

Automatic Colorization and Restoring of gray scale images using Deep Learning

A.S.Sahil Hussain, A.Vignesh, S.Meenatchi Sundaram, UG Scholar
A.Neela madheswari, Associate Professor, neela.madheswari@gmail.com
K.Rajeswari, Assistant Professor, rajeswarik@mahendra.info
Mahendra Engineering College, Namakkal, Tamilnadu, India

Abstract- Color plays an important role in our day-to-day life. If the colors are used proper, it can results in improved health zone, mind relief also for humans. When used in the right ways, color can save on energy consumption also. The main goal of this work is to bring new life to old photo and videos by colorizing them. The proposed deep neural network has a fusion layer that allows us to elegantly merge local information dependent on small image patches with global priors computed using the entire image, Furthermore, proposed architecture can process images of any resolution, unlike most existing approaches based on CNN. An existing large-scale scene database is used to train the proposed model to learn the global priors and classify the objects of image to be able to map color. It demonstrates method extensively on many different types of images, including black-and-white photography from over a hundred years ago, and show realistic colorizations. The colorization of our proposed approach is considered “natural” more than 80% for images.

Keywords: image colorization, convolutional neural network, deep learning, computer vision

1. INTRODUCTION

Image colorization is the process of assigning color values to gray scale or black and white images. There are a huge number of applications go beyond gray scale or black and white photography. Nowadays usage or applications of images and videos are very vital in every day to day activity. For example the night vision images or

videos captured from CCTVs or cameras are mostly in the form of black and white or very dim in nature. In order to find the objects in those images is also challenging. These kinds of videos or images play a very important role in the security analysis of any concern or government today. We have to extract a good picture in order to evaluate the exact information hidden behind the images. Cameras were first introduced during 1800s and produce images in only black and white colors. Till 1900s, we cannot get color images using photographs. Hence there is a large source of black and white images people want to see in colors [1].

2. COLORS AND ITS COMPONENTS

Color is considered as one of the most important information about image, as though there are other important information can be extracted from an image to define the difference between one image and another. Existing of colors is important and there are many color spaces and models used to represent the image information. There are various color components available. They are listed below.

- a) Brightness: The human sensation by which an area exhibits more or less light.
- b) Colorfulness: The human sensation according to which an area appears to be similar to one or proportions of two, of the perceived colors red, yellow, green and blue.
- c) Lightness: The sensation of an area's brightness relative to a reference white in the scene.
- d) Chroma: The colorfulness of an area relative to the brightness of a reference white.

- e) Saturation: The colorfulness of an area relative to its brightness.
- f) Luminance: It is a photometric measure of the luminous intensity per unit area of light. It describes the amount of light that passes through, is emitted or reflected from a particular area.

Chroma and luminance are considered as important information to represent image colors. Consider an example figure given in figure 1.



Figure 1: Luminance only, Chrominance only, full color image with Luminance and Chrominance

Color space is a method by which we can specify, create and visualize color. Human can define a color by its attributes of brightness, hue and colorfulness. A color is usually specified using red, green and blue emission. Different color spaces are better for different applications [2]. The Lab color space describes mathematically all perceivable colors in the three dimensions L for lightness and a and b for the color opponents green-red and blue-yellow [3].

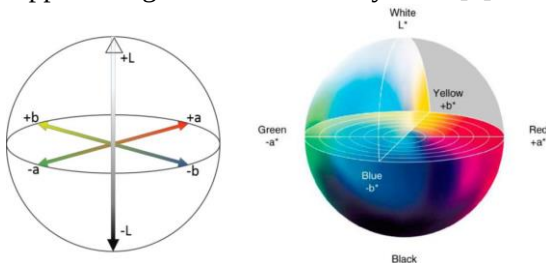


Figure 2: Representation of human recognizable colors in $L^*a^*b^*$ color space
The L^* coordinate of an object is the lightness intensity as measured on a scale from 0 to 100, where 0 represents black and 100 represents white. The a^* coordinate of an object represents the position of the object's color on a pure green and pure red

scale, where -127 represents pure green and +127 represents pure red. The b^* coordinate represents the position of the object's color on a pure blue and pure yellow scale, where -127 represents pure blue and +127 represents pure yellow. L^* component represent the luminance of pixel, while a^* and b^* represent the chrominance components of the same pixel. Image consists of pixels and each pixel had its own color. There are different color spaces and different way to represent the colors of images.

Features of images give us a clear and important information about images which help us to classify images and compare between them and their content to apply many operations such as finding the difference between two images, know the content of image, image matching, object detection and many other operations like image processing and computer vision field.

In computer vision field, features classified into three levels namely, i) low level feature, ii) middle level feature, and iii) high level feature. Consider an example picture given in figure 3.

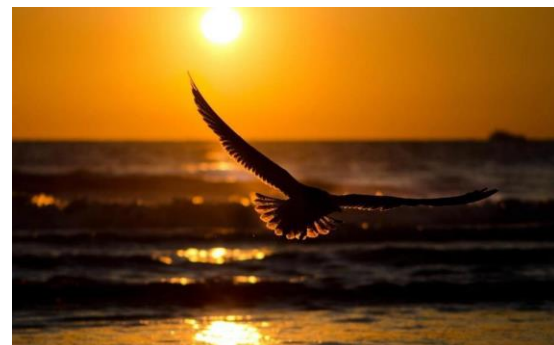


Figure 3: A flying bird with Sun and Sea background

The identification of the picture is 'flying bird with Sun and Sea at the background'. This identification is based on the whole scene. It is not possible to look separately the sun, bird, sea and then try to classify the image or to the edges inside the images and corners. Low level features describe the

corners and edges inside the images. Middle level features describe the whole objects inside the images. High level features describe the overall meaning of the image.

3. LITERATURE BACKGROUND

Colorization technique is divided into semi-automated technique and full-automated technique. Semi-automated technique is classified into two type namely: i) Scribble based, and ii) Color by example. Full-automated technique is classified into two types namely, i) Shallow neural network, and ii) deep learning colorization.

One of the more traditional approaches to image colorization consists of propagating colored user-specific scribbles to the whole image. Gray scale image normally requires human-labelled colors scribbled on the gray scale target image. The color information on the scribbles are then propagated to the rest of the target image. It is shown in figure 4 [4]. An optimization-based framework is proposed for colorizing a gray scale image. This was done by solving a quadratic cost function derived from differences of intensities between a pixel and its neighboring pixels [5].



Figure 4: Scribbled colors turn into colored image [5]

A fast colorization technique using chrominance blending based on weighted distances is proposed in [6].

Texture similarity for more effective color propagation is proposed in [7]. Texture classification has also been used for cartoon colorization [8]. A flexible colorization tool is proposed for hand-drawn cartoons based on a graph-cut-based optimization

framework that is applicable to various drawing style [9].

To enable long-range propagation for image re-colorization or tonal-editing, various affinity-based methods have also been proposed, such as global optimization with all-pair constraints [10], [11]. Manifold learning [12], and stochastic modeling of appearance similarities between user-specified pixels and other pixels [13]. However, these methods heavily depend on user input and require trial and error to obtain an acceptable result.

But scribble based technique had two demerits such as: i) heavily depend on user input, ii) require trial and error to obtain an acceptable result. Scribble-based methods utilize user-supplied colors whereas example-based colorization techniques exploit the colors of a reference image that are similar to the input image. For recoloring a color image, color transfer techniques [14] are widely used. These compute color statistics in both input and reference images and then establish mapping functions that map the color distribution of a reference image to the input image.

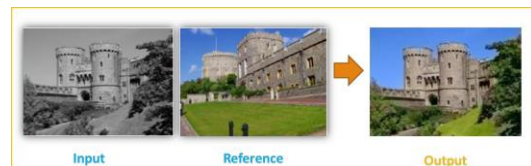


Figure 5: Colorize gray scale images

Example-based colorization technique had two demerits such as: i) require the user to supply similar reference images to the input image, which is time and effort consuming work, ii) performance highly depends on selected reference image. A fully automatic approach is proposed in which various features are extracted and the different patches of the image are colored using a small neural network [15].

Joint bilateral filtering is used to improve the results. Unlike our approach, they use very little training data which greatly limits

the type of images it is applicable. Furthermore, they require a high performance segmentation model to provide a segmentation of the image. Because of this dependency on segmentation, result is very poor for images in which none of the segmentation classes appear. This limits the application of the approach to simple outdoor scenes.

4. SYSTEM ARCHITECTURE

Convolutional neural networks are widely used tools for deep learning. They are suitable for images as inputs, although they are also used for other applications such as text, signals, and other continuous responses. They are made up of neurons that have learnable weights and biases. Each neuron receives some inputs, performs a dot product and optionally follows it with a non-linearity. The whole network still expresses a single differentiable score function: from the raw image pixels on one end to class scores at the other. And they still have a loss function on the last layer and all the tips/tricks we developed for learning regular Neural Networks still apply.

The proposed approach is based on Convolutional Neural Networks, which have a strong capacity for learning. The proposed architecture can jointly extract global and local features from an image and then fuse them together to perform the final colorization.

The proposed architecture consists of four main components: a low-level features network, a mid-level features network, a global features network, and a colorization network. Conceptually, these networks function as follows: First, a common set of shared low-level features are extracted from the image. Using these features, a set of global image features and mid-level image features are computed. Then, the midlevel and the global features are both fused by our proposed “fusion layer” and used as the input to a colorization network that outputs the final chrominance map. Needless to say, this is not explicitly implemented as a

sequential procedure; rather, it is realized as a single network. It is noted that no pre-processing nor post-processing is done: it is all computed in a single step. Additionally, as a side product of our approach, we can also perform classification of the scene. While the global features are computed using fixed-sized images, our novel approach for fusing the global and local features allows our model to be run on input images of arbitrary resolutions.

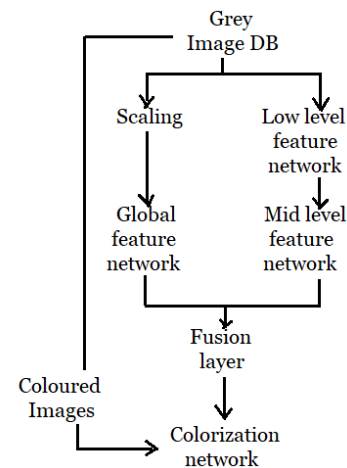


Figure 6: Overview model for Automatic colorization of gray scale images

4.1 Low level feature network

The network is consisting of 6-layer Convolutional Neural Network obtains low-level features directly from the input image. Instead of using max-pooling layers to reduce the size of the feature maps, convolution layers are used with increased strides. This is also important for increasing the spatial support of each layer.

4.2 Mid level feature network

The network is consisting of 2-layer Convolutional Neural Network which obtained by processing the low-level features further with two convolutional layers. The output is bottlenecked from the original 512-channel low-level features to 256-channel mid-level features. Note that unlike the global image features, the low-level and mid-level features networks are fully convolutional networks, such that the output is a scaled version of the input. In

particular the output of the mid-level features network is a volume of size $H/8 * W/8 * 256$, where H and W are the original image height and width respectively. This network is responsible for Extract mid-level features such textures and objects.

4.3 Global level feature network

The network consists of 4-layer convolutional neural network followed by three fully-connected layers. The global image features are obtained by further processing the low-level features. This results in a 256-dimensional vector representation of the image. Due to the nature of the linear layers in this network, it requires the input of the low-level features network to be of fixed size of $224 * 224$ pixels.

4.4 Fusion Layer

Fusion layer is a layer responsible for fusing both global and local features together. The global features act as an image prior on the local features to indicate what type of image the input is. For example, if the global features indicate that it is an indoor image, the local features will be biased to not attempt to add sky colors or grass colors to the image, but instead will add colors suitable for furniture. In order to improve the model efficiency, both networks use a number of common shared low-level features.

4.5 Colorization network

Once the features are fused, they are processed by a set of convolutions and up sampling layers, the latter which consist of simply up sampling the input by using the nearest neighbor technique so that the output is twice as wide and twice as tall.

These layers are alternated until the output is half the size of the original input. The output layer of the colorization network consists of a convolutional layer with a sigmoid transfer function that outputs the chrominance of the input gray scale image.

Finally, the computed chrominance is combined with the input intensity image to produce the resulting color image. In order to train the network, Mean Square Error (MSE) criterion is used. Given a color image for training, the image is converted to gray scale and CIE $L^*a^*b^*$ color space. The input of the model is the gray scale image while the target output is a^*b^* components of the CIE $L^*a^*b^*$ color space.

The a^*b^* components are globally normalized so they lie in the $[0; 1]$ range of the Sigmoid transfer function. We then scale the target output to the size of the output of the colorization network and compute the MSE between the output and target output as the loss.

5. EXPERIMENTAL RESULTS

Once the model has been structured, it is ready to be trained. We train our model by making it look at a large dataset that contains from gray scale images and their ground truth ones to be able to predict any gray scale image colors either encountered or not encountered during the training. It is assumed that CPU is enough for very simple models, but to train massive deep convolutional neural network model GPU is not a choice, it is a must.

The main computations in the deep convolutional neural networks are that the matrices multiplications, mostly of those multiplications operations can be massively parallelized, which implies speed up on GPUs. This is because GPUs were designed to handle those matrices operations in parallel.

So the mere reason why GPUs are well-suited to deep learning is because of the nature of computations they were designed to accelerate and handle happens to be the same as those encountered in deep learning. Thus, deep learning algorithms run several times faster on a GPU compared to a CPU, learning times can be reduced from months to weeks or even a day.

When the input image size is of a different resolution, while the low-level feature weights are shared, a rescaled image of size 224×224 must be used for the global features network. This requires processing both the original image and the rescaled image through model. In Test Model function where we put our testing model networks we used 2 low level networks and both uses the shared low-level feature weights but one of them takes the original gray scale image and feed the mid-level network with its output layers while the other network uses a rescaled image of size 224×224 and feed the global level network.

The dataset plays a vital role in the proposed approach. A large dataset with a huge number of categories is needed to make a good prediction. MIT Places dataset is used which consists of 2,448,872 training images and 20,500 validation images. In total, there are 205 classes corresponding to the types of the scene such as abbey, conference center, or volcano. Landscape category with 4000 images is used because huge number of images in training phase requires a huge number of resources and it is limited in resources in the proposed system.



Figure 7: MIT Places data set

The dataset is divided into a number patches and each one has a predefined number like 10 images. Our dataset contains gray scale images and their colored images which are used as labels to the gray scale because it is a supervised learning process. The result is shown in Table below with input image and output image.

Input image	Output image
-------------	--------------



6. CONCLUSION

Traditional colorization requires significant user interaction whether in the form of placing numerous color scribbles, looking at related images, or performing segmentation. A novel architecture is presented for the colorization of gray scale images by fusing both global and local feature. Our approach is based on CNN (convolutional neural networks) and can perform the colorization without any user intervention.

References:

1. Ross Daly, Tian Zhao, CNN assisted Colorization of Gray-scale images, 2015, pp.no. 4321-4327.
2. Quantization methodology of Hue, Saturation and Value, https://shodhganga.inflibnet.ac.in/bitstream/10603/185986/11/11_chapter%204.pdf, accessed on Jan 2019.
3. Lab Color space, <https://www.hisour.com/lab-color-space-24846/>, accessed on Jan 2019.
4. Satoshi Iizuka, Edgar Simo-Serra, Hiroshi Ishikawa, Let there be Color!: Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification, http://iizuka.cs.tsukuba.ac.jp/projects/colorization/data/colorization_sig2016.pdf, accessed on Jun 2018.
5. Anat Levin, Dani Lischinski, Yair Weiss, Colorization using Optimization, ACM Transactions on Graphics, 2004.
6. L.Yatziv, G.Sapiro, Fast image and video colorization using chrominance blending, IEEE Transactions on Image Processing, 15:5, 1120-1129, 2006.
7. Qing Luan, Fang Wen, Daniel Cohen-Or, Lin Liang, Ying-Qing Xu, Heung-Yeung Shum, Natural Image Colorization, Proceedings of the 18th Eurographics conference on Rendering Techniques, 309-320, 2007.
8. Yingge Qu, Tien-Tsin Wong, Pheng-Ann Heng, ACM Transactions on Graphics, 25(3), 1214-1220, 2006.
9. Daniel Sykora, John Dingliana, Steven Collins, LazyBrush: Flexible Painting Tool for Hand-drawn Cartoons, 28(3), EuroGraphics 2009.
10. X.An, F.Pellacini, APPROP: All-Pairs Appearance-Space Edit Propagation, ACM Transactions on Graphics SIGGRAPH 2008.
11. Kun Xu, Yong Li, Tao Ju, Shi-Min Hu, Tian-Qiang Liu, Efficient affinity-based edit propagation using K-D tree, ACM Transactions on Graphics, 28(5), 2009.
12. Du B & Zhang, Liangpei & Chen, Tao & Wu Ke, A discriminative manifold learning based dimension reduction method for hyperspectral classification, International Journal of Fuzzy Systems, 14, 272-277, 2012.
13. Seifollahi, Sattar and Dowd, Peter and Xu, Chaoshui, Fadakar Alghalandis, Younes, A Spatial Clustering Approach for Stochastic Fracture Network Modeling, Rock Mechanics and Rock Engineering, 2013.
14. Wu F, Dong W, Kong Y, Mei X, Paul J C, Zhang X, Content-based colour transfer, Computer Graphics Forum, 32(1), 2013.
15. Cheng Z, Yang Q, Sheng B, Deep Colorization, Proceedings of ICCV 2015, 29-43.