Bringing Images to Life with Convolutional Neural Networks

Anh Phung, James Shin (Team 15)
Deep Learning Project Proposal, COMS W4995 Prof. Iddo Drori
{adp2161@cs.columbia.edu, js4785@columbia.edu}

Abstract

We introduce a novel colorization system using convolutional neural networks. We aim to enhance existing systems to achieve higher colorization accuracy in order to develop a better computational understanding of images in image processing.

Keywords: image colorization, convolutional neural networks, computer vision

1. Introduction

Colorization is a highly practical field with various applications, including correcting color in legacy images and colorizing historical grayscale photographs and movie archives. Automatic colorization also leads to insights for other image processing techniques in computer vision by serving as a proxy measure for high-level visual understanding and a promising self-supervised representation learning method.

2. Previous Work

Many attempts have been made to build an efficient automatic image colorization system. Traditional methods tend to depend on user interactions and extremely large databases of image references. In the last few years, Convolutional Neural Networks (ConvNets) have revolutionized the way researchers approach computer vision and image processing. Recent works have proven that with the help of novel deep learning techniques, we can develop fully automatic colorization systems which produce natural, vibrant and lifelike results to unsaturated images.

Some state-of-the-art work include Zhang (Zhang et al., 2016), who presents a colorization system that is a feed-forward pass in a CNN and is trained on over a million colored images. They focus on handling the problem of multimodal uncertainty in colorization to capture a wider diversity of colors, and on building a testing framework to test colorization problems. Unlike Zhang (Zhang et al., 2016), Iizuka (Iizuka et al., 2016) does not focus on the multimodal nature of colorization, but rather on the incorporation of low and high-level features. Their approach greatly outperforms other approaches that use local features only, and are able to exploit semantic class labels during training to learn more global features. By combining low and high-level features, they manage to leverage the semantic information about dynamic objects like human skin or hair without human interaction.

Other work in image colorization do not necessarily rely on ConvNet techniques (Gupta et al.,), but rather the input of a semantically-similar reference image supplied by the user. The algorithm extracts features of the image and generates superpixels, which are perceptually-similar pixels that greatly reduce the properties of analysis of the image for future image processing steps. The extracted features of the reference image are then imposed onto the input image. The advantage to the algorithm is no training time and independence of large data sets for analysis, at the expense of high dependence on the input image and lower accuracy.

3. Method proposed

We propose a new method of combining the previous work in colorizing images. We will still use a CNN similar to Zhang (Zhang et al., 2016) or Iizuka (Iizuka et al., 2016), but at the final step, the network will take into account of user reference images. This method will combine both the state-of-the-art CNN model to capture semantically meaningful colors as well as the ability to consistently do so under a user's request. This consistency is important, especially when we move on to deal with colorizing black and white videos, which consist of sequence of images and require the colors of a scene or a character to be mostly the same. We will use ImageNet as the training and test set for our model.

4. Evaluation criteria

4.1. Raw Accuracy

This test will essentially check how different the ground truth image is with an image produced by our network on a black-and-white version of the same image. Raw accuracy will be calculated with respect to the RGB color value of each pixel. The raw accuracy is computed as the percentage of predicted pixel colors within a thresholded L_2 distance from the ground truth.

4.2. Semantic Interpretability

This metric tests if the colorization of images is realistic enough to be interpretable to an image classifier, and can be used as training or test data to the object classifier.

5. Bibliographical References

Gupta, R. K., Chia, A. Y.-S., Rajan, D., Ng, E. S., and Zhiyong, H.). Image colorization using similar im- ages. *ACM international conference on Multime- dia*.

Iizuka, S., Simo-Serra, E., and Ishikawa, H. (2016). 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).

Zhang, R., Isola, P., and Efros, A. A. (2016). Colorful image colorization. *CoRR*, abs/1603.08511.