International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X
IMPACT FACTOR: 7.056

IJCSMC, Vol. 11, Issue. 1, January 2022, pg.116 – 121

Colourization of Black and White Images using Deep Learning

Guda Pranay Netha¹; M.S.S.Manohar²; M. Sai Amartya Maruth³

^{1,2,3}BE Scholars, Dept. of Electronics and Communications Engineering, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana

Dr. Ganjikunta Ganesh Kumar⁴

⁴(Mentor), Ph.D Associate Professor, Dept. of Electronics and Communications Engineering, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana

DOI: 10.47760/ijcsmc.2022.v11i01.014

Abstract: Colorization is the process of transforming grayscale photos into colour images that are aesthetically appealing. The basic objective is to persuade the spectator that the outcome is genuine. The majority of grayscale photographs that need to be colourized are of nature situations. Over the last 20 years, a broad range of colorization methods have been created, ranging from algorithmically simple but time- and energy-consuming procedures due to inescapable human participation to more difficult but also more automated ones. Automatic conversion has evolved into a difficult field that mixes machine learning and deep learning with art. The purpose of this study is to provide an overview and assessment of grayscale picture colorization methods and techniques used on natural photos. The study categorises existing colorization approaches, discusses the ideas underlying them, and highlights their benefits and drawbacks. Deep learning methods are given special consideration. The picture quality and processing time of relevant approaches are compared. Different measures are used to judge the quality of colour images. Because of the complexity of the human visual system, measuring the perceived quality of a colour image is difficult. Multiple metrics used to assess colorization systems provide results by calculating the difference between the expected colour value and the ground truth, which is not always consistent with image plausibility. According to the findings, user-guided neural networks are the most promising category for colorization since they successfully blend human participation with machine learning and neural network automation.

INTRODUCTION

Historical black-and-white images are regarded as irreplaceable, with exceptional artistic value. However, by looking at them it is impossible to fully imagine the actual scene, as color is a very important component of visual representation. The colorization of black-and-white images greatly alters the perspective of the viewer. The time gap between the past and the present fades away while making the scene more conceivable. However, insight into the authentic colors of early photographs is often unavailable, making satisfactory reconstruction difficult. Nevertheless, the aim of colorization is to deceive the viewer, to make him believe in the authenticity of the colorized image, and not to reconstruct the color accurately. Main

applications of colorization include the revival of historical black-and-white images, movie restoration and coloring astronomy photographs.

Color is a subjective response of the human visual system to electromagnetic radiation of the visible spectrum with wavelengths between 380 nm and 780 nm. It is a property of an object that can be described by hue, lightness, and saturation. Understanding color perception involves physics, physiology, and psychology. The perception of color depends on vision, light and individual interpretation.

The main aim of Artificial Intelligence is to make a robot think like a human. It can be accomplished by studying how a human brain thinks, and how humans learn, decide, and work when they are trying to solve a critical problem.

The goals used for creating AI are, first to create an expert system and then implementing human intelligence in machines. The contributions to AI are in various fields, such as, maths, biology, neuroscience, psychology, computer science, philosophy, sociology, etc.,. AI knowledge has few unwelcomed properties in the real world. They are, it has a huge volume which is next to unimaginable, knowledge is not well-organized and the knowledge constantly keeps changing. To use this knowledge efficiently and organise, we use AI techniques. To correct errors it is ought to be easily modifiable. Though the techniques are incomplete or inaccurate, they should be functional in numerous situations. The complex programs equipped with AI need to be executed at an elevated speed by using AI techniques.

The RGB colour space is made up of three channels. CIE Lab colour space is comparable to RGB colour space, with the exception that colour information is encoded solely in the CIE Lab colour space. Channels "a" and "b" Only the L (lightness) channel encodes information. the intensity, so we can use it as our grayscale input to the a network of neurons. Ab channels will be predicted by the trained network. Now we'll merge the ab channels we've created with L. channel. Finally, we'll revert the "Lab" image to the original RGB.

RELATED WORK

Ryan Dahl's CNN-based system for colourizing images automatically. His system is based on several layers of ImageNet-trained VGG16 layers that will be integrated with a system that is an auto encoder with residual connections that aid in merging intermediate outputs generated by the encoding part of the network, which includes the VGG16 layers, with those generated by the latter decoding part of the network

Early digital colorization attempts serve as inspiration for scribble-based methods. Technicians used to select the most convenient colours, true or fictitious, for each object in an image by hand. Despite being complex and time-consuming, manual image segmentation procedures were also used. As in a colouring book, each region was assigned a suitable colour. Art connoisseurs were opposed to the colorization process due to concerns about disrupting artistic expression, which, along with financial and time costs, were the main reasons for the failure of colorization attempts. The advancement of computing power and the development of algorithms has rekindled interest in colorization, particularly in the computer vision research community.

The drawbacks of scribble-based methods can be mitigated by using a similar reference colour image for colorization. The procedure works by transferring colour information from a colour source image to the matching regions of a grayscale target image. This procedure reduces, but does not eliminate, human involvement in the colorization process.

The advancement of Internet search engines, as well as the appearance of centralised and indexable image databases, enabled the transfer of the general "atmosphere" between images. The goal is to locate pixels with corresponding luminance values in both target and reference images using statistical information from neighbouring pixels and similar texture recognition. Although local information is important for properly separating the boundaries of objects within an image, using global information greatly increases the likelihood of accurate colour assignment.

LITERATURE REVIEW

Grayscale photos were coloured in this work using several deep learning algorithms. A model based on a neural network is suggested. This model is built from the ground up, and many high-quality characteristics are extracted. There is a pre-trained model called "Inception-ResNet-v2." This study presents a specific encoder-decoder paradigm that can function on pictures of any size and ratio.

Following the calculation of the findings, the model is "publicly accepted." A separate user research is being conducted for this purpose.

Then a second menu of programmes that contains various types photographs are provided. As a result, the actions taken in this model may be stated as follows:

- High-level features are extracted using the pre-trained model specified above.
- Analysis of the architecture is carried out using CNN.
- A separate user study is carried out to check whether that model is publicly accepted or not.
- A separate set of old pictures are presented and the model is tested on that.

So, the major components are as follows: an encoder for encoding, a feature extraction engine, and a function for activation, which will be hyperbolic tangentially. This project is appropriate for carrying out certain tasks related to image colorization. It has the ability to colour components such as oceans, forests, and skies.[1]

Without any supervision, a generic framework is introduced in this research. For large picture datasets, an online method has been suggested. This type of training is less efficient than any other supervised method. End-to-end training happens, and this project seeks to learn discriminative features with relatively few assumptive characteristics, making it very easy to train. Because a different mean square function was utilised, the model may train millions of photos. As a loss function, a SoftMax function was employed. The quality of characteristics has been evaluated after various trials. The ImageNet database was utilised, and object categorization and detection were performed.[2]

Traditionally, image colouring is done by hand utilising scribbling methods. An automated technique is proposed in this research. The neural network's two unique convolutional architectures are compared and trained using various lossy functions. Each variant's outcome is obtained in the form of images or videos, which are then compared. The primary purpose of this article is to discover whether or not neural networks can be used to colourize grayscale photographs in an automated manner. The images would differ from natural photos in that they would have less textural content, making the process of acquiring information more difficult. So, for producing these many neural network variations and then comparing performance.[3]

PROBLEM DESCRIPTION

Colorization is essentially the process of assuming colour information in places where it does not exist. Technically, assigning three-dimensional RGB (Red, Green, Blue) colour information to each pixel with respect to intensity of a grayscale image in a visually acceptable, plausible manner is a difficult process. To reduce the task's complexity, the colorization process employs a conversion to a convenient luminance-chrominance colour space. Color spaces derived from RGB include YUV and CIELAB. CIELAB is a perceptually uniform colour space created by nonlinear transformation of RGB. The uniform changes in CIELAB components correspond to the uniform changes in human colour experience.

Fig:LAB individual channel images

As a result, the Euclidean distance between the corresponding points in the colour space can be used to approximate observing two distinct colours in CIELAB. YUV is a perceptually uniform colour that is derived from RGB via linear transformations. YUV and CIELAB both separate the luminance component from the colour information, allowing for the use of intensity information and easier prediction of the two remaining colour channels. The luminance component in YUV is represented by the Y component, while the chrominance components are represented by the U and V components. L is the luminance component in CIELAB, while ab components carry colour information - a represents the green-red axis, while b represents the blue-yellow axis. Different colorization methods are compatible with various colour spaces. While some writers investigate the impact of multiple colour spaces on the colorization process, many just pick the most practical one and create the approach using that colour space. In most circumstances, there is no distinct colour that can be assigned to a certain grey item (e.g., balloons, clothing, plastic objects, etc.). As a result, the existence of numerous items in the world that appear in a wide range of hues makes colorization an unsolved challenge (no unique solution exists). This intricacy keeps the research community interested and makes colorization an attractive subject.

OBJECTIVE

The main objective of the proposed model is to colourize the black and white images based on their intensity values and giving out RGB intensity values which forms colour images and videos with high efficiency and accuracy using deep learning models like auto encoders and convolution neural networks and deployed with lower computational cost.

PROPOSED WORK

In this method, we construct a deep neural network that accepts a grayscale picture as input and outputs a coloured image. We begin by converting our black and white image to 256 x 256 pixels. This is sent into our neural network as an input. By training on vivid photographs, our model has been educated to create shots with realistic hues. The RGB colour space has three channels. CIE Lab colour space is similar to RGB colour space, however the colour information is stored exclusively in the "a" and "b" channels. Since the L (lightness) channel simply represents intensity, we may utilise it as the neural network's grayscale input. Ab channels will be predicted by the trained network.

Now we will merge the ab channels that were created with the L channel. Finally, we will return the "Lab" image to the RGB colour space.

ARCHITECTURE

The architecture of the model used consists of pre-processing unit, encoder, feature extractor, latent space and a decoder.

A greyscale input is given to the model for A B channel intensity values which are to be predicted and given as output which then are merged to produce the final output colour image.

TRAINING THE MODEL

For a Model to be predicting the output values we need to train the model. Firstly to train a model we need an image dataset for this purpose. Image sets consisting of black and white images and their coloured counterparts are required to train the model. We have chosen, ImageNet data set. All the RGB color images are converted into LAB color space, the true purpose being to create an input-output training set for the prediction of model. The L channel is provided as an input And the A B channel are categorized as output.

PREPROCESSING

All the images are converted into 256x256 pixel size and the input pixels are scaled dividing by the maximum intensity value 256. This is done ensure that all the intensity values are in the range of -1 to 1, which helps to avoid underfitting.

ENCODER

The input given is (N x M) black and white image and its processed into (N/8 x M/8 x 512). In this process, 10 convolutional layers are used with (3 x 3) kernels. To preserve the input size of the layers, padding is used. The 1st, 3rd and 5th layers have stride equal to 2. This causes the output dimensions to be halved and therefore reduces the required computations.

FEATURE EXTRACTOR

For this we are using a pre trained inception model. Firstly, we scale the image to (299 x 299) and then we stack it with itself to produce a 2-channel image. Then we feed this into the network and just before the SoftMax function, we extract the output. The resultant is the (1001 x 1 x 1) embedding.

FUSION

The feature vector is replicated (MN/82) times and is attached along the depth axis to the feature volume which is the output of encoder. So, a single volume of shape (M/8 x N/8 x 1257) is obtained. Finally, an image of (M/8 x N/8 x 256) dimensions is obtained after applying 256 convolutional layers of (1 x 1) size.

DECODER

The input to the decoder is (M/8 x N/8 x 256) image. It is passed through a series of up-sampling and convolutional layers. It outputs a layer of size (M x N x 2).

Layer (type)	Output Shape	Param #
conv2d_87 (Conv2D)	(None, 128, 128, 64)	640
conv2d_88 (Conv2D)	(None, 128, 128, 128)	73856
conv2d_89 (Conv2D)	(None, 64, 64, 128)	147584
conv2d_90 (Conv2D)	(None, 64, 64, 256)	295168
conv2d_91 (Conv2D)	(None, 32, 32, 256)	590080
conv2d_92 (Conv2D)	(None, 32, 32, 512)	1180160
conv2d_93 (Conv2D)	(None, 32, 32, 512)	2359808
conv2d_94 (Conv2D)	(None, 32, 32, 256)	1179904
conv2d_95 (Conv2D)	(None, 32, 32, 128)	295040
up_sampling2d_27 (UpSampling)	(None, 64, 64, 128)	0
conv2d_96 (Conv2D)	(None, 64, 64, 64)	73792
up_sampling2d_28 (UpSampling)	(None, 128, 128, 64)	0
conv2d_97 (Conv2D)	(None, 128, 128, 32)	18464
conv2d_98 (Conv2D)	(None, 128, 128, 16)	4624
conv2d_99 (Conv2D)	(None, 128, 128, 2)	290
up_sampling2d_29 (UpSampling)	(None, 256, 256, 2)	0
Total params: 6,219,410		
Trainable params: 6,219,410		
Non-trainable params: 0		

Fig1:ARCHITECTURE OVERVIEW

OBJECTIVE FUNCTION

The model's optimal parameters are determined by minimising the objective function, which is specified over the target and estimated output. We compute the mean square error for this purpose between the genuine values of the ab component's pixel colours as well as its estimated value.

$$\text{Cost function: } C(w, b) = \frac{1}{2n} \sum_x \|y(x) - a\|^2$$

Fig:objective function

EXPERIMENTS

For the most part, the ImageNet database is used in the training process. The database contains millions of photos organised into several sets. We trained our model using 8 GB of photos. The ImageNet database's images include a variety of shapes. As a result, for encoding and inception, all pictures are rescaled to (224 x 224) and (299 x 299), respectively. A pretrained inception model was used for feature extraction. The course lasted about 6 hours. The Nvidia GeForce 1650Ti GPU was employed to accelerate the procedure.

RESULT

Following the training of our model, we attempted to colourize several black and white photographs. Nature components such as rivers, trees, grass, and so on are well coloured, however not all items are. Our model generated the most likely colours for the given objects.

Fig:outputs

CONCLUSION AND FUTURE WORK

In this research, we demonstrated an efficient method of colouring pictures using Deep CNN, as opposed to the conventional manual process. The goal of this study is to create an output image that is similar to but not identical to the input image. To avoid overfitting, several changes such as picture scaling and flipping were used. The model is used to extract high-level characteristics.

Colorization of old movies will be one of our future projects. Using this method, the old documentaries will appear more aesthetically appealing. Overall, some human interaction is necessary in the picture colorization approach, but it still has a bright future.

REFERENCES

- [1]. Federico Baldassare, Diego Gonzalez Morn and Lucas Rodes-Guirao. "Deep Koalarization: Image Colorization using CNNs and Inception-Resnet-v2". arXiv:1712.03400
- [2]. Kreahtbuhl, P, Doersch, C, Donahue, J, Darrell, Data-dependent initializations of convolutional neural networks. ICOLR (016).
- [3]. David Futschik, "Colorization of black-and-white images using deep neural networks", January 2018.
- [4]. Conference on Computer Vision. Springer, Cham, 2016. LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." nature 521.7553 (2015): 436.
- [5]. T. Welsh, M. Ashikhmin, and K. Mueller, "Transferring color to greyscale images," Proc. 29th Annu. Conf. Comput. Graph. Interact. Tech. SIGGRAPH '02, pp. 277–280, 2002.