Project Proposal: Iterative Coarse-to-Fine Image Colorization with GAN

Thomas Rüegg (16-926-396), Timo Streule (16-938-706), Marc Styger (16-922-767)

I. Summary of the Proposal

Working with high-resolution images in neural networks can be quite resource-intensive. Training these nets is therefore almost infeasible for personal computers. In this work, we'd like to propose a novel approach for coloring high-resolution greyscale images without consuming massive computational resources and without losing neighboring, contextual information. Our main contribution is to process the image at different scales (from coarse to fine) and feeding it back into the network for an improved prediction.

II. Background

The problem of colorizing greyscale images has been approached in various ways. For instance, one can rely on significant user interaction [see 9] and thus significantly slow down the process. Other approaches pose the problem as a classification task [see 8, 4] or, conversely, use a regression loss as objective [see 2].

In Isola et al. [3] the authors proposed a general solution to image-to-image tasks in deep learning by using a combination of two losses for the discriminator objective. The usage of a conditional generative adversarial network (cGAN) helps to solve the problem in an unsupervised manner and to produce good-looking colorful images that seem real. Combining this with the L1 loss, which makes it a regression task, introduces some supervision and reduces the effect of producing gray-ish images. The final objective is a λ-weighted sum of both objectives.

One major drawback of image colorization methods based on deep learning is time for computation. Especially when increasing the image resolution the process gets more time consuming. A natural solution to speed up the process is to split images in smaller patches, colorize and then reassemble all patches [5]. In this patch-wise colorization approach, the patches loose context to neighboring patches and boundary artifacts will occur. By means of a feedback loop as described in section IV we intend to overcome this issue.

III. Goal and Objectives

We intend to implement a fully convolutional network with residual blocks in a GAN-like structure to achieve more vivid/vibrant colors. Which has been proven to work quite well on small images. With our contribution our goal is to enable the network to also color larger images while achieving this without a significantly higher computing time.

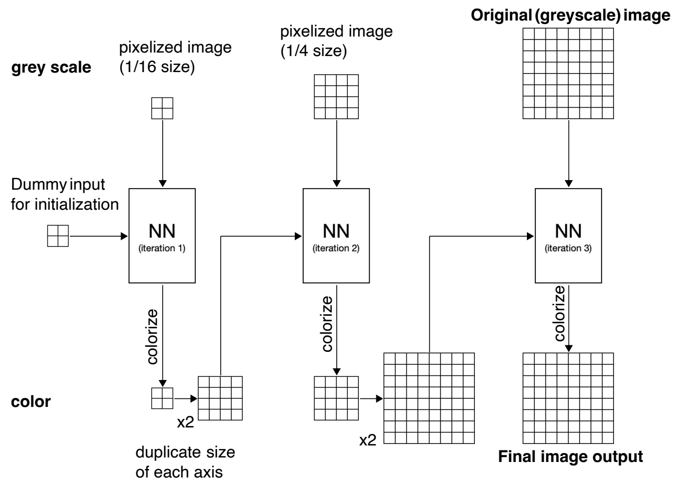

Figure 1. Schematic of the approach.

To do this we propose an iterative approach. The main idea is to process the picture at different scales, starting small and then double the size after each iteration and concatenate the (up-scaled) prediction of the last iteration to the input of the current iteration.

The scope of this project includes:

- Implement and train baseline U-Net model [7].
- Add our proposed method to the baseline model.
- Quantitative and qualitative comparison of our results to the baseline model (L1 and SSIM).

IV. Population and Methods

Our architecture is based on a U-Net because it is able to work with dynamic input sizes. As loss function we use a combination of L1 loss with an adversarial agent as proposed by Isola et al. [3]. This will be our baseline model which we can use as comparison. The good thing about the U-Net is that we can easily use a pretrained model as a backbone. This will reduce the training time significantly.

Our approach is to add a feedback loop to this architecture which will go through the same picture at different scales and concatenate the predictions of the previous iteration to the input of the current iteration.

There are different options for training the network:

- Training all iterations together where the loss gets calculated at the very end
- Training every iteration separately where the prediction concatenated to the input can be either a prediction from the network or the ground truth.

As dataset we can either use ImageNet [1] or Microsoft Coco [6].

References

[1] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, pages 248–255. Ieee, 2009.

[2] S. Iizuka, E. Simo-Serra, and H. Ishikawa. Let there be Color!: Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification. ACM Transactions on Graphics (Proc. of SIGGRAPH 2016), 35(4):110:1–110:11, 2016.

[3] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks, 2018.

[4] G. Larsson, M. Maire, and G. Shakhnarovich. Learning representations for automatic colorization, 2017.

[5] X. Liang, Z. Su, Y. Xiao, J. Guo, and X. Luo. Deep patch-wise colorization model for grayscale images. In SIGGRAPH ASIA 2016 Technical Briefs, SA '16. Association for Computing Machinery, 2016. ISBN 9781450345415. doi: 10.1145/3005358.3005375. URL https://doi.org/10.1145/3005358.3005375.

[6] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár. Microsoft coco: Common objects in context, 2015.

[7] M. Shariatnia. Colorizing black & white images with U-Net and conditional GAN – A Tutorial, Nov 2020. URL https://towardsdatascience.com/colorizing-black-white-images-with-u-net-and-conditional-gan-a-tutorial-81b2df111cd8. [Online; accessed Dec 7, 2021].

[8] R. Zhang, P. Isola, and A. A. Efros. Colorful image colorization, 2016.

[9] R. Zhang, J.-Y. Zhu, P. Isola, X. Geng, A. S. Lin, T. Yu, and A. A. Efros. Real-time user-guided image colorization with learned deep priors, 2017.