Image Colourization

Kandru Sahitya
Dept.of Computer Science,AIE
Amrita School of Engineering

Shuktij Rajesh Katta
Dept.of Computer Science,AIE
Amrita School of Engineering

Sri Vishnu Vallabh
Dept.of Computer Science,AIE
Amrita School of Engineering

Vireen Chowdary Vesangi
Dept.of Computer Science,AIE
Amrita School of Engineering

***Abstract*—The colorization of grayscale pictures is a not well presented issue, with numerous right solutions. In this paper, we propose an adversal learning colorization approach combined with semantic data. A generative network is utilized to deduce the chromaticity of a given grayscale picture molded to semantic signs. This network is outlined in an adversal model that figures out how to colorize by joining perceptual and semantic comprehension of shading and class distributions. The model is prepared through a completely selfsupervised system. Subjective and quantitative results show the limit of the proposed method to colorize pictures in a practical manner acheiving efficient results.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

Colorization is the way towards adding conceivable shading data to monochrome photos or recordings. Presently, advanced colorization of black and white visual data is a vital task in regions so diverse as promoting and entertainment worlds, photography innovations or craftsman help. Although significant advancement has been accomplished in this field, image colorization stills stays a test. Colorization is a highly undetermined issue, requiring planning a real gray scale picture to a three dimensional color one, that has not a unique solution. Prior to the rise of deep learning methods, the best techniques depended on human intercession, typically through either user given shading jots or a shading reference picture. As of late, convolution neural network procedures have advantage from the huge measure of freely accessible color pictures to consequently realize what tones normally compare to the genuine items and its parts. In this paper, we propose a completely programmed start to finish antagonistic methodology called ImgLab. It consolidates the strength of generative adverse networks with semantic class distribution learning. Accordingly, ImgLab can perceptually colorize a gray scale picture from the semantic comprehension of the caught scene. To illustrate this, Fig. 1 shows vibrant and diverse colorization frequently achieved. Moreover, ImgLab shows variability by colorizing differently some objects belonging to the same category that may have several real colors, as for example, the birds in Fig. 1. The user-based perceptual ablation study show that the effect of the generative adversarial learning is key to obtain those vivid colorization.

Identify applicable funding agency here. If none, delete this.

A. *Problem Setting and Description*

The intention of the project is to Automate the colorization of black and white images. It has been subject to much research within the computer vision and machine learning communities. Beyond simply being fascinating from an aesthetics and artificial intelligence perspective, such capability has broad practical applications ranging from video restoration to image enhancement for improved interpretability.

So far the colorization techniques are either pre defined shades of colors which we give the input to the system and it is not feasible in all cases. In order to avoid that we automate the technique of color selection of an object in the image.The main goal of this project is to minimise the maximum error between our training set and testing set using gradient decent.This will probably result in very slight difference in the colorization between our given image and the actual image.

B. *Known Methods Of Implementation*

There are methods and procedures to colorize an image.Our project is inspired by a few research papers.

First is the Image Colorization with Deep Convolutional Neural Networks by Jeff Hwang and You Zhou.Their program reads images of size 224 x 224,pixel by pixel and the three known and common channels,the RBG channels.They have converted the images into CIELUV colour space where the I cahnnel is itself given as the input during the training time and the U and V channels are extracted as output.They initialised the parts of model with a model of VGG16 architecture instance that has been pretrained on the ImageNet Data Set.The model when executed it accepts 224 x 224 x 1 black and white image.Two arrays of each of dimension $224 \times 224 \times 1$,one corresponding to U and One corresponding to V of the CIELUV colour space.Now to form the CIELUV representation of the given Black and White Image, the three channels are concatenated.

Second is Let there be Color!: Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification by Satoshi Iizuka, Edgar Simo-Serra and Hiroshi Ishikawa.Their approach to the project is based on CNN,Which have strong capacity for learning.They have proposed a Novel Architecture which can jointly extract global and local features from an image and fuse them together to perform final colourization.

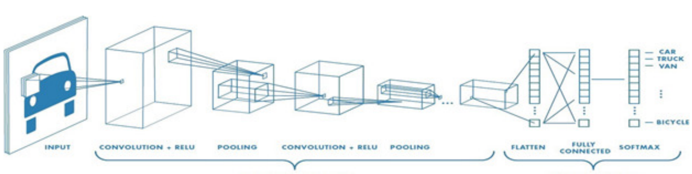
II. Theriotical Background

A. Convolution Neural Network

Convolution neural network is a deep learning algorithm which can take image as an input and assign importance or weights to each pixel based on the importance of the image and which can differentiate between each pixel. Many other algorithms cannot learn filters by themselves and need human assistance. Where as CNN are capable of self learning the filters.

The CNN is actually inspired by human brain and visual cortex of our brain and how process a given information and gives us a certain output or work which we execute.

- Convolution Layer: The layer or the matrix involved in the computing convolution operation is called the convolution layer. The filter moves on with a certain Stride Value till it parses the complete width. Moving on, it hops down to the beginning of the image with the same Stride Value and repeats the process until the entire image is traversed.
- Padding: Padding is the amount of the pixels increased or added to an image during its process of convolution layer. If the padding in a CNN is set to zero, then every pixel value that is added will be of value zero.
- Pooling Layer: The function of the pooling layer is to decrease the spatial space of a convolution layer. This will be helpful during the computation power required to process the image or the data.
- Types of Pooling:
  - Max pooling: It is an pooling operation in which selects the highest element from a region or a stride. The output of the max pooling is a feature map which has the predominant features of previous feature map.
  - Average pooling: Unlike Max pooling, Average pooling takes out the average patch of each feature map.

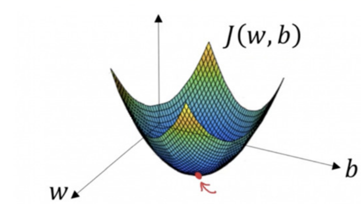


B. Gradient Descent

Gradient Descent is an optimization algorithm for finding the variables of a given function such that cost function is reduced as much as possible.

- Gradient:
  Gradient is the a measure of how much an output is varied in change in the input of that. In terms of mathematics gradient is the partial derivative with respect to inputs. Imagine hiking down a hill blindfolded. We take small steps at the top of the mountain and will begin taking larger steps during the least steep area. This process can be represented using gradient.
  b = a-$\gamma\delta f(a)$

The above equation gives how gradient descent works. is the next position and represents current position and the difference between and is the minimization part. Represents the waiting factor and represents the gradient term which means the steepest direction.
Imagine we have a machine learning problem and want to train our algorithm with gradient descent to minimize your cost-function J(w, b) and reach its local minimum by tweaking its parameters (w,b). The image below shows the horizontal axes represent the parameters (w, b), while the cost function J(w, b) is represented on the vertical axes. Gradient descent is a convex function.



III. Approach

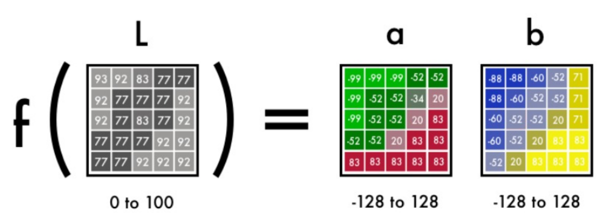
We build a learning pipeline that comprises of a nueral network and an image pre-processing front-end.

A. Our Goal

We will be having a grey scale image on which we are going to apply some computations and generate the red, green and blue channels from it . And this grey scale image is obtained when we separate the L,A and B channels from the RGB image(Color image).

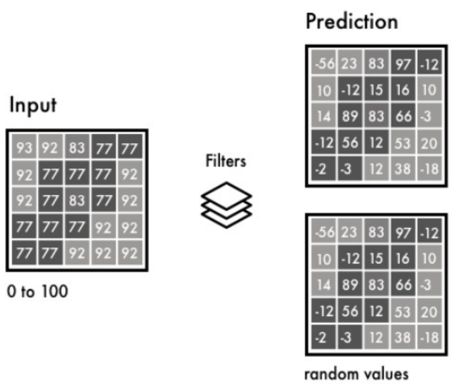
B. General Pipeline

During our training time,Our program reads,compares and predicts the output channels from CNN with the two channels corresponding to the red , green and blue, yellow color space that we have generarted from the RGB image. And it shows the error and backpropagates to decrease the error and finally it exports. The three channels are then Concatenated together to form the RGB image.



C. Separating the Channels using CNN

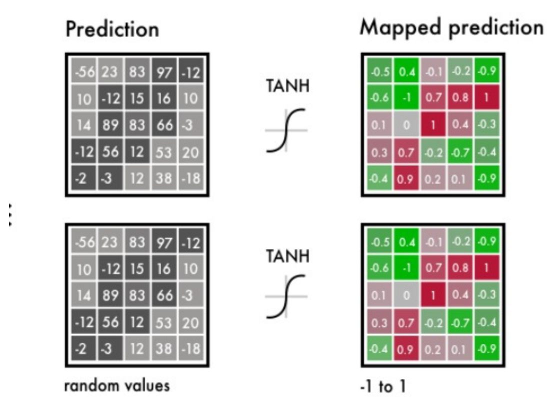
The L channel (Greyscale) image is fed to the Convolutional Neural Network to obtain two prediction channels Which are obtained by Using filters as they get applied to each channel and as we increase the number of filters more layers gets formed on the L channels hence extracting broader information from the images.



D. Activation Function

It is used to determine the output of neural network like yes or no. It maps the resulting values in between 0 to 1 or -1 to 1 . In our case the output from the CNN are the 2 predicted channels So this tanh function replaces the -1 with 1 in the predicted images.It is just llike normalization and it is done to keep the image numerically well behaved.

So the Baseline model comprises of all the above mentioned points.

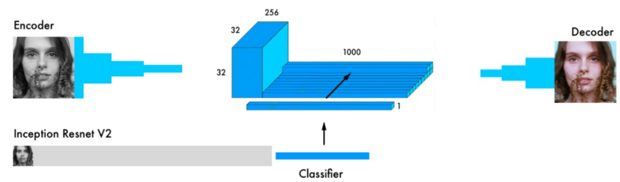


E. Final Model

We use a classifier in our final model because in our baseline model if we input images comprising of the walls the model gets used to it and if we change and give a different image the prediction can go wrong in order to increase the scope of prediction we used a classifier.

We have used a pre-trained classifier which has been run on 2.5M images and thousands of different objects. The classifier which we are using is Inception-Res-Net-v2 classifier.

The classifier takes the input from the encoder network and flattens the image and tries to obtain the context of the image like if it is a sky or a man then it also decides which colour has to be filled in the image after this the decoder will up sample to get the final filter and then it extracts it.



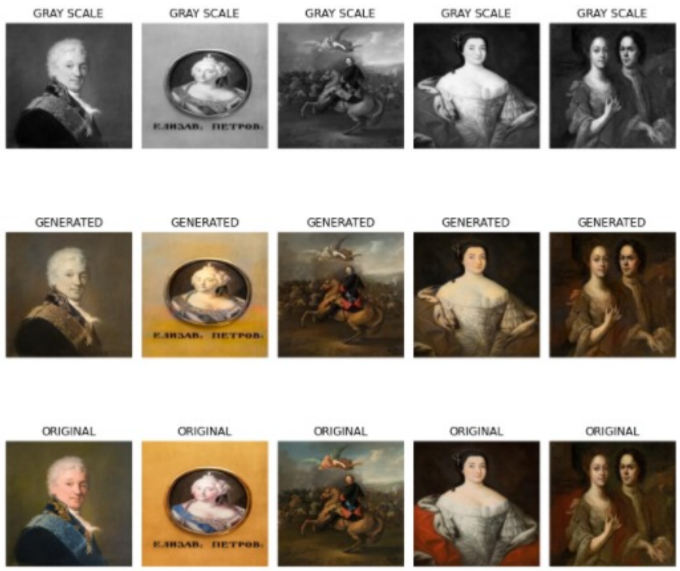
IV. Data Set

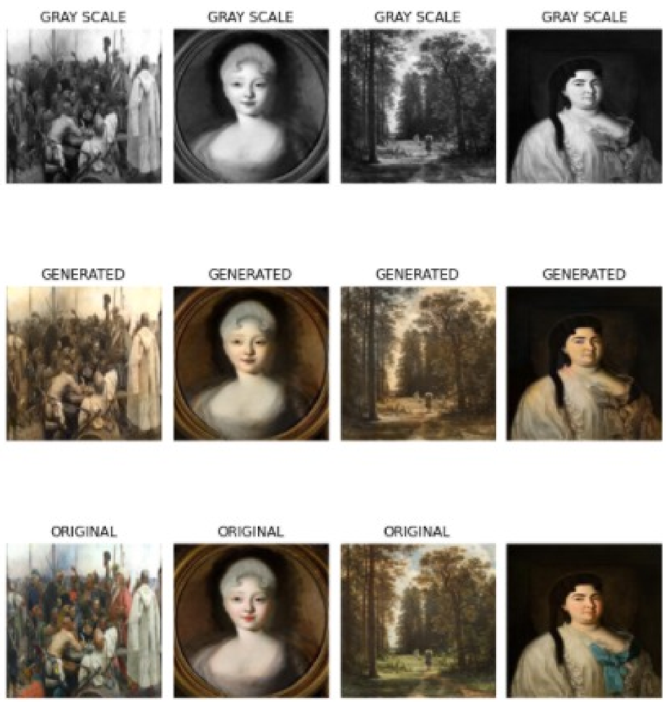
The Data set used in this project is a file named paintings, consisting 2000 images of old paintings.We have taken this data set from Kaggle Art Images:Drawing/Painting/Sculptures/Engravings.In this there are 9000 images consisting Drawings and watercolours,Works of painting,Sculpture,Graphic Art,Iconography (old Russian art). All these images are sorted in different folders according to their categories .Among those we have taken the painting data set because the data in the file are paintings of humans along with the background i.e., sky,the vegetation etc., The data set portrays the ability of the model so hence our data set contains images/paintings with different set of images,challenges the model. Even if we supply different images our model performance has not changed. so those details makes the predicted output more generalized towards the human understanding.After finding the correct data set we have to pre process the data i.e., we have to make the data set suitable for our project.We have to resize the image and we have to remove blank images and undefined (error) images.So we can ensure that our model is not learning faulty images.We also flip the image horizontally, shear the image by certain values, this is called augmentation,we perform this process to expand the details available in our data set.

V. Experimental Results and Discussion

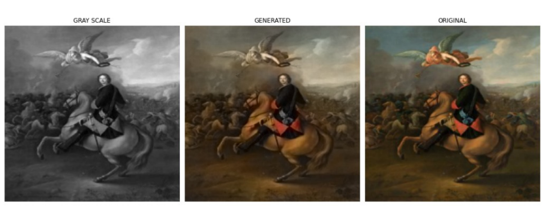
We have experimented our program on different data sets,one is which we have used for training and testing for our final model.One is the Historic colour-ECCV 2012, which contains the images of different sets corresponding to a particular year staring from 1930 to 1970.We have observer that the images are more likely leaning towards the brown colour because as our model is trained on the paintings data set.

In the Results section we just output the black and white image, the predicted image, and the ground truth, so that we can see the difference between all the three images and compare the prediction and ground truth.





The generated colourized images are accurate up to some extent.However one may find errors in predicting certain images but at the same time even if that error exists,one can understand the image and its context .



We represent the three pictures once again, but this time We output images in big size to understand the output and also can compare the images easily and find out the difference between the predicted and the original images.

Conclusion

We finally hereby conclude that we have used the Convolution Neural Network to solve our problem i.e., we colourized a black and white image. Our approach is based on convolutional neural networks and is able to perform the colorization without any user intervention.This model can be further trained in the future to get more and more accurate results.We can use Inception-resnet v2 classifier in order to increase the color accuracy of the image.

Acknowledgment

We would like to thank professor GopaKumar sir for his excellence in lecturing the topics throughout the semester.The assignments were very helpful for us which made easier for us to understand and implementation.

References

[1] Patricia Vitoria, Lara Raad and Coloma Ballester, “ChromaGAN: Adversarial Picture Colorization with Semantic Class Distribution,” University Pompeu Fabra, Barcelona, Spain.

[2] Frank Palermo1, James Hays2, and Alexei A. Efros1, Dating Historical Color Images, Carnegie Mellon University, Pittsburgh PA 15213, USA Brown University, Providence RI 02912, USA

[3] Jeff Hwang “Image Colorization with Deep Convolutional Neural Networks”.

[4] Satoshi lizuka,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”.ACM Transactions on Graphics july 2016 article 110