International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

Automatic Image Colorization using Deep Learning

Abhishek Pandey, Rohit Sahay, C. Jayavarthini

Abstract: ***Image colorization is a fascinating topic and has become an area of research in the recent years. In this project, we are going to colorize black and white images with the help of Deep Learning techniques. Some previous approaches required human involvement or resulted in the development of desaturated images. We are building a Deep Convolutional Neural Network (CNN) which will be trained on over a million images. The output generated by the model is fully dependent on the images it has been trained from and requires no human help. The images are taken from different sources like ResNet, Reddit, etc. The model will include many hidden layers to make the output more accurate. This will be a fully automatic model and will produce images with accurate colors and contrast. Finally, the goal of this project is to produce realistic and color accurate images that can easily fool the viewer. The viewer wouldn't be able to differentiate between the photo which the model produced and the real photo. Our project has wide practical applications like historical image/video restoration, image enhancement for better interpretability, frame by frame colorization of black and white documentaries, etc.***

Keywords: CNN, CSV, DL, ML, GAN

I. INTRODUCTION

The colorization of black and white images can impact a huge variety of domains. Some of the applications of black and white image colorization are remastering of historical images and surveillance feed improvement. The content of black and white images is very limited. So, if we add color components, we can improve the semantics of the image. Prebuilt models like Inception and ResNet are trained using datasets of colored images. When we apply these neural networks on black and white images, we can improve the results if we colorize them beforehand. How-ever it is very challenging nowadays to design and implement a system that is both active and reliable to automate the whole colorization process. In this approach, we build a deep convolutional neural network that takes a grayscale image as an input and produces a colorized image. Firstly, we convert our black and white image in 256 x 256 pixels. We give this as an input to our neural network. Our model is trained to produce photos with realistic colors by training on colorful images. The images produced would easily fool a viewer.

The RGB color space is a 3-channel color space. CIE Lab color space is similar to RGB color space but the only difference is that the color information is encoded only in the "a" and "b" channels. The L (lightness) channel only encodes the intensity, so we can use it as our grayscale input to the neural network. The trained network will predict ab channels. Now we will combine the produced ab channels with L channel. Finally, we will convert the "Lab" image back to the RGB color space.

II. LITERATURE REVIEW

In this paper gray-scale images have been colored using various deep learning approaches. A model is proposed which is based on a neural network. This model starts from scratch and various high-quality features are extracted. There is a pre-trained model by the name of "Inception-ResNet-v2". In this paper, a particular encoder-decoder model is present which can work on images that can be of any size and ratio. After calculating the results "public acceptance" of the model is carried out. A separate user study is carried out for this. Then a separate menu of applications that has different types of images is presented. So, the steps which are carried out in this model can be summarized 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 main components are: An encoder for encoding and a feature extraction engine, and then there is a function for activation, a function that will be hyperbolic tangentially.

This project is suitable for carrying out certain tasks for colorizing the images. It is able to color elements such as oceans, forest, and sky [1].

In this paper the styles and formats of colored images are mixed with contents of grayscale-images and then the result is obtained as colorized-grayscale pictures. CNN is used which will extract some colored information from a particular picture and then it is transferred to another picture.

Manuscript received on February 10, 2020.
Revised Manuscript received on February 20, 2020.
Manuscript published on March 30, 2020.
* Correspondence Author
Abhishek Pandey*, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, India. Email: apandey287@gmail.com
Rohit Sahay, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, India. Email: rohitsahay_rakesh@srmuniv.edu.in
Mrs. C. Jayavarthini, Faculty of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, India. Email: jayavarthini.c@ktr.srmuniv.ac.in

The approach used is that both the content-image and the style-image are passed into a CNN network that will be pre-trained and then the formats of content representation and styled representation will be extracted. Then the same will be done to a noisy picture. For optimization L-BFGS was used and then the parameters were properly tuned and images produced were much better than the method of using stochastic gradient descent [2].

In this model architecture is proposed which is based on neural networks. It will color black and white pictures without the use of any human-interference. Many network models, problem objectives are focused on. The final architecture will produce colored pictures that will be more useful and pleasing than the previously made base-line regression models. This system uses various datasets. There are several 1000s of pictures divided into 8 categories in the MIT CVCL Urban and natural scene categories dataset. About 411 pictures were experimented to check the reliability of the system. A pipeline is built by making the program read pictures of certain constrained dimensions and in red, green, blue color spaces. This pipeline consists of a neural network. This model also solved issues of image-inconsistency. The system can be made to learn to produce pictures that could be compared with real images [3].

Traditionally picture-colorization is done using scribbling methods that work manually. In this paper, an automated method is proposed. Two distinct convolutional architectures of the neural network are compared and trained using various lossy functions. Each variant result is obtained in the form of pictures, videos and then compared. The main goal of this paper is to determine whether any possibility is there to use neural networks for the colorization of grayscale images in an automated manner or not. The images would be different from natural images by containing less amount of textural material so making the process of obtaining information harder. So, for obtaining these several variants of neural networks and then performances compared. Two architectures are considered-one will be a traditional and plain network and another one will be inspired by a residual network that has not been put to use previously [4].

The aim of this paper is to make an output image a realistic picture like the input but not necessarily the same as the original. A neural network is explored first. Then the model is combined with a classifier named Inception ResNet V2 which has been trained using 1.2 million pictures in order to obtain a more realistic output. CNN has been used to color images. This model had advantages compared to earlier models which used mean squared errors. And that led to photos of which was desaturated. New models such as colorful image colorization help to encourage more bold pixel choices as compared to those which were more conservative. The dataset used is that of Unsplash which consists of 10,000 pictures, 95% of which is used in the training set and 2.5% in the development set and 2.5% in the test set. Various transformations such as image zooming and flipping were also performed to avoid overfitting. A very simple survey was done at the end to determine the frequency of colorings which were accurate but this approach proved to be too slow and blunt [5].

In this paper, the ImageNet database has been used which contains 7 million pictures. Using the concept of CNN density and diversity of datasets have been evaluated. The difference between object-centric and scene-centric networks has been shown. Linear SVM and pre-trained ImageNet database has been used. Visualization of CNN layers have been used to show object-centric and scene-centric networks. The Places database is very large it achieves the best performance when the whole of the set is used as a training set. The workers were presented with different sets of images and had to choose the set which is so similar. The deep features obtained from ImageNet were not enough competitive to perform the tasks. The Places dataset was 60 times larger than the SUN database [6].

In this paper, a generic framework has been introduced without any supervision. An online algorithm has been proposed for big image databases. This approach is less efficient for training rather any other supervised method. End to end training occurs and this project aims at learning discriminative features and very few assumptive features are there and so this is very easy to train. A separate mean square function has been used and so the model is used to train millions of images. A SoftMax function as a loss function has been used. After performing several experiments, the quality of features has been evaluated. ImageNet database has been used and object classification and detection have been done [7].

III. PROPOSED WORK

In this approach, we build a deep convolutional neural network that takes a grayscale image as an input and produces a colorized image. Firstly, we convert our black and white image in 256 x 256 pixels. We give this as an input to our neural network. Our model is trained to produce photos with realistic colors by training on colorful images. The images produced would easily fool a viewer.

The RGB color space is a 3-channel color space. CIE Lab color space is similar to RGB color space but the only difference is that the color information is encoded only in the "a" and "b" channels. The L (lightness) channel only encodes the intensity, so we can use it as our grayscale input to the neural network. The trained network will predict ab channels. Now we will combine the produced ab channels with L channel. Finally, we will convert the "Lab" image back to the RGB color space.

IV. ARCHITECTURE

On giving our model the l component of an image as an input, it calculates the ab components. It then combines it with the input to form the colored image. The architecture is proposed in the Fig. 1.

The CNN is divided into 4 parts. The encoder component produces mid-level features and the feature extraction component produces high level features. These are then merged into the fusion layer. Finally, the output is generated with the help of decoder component.

A. Preprocessing

The pixel size of the images is scaled between (-1,1) for correct learning.

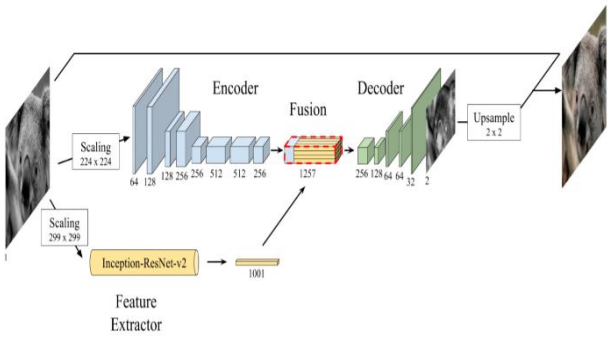

Fig. 1. Architecture

B. Encoder

The input given is (H x W) black and white image and its processed into (H/8 x W/8 x 512). In this process, 8 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.

C. 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.

D. Fusion

The feature vector is replicated (HW/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 (H/8 x H/8 x 1257) is obtained. Finally, an image of (H/8 x W/8 x 256) dimensions is obtained after applying 256 convolutional layers of (1 x 1) size.

E. Decoder

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

V. OBJECTIVE FUNCTION

Optimal values for the model are calculated by minimizing the objective function which is defined over the target and estimated output. For this, we calculate the mean square error between the real value of the pixel colors of the ab component and its estimated value. It is given by:

$$C(\mathbf{X},\boldsymbol{\theta}) = \frac{1}{2HW} \sum_{k\in\{a,b\}} \sum_{i=1}^{H} \sum_{j=1}^{W} (X_{k_{i,j}} - \tilde{X}_{k_{i,j}})^2,$$

Θ: Model Parameters.

$X_{k(i,j)}$: ij:th pixel value of the k:th component of the target.

$X\tilde{}_{(ki,j)}$: ij:th pixel value of the k:th component of the reconstructed image.

Adam optimizer is used while training so that the loss is back gets backpropagated and the model parameters gets updated.

VI. EXPERIMENTS

ImageNet database is used for the most part of the training process. The database consists of images which are in millions and they come in different sets. We have trained our model on 18 gigabytes of the images. The shape of the pictures of the ImageNet database are heterogeneous. So, all the images are rescaled to (224 x 224) and (299 x 299) for encoding and inception respectively. The training time was around 8hrs. Nvidia GeForce 1050Ti GPU was used for speeding up the process.

VII. RESULT

After training our model, we tried colorizing some black and white images. The nature elements like rivers, trees, grass, etc. are colorized well but some of the objects are not always. For those objects, our model has produced next probable colors.

We have compared our results with Zhang's who has used the same training set of images. We both have used different loss functions. We observed that although the results were good most of the time but some of the results were low saturated because of less diverse data set.

VIII. CONCLUSION AND FUTURE WORK

In this project, we have presented an efficient way of coloring images using Deep CNN unlike the older manual procedure. The aim of this paper is to make an output image a realistic picture like the input but not necessarily the same as the original. Various transformations such as image zooming and flipping were also performed to avoid overfitting. High-level features are extracted using the model.

Our future work will include the colorization of historical videos. This technique will cause the old documentaries look visually appealing.
Altogether, some human intervention is required in image colorization technique but still it has a great future potential.

REFERENCES

1. Federico Baldassare, Diego Gonzalez Morn and Lucas Rodes-Guirao. "Deep Koalarization: Image Colorization using CNNs and Inception-Resnet-v2". arXiv:1712.03400, 2017.
2. Tung Nguyen, Kazuki Mori and Ruck Thawonmas, "Image Colorization Using a Deep Convolutional Neural Network". arXiv:1604.07904, April 2016.
3. Jeff Hwang and You Zhou, "Image Colorization with Deep Convolutional Neural Networks", Stanford University.
4. David Futschik, "Colorization of black-and-white images using deep neural networks", January 2018.
5. Alex Avery and Dhruv Amin, "Image Colorization".CS230 – Winter 2018.
6. Zhou, B. Lapedriza, A., Xiao, J., Torralba, A., Oliva, A.: Learning deep features for scene recognition using places database.
7. Kreahenbuhl, P, Doersch, C, Donahue, J, Darrell, Data-dependent initializations of convolutional neural networks. ICOLR (2016).

AUTHORS PROFILE

Abhishek Pandey, I am a B. Tech. Computer Science and Engineering student from SRM Institute of Science and Technology. I am currently in 8th semester. I have scored 91.2% in class 10th and 87% in class 12th. I have done internships in the field of web development and data science. I have a keen interest in digital image processing and deep learning.

Rohit Sahay, I am a B. Tech. Computer Science and Engineering student from SRM Institute of Science and Technology. I am currently in 8th semester. I have scored 87.4% in class 10th and 85% in class 12th. I have done internships in the field of app development, machine learning and deep learning. I have a keen interest in integrating app development and deep learning.

Mrs. C. Jayavarthini , I am an Assistant Professor in the faculty of Computer Science and Engineering SRM Institute of Science and Technology. My areas of interest are Image processing, Biometrics, Data Mining and Image Annotation. I have published more than 7 research papers in many international journals. I am a member of IET and ISCA.