International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 6 June 2022, pp: 1594-1598 www.ijaem.net ISSN: 2395-5252

Image and Video Colorization

Yash Dubey, Prince, Zubair Saifi
Department Of Computer Science and Engineering
HMR Institute of Technology and Management, Delhi 110036, India

Submitted: 05-06-2022 Revised: 17-06-2022 Accepted: 20-06-2022

ABSTRACT: Techniques of using convolutional neural networks (CNNs) to colorize monochrome video have been widely researched. However, the results of automatic colorization are often different from the user's intentions and historical fact. A lot of color correction work still needs to be done in order to produce a colorized video. This is a major problem in situations such as broadcasting production where footage must be appropriately colorized in accordance with historical fact. In this we propose a practical video colorization framework that can easily reflect the user's intentions. The proposed framework uses a combination of two CNNs-a user-guided still-video-colorization CNN and a color-propagation CNN-that allows the correction work to be performed efficiently. The user-guided still-image-colorization CNN produces key frames by colorizing several monochrome frames from the target video on the basis of user-specified colors and color-boundary information. The color-propagation CNN automatically colorizes the entire video on the basis of the key frames, while suppressing discontinuous changes in color between frames. A quantitative evaluation showed that it is possible to produce color video reflecting the user's intention with less effort than with earlier methods.

Keywords :convolutional neural network, open CV, python, visual studio code

I. INTRODUCTION

1.1. Background

In this project we try to tackle the problem of colorizing grayscale images automatically using convolutional neural networks. The input to our project is a grayscale image. We then use a convolutional network to generate a 'realistic' colorization of the image.

Colorization of grayscale images may be a walk in the park for the human imagination but for a computer, it is a humongous task. A person only needs to recall that the sky is blue, the grass is green or can imagine several plausible colors for other objects. Automatic colorizations requires a very high comprehension thus making a very challenging algorithm. Colorization may have been reserved for those with artistic talent in the past, but thanks to AI, now it is possible to colorize images with outstanding quality with no human input or feedback. Colorizing gray-scale images can have a big impact in a wide variety of domains, for instance, re-mastering historical images and improvement of surveillance feeds. However, designing and implementing an effective and reliable system that automates this process remains nowadays as a challenging task.

1.2 Convolutional Neural Network

It is a very challenging task to fill colors in a grayscale image using human imagination. Humans can associate a certain color with the semantic content of the image but this recognition of content and then classifying a color is an

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 6 June 2022, pp: 1594-1598 www.ijaem.net ISSN: 2395-5252

arduous and laborious task. However, convolutional neural networks (CNN) are very capable at quick image classification and recognition with error rates below 5% on ImageNet challenge. Hence, the convolution neural networksare a natural choice to explore for the task at hand.

1.3 RELATED WORK

Our project was inspired by Ryan Dahl's [4] CNN based system for colorizing grayscale images. Dahl was one of the first to explore the possibility of using a convolution neural network for image classification to produce full colored channels for the input image. His system relies on several trained layer of ImageNet which integrates into a system with residual connections that merge the intermediate encoded output of network having VGG16 layers with the output produced by the other portion of the network [3]. The ResNet system built by He et al who won ImageNet challenge 2015 inspired the residual connections in network [5]. The residual connections connect the network edges thus reducing the training convergence time and train deeper networks more reliably.

In terms of result, Dahl's approach generated mixed results. While the predicted colors were almost always reasonable, however numerous images are sepia colored and desaturated, owing to the "average effect" of L2 loss. Recent works approaches this problem either as regression on continuous color space [4, 6, 7] or classification of quantized colorvalues[9]. More recent work in loss function has made it reliable to capture the multimodal nature of the problem. For instance, Hwang et al. [1] use a tailored loss function which predicts a distribution of possible colors for each pixel, and re-weight the loss at training time to emphasize rare colors ("class rebalancing"). Larsson et al. [8] and Iizuka et al. [10] have developed similar systems with subtle changes in the architecture and methods. While Larsson et al. use an un-rebalanced classification loss, Hwang et al. [1] use a classification loss, with rebalanced rare classes, and Iizuka et al. use a regression loss. Larsson et al. use hyper-columns
[11] on a VGG network [12], Iizuka et al. use a two- stream architecture in which they fuse global and local features. The three papers all produce competitive results with pros and cons for each method.

II. METHODOLOGY

2.1 Colorization

The problem of colorization can be tackled with either RGB color space or CIE color space.
A RGB color space is any additive color space based on the RGB color model. A specific RGB color space is characterized by the three chromaticity of the red, green, and blue added substance primaries, and can deliver any chromaticity that is the triangle characterized by those essential colors. In RGB color space the images can be recognized as a grid of pixels having a range of 0 representing black and 255 indicating white.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 6 June 2022, pp: 1594-1598 www.ijaem.net ISSN: 2395-5252

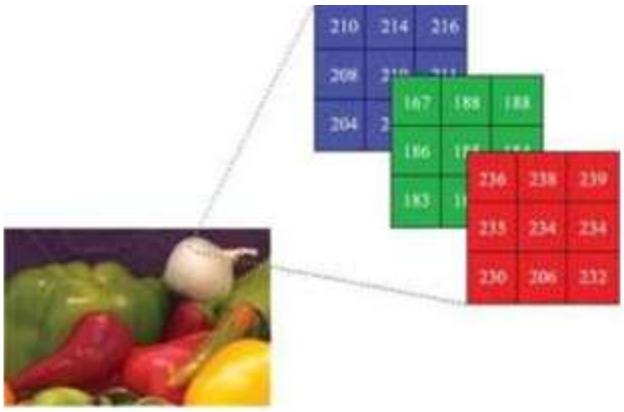


Fig: Representing pixel values for the part of image in RGB color space

As shown in the figure, in RGB color space equal distribution of all colors is required to produce a white color. A brighter color can be achieved by adding equal amounts of RGB colors. Thus, color and contrast can be manipulated using the RGB layers.

Let us now explore the CIE Lab color space of colorization problem. The similarity between RGB and CIE color space is that they both are 3-channel colored spaces, but unlike the RGB color space, The CIE color space comprises only of 2 channels namely 'a'(red-green channel) and 'b'(blue-yellow channel) to store the encoded color information and the 'L'(lightness) channel stores the information pertaining to intensity encoding. As depicted in the figure below, the entire a-b space is quantized into 313 bins. This quantization makes the calculations easier i.e., we will effortlessly search for a bin number confined in range [0,312] instead of finding the a and b values for each pixel. The greyscale image already has an 'L' value which lies in range [0-255] and the value of a-b channel rangingbetween [0-313] needs to be known. Now the color prediction task can be viewed as selecting a bin from 313 classes for every grey pixel which already comprises of its 'L' value thus transforming it into a monic polynomial classification problem.

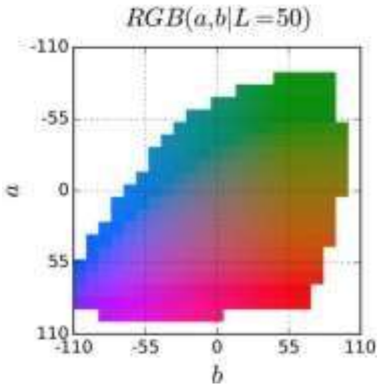
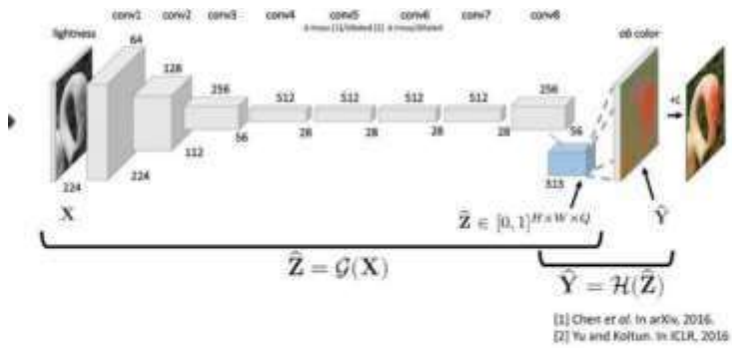


Fig. Quantized colors in ab space

1.4 Architecture

The CNN architecture resembles a VGG-styled network with multiple convolutional blocks which is proposed by Zhang et al. Every block comprises of 2 or 3 layers, the first part has 2 or 3 internal convolutional layers, second layer constitutes of Rectified Linear Unit (ReLU) and the third layer is a Batch Normalization layer. This architecture lacks a pooling layer.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 6 June 2022, pp: 1594-1598 www.ijaem.net ISSN: 2395-5252



The input image 'X' is rescaled to 224x224.' X' is passed through the neural network and gets transformed into 'Z^'. The following transformation can be depicted with 'G' and can be mathematically formulated as:

$Z\hat{} = G(X)$

$H\times W\times Q$ is the dimensions of Z^, where H (=56) and W(=56) renders the height and width of the output that is generated in the last convolution layer. Z comprises of a vector Q (=313) which represents the value for each of the $H\times W$ pixels. Value of Q signifies the probability of the pixel being part of that class. We aim to find a single pair of ab channel values for every probability distribution $Z^{(h,w)}$.

1.5 Color recovery from Z^

Group of distributions in Z^ from the resized input image X is portrayed in the above figure. Now we are required to recover a single a-b value pair from each distribution in Z^.

We cannot simply take the mean of the distribution and choose the a-b pair corresponding to the nearest bin center as the distribution is not Gaussian and also the mean distribution simply pertains to an unnatural desaturated color.

The a-b pair correlate to the annealed-mean of the distribution Z^h,w is represented in Yh,w, which can be mathematically formulated as a transformation of the original distribution Z^h,w

$Y = H(Z\hat{})$

To obtain the colored image, it is unsampled to original image size and then incorporated with the intensity factor 'L', to produce final colored image.

1.6 Loss Function

The neural networks are trained by defining a loss function. The Purpose of training process is to topple the loss over training set. Training data comprises of thousands of images colored images and greyscale counterpart versions. The output of the CNN is Z^ given an input image X. There is a dire requirementtotransform all colored images in the training set to the matching Z values.

$Z = H^{-1}(Y)$

In an output image Y, for every pixel Y we can effortlessly find the nearest a-b bin and representsZh,w as a one vector, in which we assign '1' to the closest a-b bin and on contrary '0' to all remaining 312 bins . But for a comparatively better result, the 5-nearest neighbours are taken into consideration and a Gaussian distribution is taken into use to compute the distribution Zh,w depending on the distance from the ground truth.

$$L(\hat{Z}, Z) = -\frac{1}{HW}\sum_{h,w}\sum_{q} Z_{h,w,q}\log(\hat{Z}_{h,w,q})$$

The result very dull colors are produced when the above loss function is used which is the outcome of heavy distribution of colors surrounding greyscale.

$$L(\hat{Z}, Z) = -\frac{1}{HW}\sum_{h,w} v(Z_{h,w})\sum_{q} Z_{h,w,q}\log(\hat{Z}_{h,w,q})$$

V(.) represents the color rebalancing factor. This factor is used to rebalance the loss framed on the rarity of color class.

2.5 Rebalancing the color

To produce a vibrant denomination, loss function can be further altered as:

III. CONCLUSION

While video colorization could be a boutique computer graphics assignment, it is additionally an occasion of a troublesome pixel forecast issue in computer vision. Here we tried to

generate a model which can color a video and represent that colorization achieved by deep learning and convolutional neural network in junction with efficient objective function can produce results highly similar to the original colored videos. Although it is trained in coloring our model can perform very well on classification and detection problem
.

REFERENCES

[1]. Zhang, Richard, Phillip Isola, and Alexei A. Efros. "Colorful image colorization." European Conference on Computer Vision. Springer International Publishing, 2016.

[2]. Hwang, J. and Zhou, Y., 2016. Image colorization with deep convolutional neural networks. In Stanford University, Tech. Rep..

[3]. Larsson, G., Maire, M. and Shakhnarovich, G., 2016, October. Learning representations for automatic colorization. In European conference on computer vision (pp. 577-593). Springer, Cham.

[4]. Dahl, Ryan. "Automatic colorization." (2016)

[5]. K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scaleimagerecognition.arXiv preprintarXiv:1409.1556, 2014.

[6]. Cheng, Zezhou, Qingxiong Yang, and Bin Sheng. "Deep colorization." Proceedings of the IEEE International Conference on Computer Vision. 2015.

[7]. Deshpande, Aditya, Jason Rock, and David Forsyth. "Learning large-scale automatic image colorization." Proceedings of the IEEE International Conference on Computer Vision. 2015.

[8]. Larsson, Gustav, Michael Maire, and Gregory Shakhnarovich. "Learning representations for automatic colorization." European Conference on Computer Vision. Springer International Publishing, 2016

[9]. Charpiat, Guillaume, Matthias Hofmann, and Bernhard Schölkopf. "Automatic image colorizationvia multimodal predictions." Computer Vision– ECCV 2008 (2008): 126-139.

[10]. Iizuka, Satoshi, Edgar Simo-Serra, and Hiroshi 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 (TOG) 35.4 (2016):110.

[11]. Hariharan, Bharath, et al. "Hypercolumns for object segmentation and fine-grained localization." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

[12]. Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[13]. Sindhuja Kotala; Srividya Tirumalasetti; VudaruNemitha; Swapna Munigala "Automatic Colorization of Black and White Images using Deep Learning" IJSCN Journal Volume 8, Issue 2 (April 2019).

[14]. Agrawal, M., & Sawhney, K. (2016). Exploring convolutional neural networks for automatic image colorization. Stanford University.