Automatic image colorization: a comparative overview

Bigarella Chiara
Student nr. 2004248

Poletti Silvia
Student nr. 1239133

Abstract

The ABSTRACT is to be in fully-justified italicized text, at the top of the left-hand column, below the author and affiliation information. Use the word "Abstract" as the title, in 12-point Times, boldface type, centered relative to the column, initially capitalized. The abstract is to be in 10-point, single-spaced type. Leave two blank lines after the Abstract, then begin the main text. Abstract should be no longer than 300 words.

1. Introduction

Introduction (10%): describe the problem you are working on, why it's important, and an overview of your results.

2. Related Work

Related Work (10%): discuss published work or similar apps that relates to your project. How is your approach similar or different from others?

Papers are: [12], [11], [14], [8],[10],[13],[9].

3. Dataset

We considered three types of images: 4023 originally colored images from five different datasets, 18 originally black and white images from various artists and 180 filtered images (see more details in Experiments section) obtained starting from 18 originally colored images.

Indeed, our data includes heterogeneous images, representing many different environments, situations and subjects. For what concerns the originally colored images, we considered various sources:

- subset of ImageNet made of 12 classes (200 images each) taken from [6], ten of which are easily classified classes (tench, English springer, cassette player, chain saw, church, French horn, garbage truck, gas pump, golf ball and parachute) while the other two are not so easy to classify (Samoyed and Rhodesian ridgeback);
- subset of 100 randomly selected images from Pascal VOC [2] representing realistic scenes in which the subjects could be animals, human beeings, plants, rooms, landscapes, various objects and vehicles;
- subset of 200 randomly selected images form Places205 [3] reguarding mountain, desert, sea, beach and island landscapes.
- subset of 325 Bird Species [4] made of 8 classes (100 images each), which were selected to depict those birds having the most unusual colors (Cuban Tody, Fire Tailed Myzornis, Flamingo, Nicobar Pigeon and Pink Robin) and those that are well-known by the majority of people (Bald Eagle, Ostrich and Touchan);
- subset of 102 Category Flowers [1] made of 6 classes (from 50 to 100 images each), which were selected to depict those flowers having the most unusual colors and shapes (Purple Coneflower, Grape Hyacinth, Hibiscus) and those that are well-known by the majority of people (Rose, Water Lily and Giant White Arum Lily).

The images have been treated by using OpenCV (Chromagan and InstColorization) or Pillow combined with Skimage (Baseline, Dahl, Zhang, Siggraph).

The images have been reshaped to various formats ($256 \times 256 \times 3$ for Baseline, Zhang, Siggraph and InstColorization and $224 \times 224 \times 3$ for Dahl and Chromagan) and Dahl also required center cropping and desaturation. Despite the preliminar reshape, Zhang, Siggraph and Chromagan models are built in a way that allows to obtain colorized images having the original shape.

Given an RGB image (additive colour model in which red, green and blue primary colour channels are added together) we obtain the corrisponding image in the *Lab* color space, in which colors are expressed through 3 new channels: L for perceptual lightness ($L = 0$ corresponds to white, $L = 100$ corresponds to black), a and b for four primary colors ($a = \pm 100$ correspond to red and green, $b = \pm 100$ correspond to yellow and blue).

Our models get only the L channel as input (greyscale images) with the goal of predicting the a and b channels. Then, the resulting images are projected again in the RGB color space.

Moreover, the classification with AlexNet required the normalization of the images' RGB channels in the range $[0, 1]$ and a further standardization of the images according to the mean and standard deviation of the training set images.

On the other hand, the LPIPS metric required the normalization of the images' RGB channels in the range $[-1, 1]$ and the dataset reshaping from $N \times H \times W \times 3$ to $N \times 3 \times H \times W$, where N is the number of images.

4. Methods

In order to carry out a comparative overview about automatic image colorization, we built, trained and tested a simple autoencoder based on cartoonization, to be considered as baseline. Then, we tested some state-of-the-art pre-trained models taken from the literature: Dahl, Zhang and its upgraded version Siggraph, Chromagan and InstColorization.

4.1. Baseline

As a baseline, we built with Keras a simple autoencoder having 8 Convolutional layers for the encoding part (relu activations, zero-padding, 3×3 kernels and sometimes 2×2 strides), while the decoding part consisted in a combination of 5 Convolutional layers (relu activations except for the last layer, zero-padding and 3×3 kernel) and 3 UpSampling layers of size 2×2. The encoder learns a compact representation of the black and white input image and the decoder generates the corresponding novel coloured image.

The model was trained (50 epochs) on a *mixed dataset* containing a fraction of the ImageNet subset and all the other data described in the Dataset section.

Moreover, we enriched this model with a novel approach: instead of using the original dataset, we fed the model with the cartoonized (black and white) version of the images, computed with the pre-trained GAN cartoonization model by [13]. This cartoonization provides fine-grained results (we don't miss much information) and synthesizes the original images in order to exclude noisy elements that could interfer with the colorization task.

As a result, the model produces cartoonized RGB colored images, which are then converted as Lab images. After extracting the a and b channels from those images, we combine them with the L channel of the original Lab images. Therefore, we mantain the original details of the pictures, while producing a more precise and sectorial colorization.

For comparison, we also include in our experiments the Baseline without cartoonization (Baseline w/c).

4.2. Dahl

4.3. Zhang and Siggraph

4.4. Chromagan

4.5. InstColorization

5. Experiments

To compare the results of each model, we computed several metrics: classification with AlexNet, LPIPS, PSNR and SSIM (all quantitative metrics) and a Turing test on few images (qualitative metric). Finally, we applied image filtering to evaluate possible improvements in the performances.

5.1. Classification with AlexNet

First, we considered the AlexNet classifier pre-trained on ImageNet and tested on the ImageNet subset in its original, black and white and re-colorized versions.

Table 1 reports the AlexNet classification accuracy in this setting and in other two settings that we will discuss later in this section. Note that the Baseline without cartoonization (Baseline w/c) always reaches a slightly worse accuracy than the Baseline combined with cartoonization, meaning that our approach is valid and can actually improve the colorization performance.

The great gap in the accuracies computed on the original and the black and white versions of the images suggests that colors play an important role in image classification.

The best colorizations according to this experiment are given by Chromagan and InstColorization, while the Baseline and Dahl are not even able to improve the accuracy with respect to the black and white images.

Overall, the accuracy on the models' colorizations is much lower than the one computed on the original images and the latter is relatively low. Therefore we applied feature extraction to better focus on our ImageNet subset: we used the pre-trained AlexNet as a fixed feature-extractor, and only updated the final layer (for 2 epochs) in order to consider just our 12 ImageNet classes. This resulted in more reliable accuracy values and all the models except the Baseline are able to outperform the black and white images.

For a further comparison, we applied finetuning to perform classification on the birds and flowers images, which present more vibrant and various colors than our ImageNet subset: we updated (for 2 epochs) all the AlexNet parameters for the new task. In this new setting we have, as expected, a greater gap than before between the original and the black and white accuracies, meaning that the color is much more relevant. Indeed, all the models including the Baseline with cartoonization are able to improve the accuracy with respect to the black and white images.

Figure 1: Colorization comparison on two images from ImageNet Church (first row) and 325 Birds Species Flamingo (second row).

The best colorizations according to this experiment are given by the Zhang and Siggraph models, which are able to generalize better across different datasets.

In this last setting, we can notice a general decreasing in the accuracy (except for the original images) with respect to the feature extraction using the ImageNet subset. This is due to the fact that our pre-trained models have been trained on Image-Net and their colorization of the birds and flowers images are overall bad. However, looking at our results, a badly colored image generally seems more distinguishable than its black and white version.

To conclude, the colorizations of two images depicting a church (ImageNet) and a flamingo (325 Bird Species) are reported as an example in Figure 1.

5.2. LPIPS, PSNR and SSIM Metrics

[5] and [7]

5.3. Turing Test

5.4. Image filtering

6. Conclusion

Conclusion (5%): summarize your key results; what have you learned? Suggest ideas for future extensions.

References

[1] 102 category flowes dataset. https://www.robots.ox.ac.uk/~vgg/data/flowers/102/, 2008.

[2] Pascal voc dataset. https://deepai.org/dataset/pascal-voc, 2012.

[3] Places205 dataset. https://paperswithcode.com/dataset/places205, 2014.

[4] 325 bird species dataset. https://www.kaggle.com/gpiosenka/100-bird-species, 2019.

[5] Peak signal-to-noise ratio and structural similarity metrics. https://cvnote.ddlee.cc/2019/09/12/psnr-ssim-python, 2019.

[6] Imagenette and imagewoof datasets. https://github.com/fastai/imagenette, 2021.

[7] Learned perceptual image patch similarity metric. https://github.com/richzhang/PerceptualSimilarity, 2021.

[8] Ryan Dahl. Automatic colorization, 2016.

[9] Jianbo Chen et al. Language-based image editing with recurrent attentive models, 2018.

[10] Seungjoo Yoo et al. Coloring with limited data: Few-shot colorization via memory-augmented networks, 2019.

[11] Jheng-Wei Su, Hung-Kuo Chu, and Jia-Bin Huang. Instance-aware image colorizationl, 2020.

[12] Patricia Vitoria, Lara Raad, and Coloma Ballester. Chromagan: Adversarial picture colorization with semantic class distribution, 2020.

[13] Xinrui Wang and Jinze Yu. Learning to cartoonize using white-box cartoon representations, 2020.

[14] Richard Zhang, Phillip Isola, and Alexei A. Efros. Colorful image colorizationn, 2016.

	Original	B&W	Baseline w/c	Baseline	Dahl	Zhang	Siggraph	Chromagan	InstColorization
Pre-trained	74.5%	43.1%	32.5%	34.0%	39.8%	42.7%	43.2%	46.8%	49.5%
Feature Extraction	97.2%	84.2%	79.4%	80.4%	87.8%	87.6%	88.9%	90.2%	90.0%
Finetuning	99.7%	63.0%	58.4%	63.6%	64.4%	80.9%	79.9%	77.9%	73.7%

Table 1: Summary of the classification accuracy of Alexnet in three different settings: Alexnet pre-trained on Imagenet, Alexnet feature extraction for the ImageNet subset, AlexNet finetuning for the Birds and Flowers dataset.