output of the network on this frame as a guideline, which enhances temporal consistency. However, the adaptation of such an algorithm to the much more complex task of image colorization is far beyond the scope of this project.

Actually, one promising way to perform image colorization is to be able to learn meaningful color-related representations for the images (which often involves using very deep and heavy or pretrained architecture as in [7]) and then ensure the temporal consistency of them.

Given our commitment to developing a lightweight model, we preferred focusing on an efficient colorization for images and then add a layer of temporal convolution for videos.

2. Methods

3. Dataset and Features

4. Results and discussion

5. Conclusion and perspectives

Contributions and acknowledgements

References

- [1] G. Charpiat, M. Hofmann, and B. Schölkopf. Automatic image colorization via multimodal predictions. In *European conference on computer vision*, pages 126–139. Springer, 2008.
- [2] Z. Cheng, Q. Yang, and B. Sheng. Deep colorization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 415–423, 2015.
- [3] R. Dahl. Automatic colorization. <http://tinyclouds.org/colorize/>, 2016.
- [4] A. Deshpande, J. Rock, and D. Forsyth. Learning large-scale automatic image colorization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 567–575, 2015.
- [5] M. He, J. Liao, L. Yuan, and P. V. Sander. Neural color transfer between images. *CoRR*, abs/1710.00756, 2017.
- [6] V. Jampani, R. Gadde, and P. V. Gehler. Video propagation networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, 2017.
- [7] G. Larsson, M. Maire, and G. Shakhnarovich. Learning representations for automatic colorization. *CoRR*, abs/1603.06668, 2016.
- [8] A. Levin, D. Lischinski, and Y. Weiss. Colorization using optimization. In *ACM Transactions on Graphics (ToG)*, volume 23, pages 689–694. ACM, 2004.
- [9] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.
- [10] R. Zhang, P. Isola, and A. A. Efros. Colorful image colorization. In *European Conference on Computer Vision*, pages 649–666. Springer, 2016.
- [11] F. Zhu, Z. Yan, J. Bu, and Y. Yu. Exemplar-based image and video stylization using fully convolutional semantic features. 26:3542–3555, 07 2017.