Image Colorization

Markéta Minářová

November 30, 2022

1 Introduction

Image colorization is one of the many current challenges in image processing. Given a grayscale image, the goal is to estimate colours for all of the image's pixels. Deep learning techniques have been broadly used throughout the time and some of them resulted in pretty solid solutions. However, it is an ambiguous task sometimes as not all objects tend to possess one colour shade – for example, the colorization is quite straightforward when we are given a tooth - it will most probably be white (or some shade of white to yellow), but when given a car, it could be any color.

The reason for colouring images can be for example restoring old black and white images and enhancing them by giving them colors. Most of the works that I reviewed focused on coloring houses, landscapes, people - sometimes pretty straightforward colouring. However, I was interested in what the models can do when faced with paintings and portraits, art throughout the time periods. Therefore, I downloaded a representative art dataset containing paintings of landscapes, portraits and more. In the next part, I will review some of the methods that have been utilized by others.

2 Related literature

The first paper I read was [1]. The authors use CNN to colorize black and white images from the ImageNet dataset [2]. The images were represented in the Lab colorspace, which expresses color as three values: Lightness, a and b, which expresses four colors – red, green, yellow and blue. The lightness defines black at 0 and white at 100 and is the one channel that we want when representing a grayscale image. The combinations of a and b axis make the whole colour. Negative values at axis a tilts more toward green and positive ones tilt to red. Negative values at axis b tilt toward blue and the positive ones towards yellow.

Their network architecture consists of 8 convolutional layers, while each of them is repeated twice of thrice. There are also ReLU layers and a BatchNorm layer, which speeds up the training.

As the colorization task is multimodal – meaning that there can be several 'right' colors for an object, it is not ideal to utilize standard loss functions, where

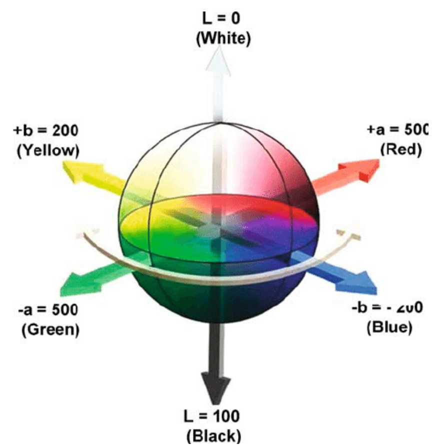

Figure 1: The LAB color space. Source: `https://www.researchgate.net/figure/The-cubical-CIE-Lab-color-space_fig3_23789543`

the goal is to minimize the error between a prediction and the real value. Thus, the authors decided to predict a distribution of possible colors for each pixel. They also re-weight the loss while training to make sure some non-generic colours have a shot. The final colorization process takes a mean of the distribution for each pixel.

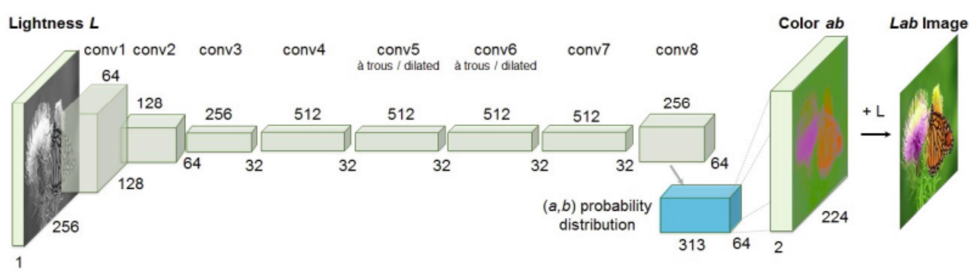

Figure 2: The Architecture. Source: [1]

They also came up with a "colorization Turing test" in which they showed people real images and artificially coloured ones to see if they can recognize them from one another. 32 % of the artificially coloured images were classified as real ones.

The next paper I read was one that uses uses the Split-Brain Autoencoders for the colorization task [3]. The split in this case means that the network is divided into two sub-networks, where each one is trained to estimate values from the other one. This is called the cross-channel prediction.

They compare two approaches regarding the loss function. Firstly, they simply use the L_2 distance and the Sum of Squared Errors, letting the layers

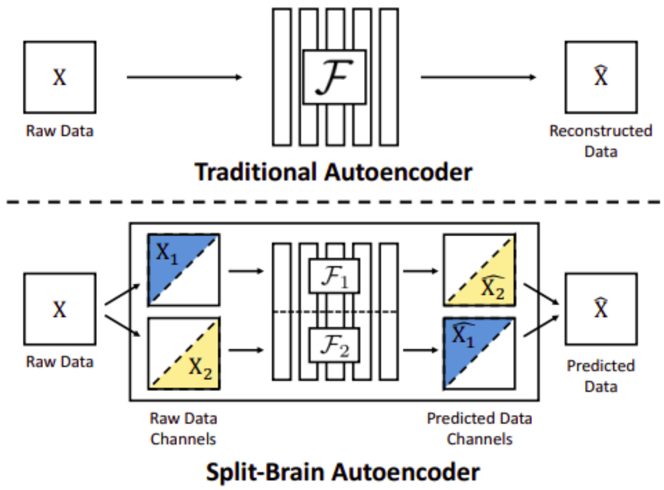

Figure 3: Difference between autoencoder and split-brain autoencoder. Source: [3]

predict color pixels of one another. In the second approach, they let the layers predict the distribution of possible colours. To measure the errors, they use the standard cross-entropy loss between the real and predicted distributions.

They tested their model on Lab images and RGB-D Imaged (combination of an RGB image and their depth image). They conclude that their approach is comparable to the state-of-the-art approaches, if not a little better.

3 Dataset

I downloaded several separate art datasets and then concatenated them into a folder of 6883 images. First dataset that I downloaded is from `https://www.kaggle.com/datasets/ikarus777/best-artworks-of-all-time`. The author downloaded paintings from the top 50 most influentual artists from artchallenge.ru website. The second dataset I downloaded was from `https://data.mendeley.com/datasets/289kxpnp57/1`, which is a collection of of portraits.

I also converted these images to grayscale images, thus creating a parallel folder.

4 Models

I will implement and compare three different models to find out which one tackles the problem of colorization the best.

The images are represented in the Lab color space and the predicted channels are a and b, so it is two instead of three as would be in the RGB color space. I would like to predict the color distribution rather than the specific color.

4.1 CNN

The first model is CNN, inspired by the article that I read [1].

4.2 Autoencoders

4.3 GAN

5 Results

6 Conclusion

References

1. ZHANG, Richard; ISOLA, Phillip; EFROS, Alexei A. Colorful image colorization. In: *European conference on computer vision*. Springer, 2016, pp. 649–666.
2. RUSSAKOVSKY, Olga; DENG, Jia; SU, Hao; KRAUSE, Jonathan; SATHEESH, Sanjeev; MA, Sean; HUANG, Zhiheng; KARPATHY, Andrej; KHOSLA, Aditya; BERNSTEIN, Michael, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*. 2015, vol. 115, no. 3, pp. 211–252.
3. ZHANG, Richard; ISOLA, Phillip; EFROS, Alexei A. Split-brain autoencoders: Unsupervised learning by cross-channel prediction. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 1058–1067.