Image Colorization

Bigarella Chiara
chiara.bigarella.1@studenti.unipd.it

Poletti Silvia
silvia.poletti@studenti.unipd.it

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: [10], [9], [12], [6],[8],[11],[7].

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 [5], 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 Landscape Pictures [4] reguarding montain, desert, sea, beach, island and Japan landscapes.
- subset of 325 Bird Species [3] 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

Method (30%): discuss your approach for solving the problems that you set up in the introduction. Why is your approach the right thing to do? Did you consider alternative approaches? It may be helpful to include figures, diagrams, or tables to describe your method or compare it with others.

- Baseline - Dahl - Zhang - Siggraph - ChromaGAN - Su

5. Experiments

Experiments (30%): discuss the experiments that you performed. The exact experiments will vary depending on the project, but you might compare with prior work, perform an ablation study to determine the impact of various components of your system, experiment with different hyperparameters or architectural choices. You should include graphs, tables, or other figures to illustrate your experimental results.

- Ricolorare immagini - Metriche - Filtri - Baseline con cartoonization - Turing Test

- Classification con 1) AlexNet pretrained: l'accuracy su B&W è minore rispetto alle immagini originali e rispetto alla maggiorparte dei nostri modelli. l'accuracy delle immagini ricolorate dai nostri modelli è comunque molto più bassa rispetto all'accuracy delle immagini originali. quindi abbiamo provato a fare feature extraction sulle nostre classi selezionate da ImageNet per avere un risultato più reliable per quanto riguarda l'accuratezza dei nostri modelli

2) feature extraction -> accuracy di alexnet sulle classi selezionate di ImageNet: confronto tra black and white e colorized. AlexNet performa peggio sulle B&W, ma non tanto peggio, perchè il colore non è rilevante per le immagini di ImageNet che abbiamo scelto => abbiamo provato finetuning con un altro dataset contenente fiori e uccelli

3) finetuning -> accuracy di alexnet sulle classi del nuovo dataset: confronto tra black and white e colorized. AlexNet performa tanto peggio sulle B&W perchè questa volta il colore è molto più rilevante. Tuttavia, l'accuracy generale è più bassa rispetto a feature extraction perchè: 1) il nuovo dataset è molto più piccolo di ImageNet 2) il numero di epoch (2) è troppo piccolo rispetto alla grande differenza che c'è tra i due dataset 3) i nostri modelli non colorano molto bene queste immagini perchè sono pretrained su ImageNet

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

[4] Landscape pictures dataset. `https://www.kaggle.com/arnaud58/landscape-pictures?select=00000000.jpg`, 2019.

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

[6] Ryan Dahl. Automatic colorization, 2016.

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

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

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

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

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

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