RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

CS536 Final Project

Colourization of Grayscale Images

Deepak Vagish K

dk1156

Sivakalyan S

ss4356

Raghav Ramkumar

rr1268

Shirisha Sudhakar Rao

ss4313

Darshan S

ds1992

April 28, 2023

Abstract

Automatic image colorization has been a popular research area in the past decade, with applications in restoring older grayscale photograph. However, the problem is highly ill-posed due to the many possible color assignments for each pixel. This research paper tackles the challenge of generating a color version of a grayscale photograph, without relying on significant user input. Previous approaches have struggled with this problem, resulting in desaturated colorizations or requiring a lot of user interaction. In contrast, our proposed fully automatic approach produces vibrant and realistic colorizations. To account for the inherent uncertainty in this task, we formulated it as a regression problem. We trained our system using subset of ImageNet data, and at test time, it uses a feed-forward pass in a VGG Network and GAN Network. Based on the results we have extended the paper from Image to Video.

1 Introduction

Vintage images which are in black and white holds its own charm and it's artistic value. It captures the soul of an image and evokes a feeling of dramatic intensity especially conveying the emotion of relaxation. On the other hand, coloured images has an appeal on its own as colour is most often the first thing that draws the eye in and naturally makes the viewer feel more empathetic since humans see the world in colour. Thereby colourization of a black and white image significantly alters the perspective of the user. The aim is to not accurately colour each part of the image but instead color it to deceive the user into believing its authenticity. This has been an active topic of research for the last few decades. In particular deep learning techniques have been giving state of the art results. The main reason for that is that colourization requires automatic learning of colours that naturally resemble real world objects. Adding to the same is the availability of a huge repository of data which is imperative in training a deep neural network. Over time, there came a requirement for better results which brought in Generative Adversarial Networks (GAN) in this space.

The role of the generator network is to produce results indistinguishable from real data. On the other hand the discriminator network classifies whether the result comes from a generator or the original data. During training, the networks - both of which are multilayer perceptrons, are pitted against each other where the generator tries to produce data that can fool the discriminator, while the discriminator tries to correctly classify the real and fake data. Among the many GAN architectures the fully connected GANs was initially used for simple datasets such as MNIST, CIFAR, etc. Later when moving towards applying the same to convolutional neural networks the results were giving poor accuracy compared to fully connected neural networks. It was deep convolutional GAN (DCGAN) which gave better results where a pair of deep convolutional generator and discriminator networks where utilized. It made use of strided and fractionally strided convolutions allowing spatial downsampling and upsampling to be learned during training.

2 Litreature Survey

Larsson et al. [4] used VGG to extract low dimensional features and semantic information from images. They improved the colourization effect by predicting the colour histogram of each pixel, but it still suffers from colour overflow. Zhang et al. [5] transformed the colouring problem into a regression problem. Their approach first extracted features from grayscale images, then it predicted the colour distribution of each pixel, and performed rebalancing operations on the predicted colour distribution according to the prior colour distribution. However, the method suffers from inconsistent colour regions. Su et al. [6] combined object detection with image colouring. The object is segmented by the object detection algorithm, and then the overall image and the object colour distribution are predicted, respectively. This method was greatly affected by the object detection algorithm and had poor colouring performance on small objects. Messaoud et al. [7] established a conditional random field-based variational auto-encoder algorithm to solve structural inconsistencies occurring in colouring tasks. It not only achieves diversity but also ensures the structural consistency of the results. Yoo et al. [8] proposed a novel colorization model of the memory augmented networks method. This method points out that high-quality colorizing results can be obtained from a small sample of training data, and the training of the memory network can be unsupervised through the threshold triplet loss without using a classification tag. In ChromaGAN, GAN, and semantic classification distribution are used as conditions and the colorization and the model training strategy is fully self-supervised.

3 Approach

3.1 Deep CNN Model

We have chosen a Deep CNN architecture as the first model to tackle our problem statement. The model is made up of 8 main layers. Where each layer refers to a block of 2-3 repeated Conv and RelU layers followed by a BatchNorm. This network has no pool layers and all changes in resolution are achieved through spatial downsampling or upsampling between conv blocks. At the start of conv8 there is a 2D transpose layer and at the end of it there is a softmax layer which is followed by a probability dist and finally upsamples to get the ab layers. which is then added with the out input to obtain the final output.

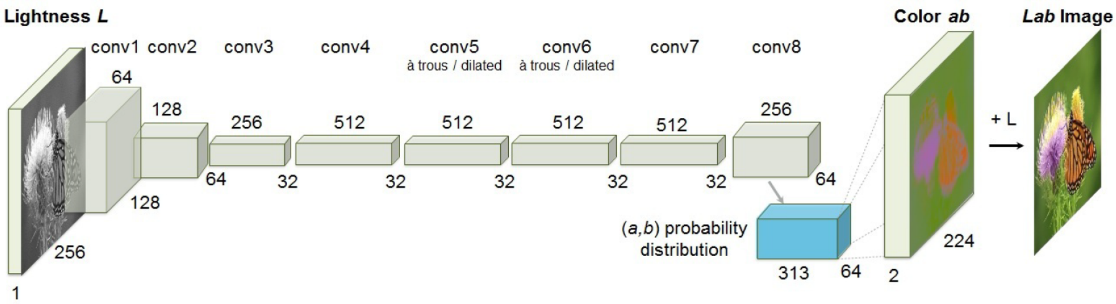

Figure 1: Deep CNN Architecture

The task of color prediction is inherently multi-modal, that is many objects can take on several plausible colorizations. For example, an apple is typically red or green but unlikely to be blue or orange. To appropriatley model the multi-modal nature of this problem, a distribution of possible colors for each pixel is predicted.

The limitations of this model, is that there is over saturation of colors from surrounding objects, as we can see from this picture, the yellow center has been evenly spread over the entire surface of the flower while the surrounding green has started to enter the edges of the flower.

3.2 GAN

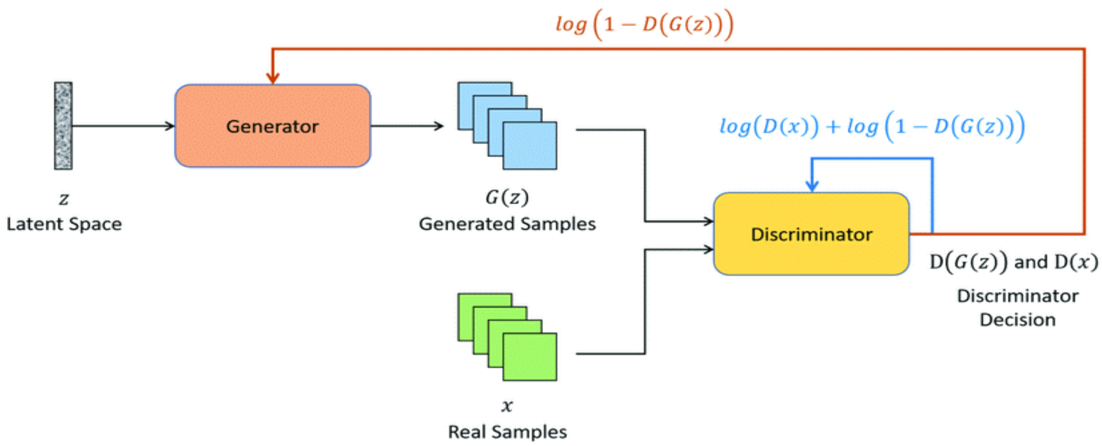

Figure 2: GAN Architecture

Traditionally in a regular GAN network the generator network initially consists of a random noise vector. However in this problem statement the same network can't be applied and therefore colours are added to images having a single channel. So to add RGB colour channel we use Conditional GANs so that the generator generates coloured images from input grayscale images and the the discriminator gets colored images from both generator and original data along with the grayscale input as the condition and tries to decide which pair contains the true coloured image.

3.3 Method

Here, a high dimensional input is mapped to a high dimensional output or in other words it's a pixel-wise regression task. The network therefore generates an output with the same spatial dimension as the input, and also colours each pixel in the grayscale input image. We use the L*a*b* color space for the colorization task. This is because L*a*b* color space contains dedicated channel to depict the brightness of the image and the color information is fully encoded in the remaining two channels which prevents any sudden variations in color and brightness.

3.4 Architecture

We use a UNET architecture which is symmetric and the contracting path consists of 4×4 convolution layers with stride 2 for downsampling, each followed by batch normalization and Leaky-ReLU. The number of channels are doubled after each step. Each unit in the expansive path consists of a 4×4 transposed convolutional layer with stride 2 followed by batch normalization and ReLU activation function. The last layer of the network is a 1×1 convolution.

We use the same architecture for generator and for the discriminator, we use similar architecture as the baselines contractive path where a series of 4×4 convolutional layers with stride 2 is used with the number of channels being doubled after each downsampling. After the last layer, a convolution is applied followed by a sigmoid function to return a probability value of the input being real or fake.

3.5 Output

4 Conclusions

In this project we automate the colouring of grayscale images using GAN. We were able to apply the model and convert a black and white video to a coloured video. The colour scale in some frames could have been improved, which can be achieved by training it on more variety of data with a network having deeper layers and better fine tuned hyper parameters. One way we could improve the quality of GAN-based image and video colorization, is to use semantic segmentation to divide the image into meaningful regions based on its content. By doing so, the GAN can colorize each region separately based on its semantic category, resulting in more accurate and visually pleasing colorization.

However, evaluating the efficacy of GANs in this task is complex due to the abstract nature of success metrics, requiring multifaceted criteria beyond traditional measures. Future work could focus on generating efficient performance metrics to evaluate the results.

References

[1] Kamyar Nazeri, Eric Ng, and Mehran Ebrahimi, *Image Colorization using Generative Adversarial Networks*, Articulated Motion and Deformable Objects pp 85–94, 2018.

[2] Saeed Anwar, Muhammad Tahir, Chongyi Li, Ajmal Mian, Fahad Shahbaz Khan, Abdul Wahab Muzaffar, *Image Colorization: A Survey and Dataset*, 2020.

[3] I. Žeger, S. Grgic, J. Vuković and G. Šišul, *Grayscale Image Colorization Methods: Overview and Evaluation*,IEEE Access, vol. 9, pp. 113326-113346, 2021.

[4] Larsson G, Maire M, Shakhnarovich G, *Learning Representations for Automatic Colorization*, European Conference on Computer Vision (ECCV), 577–593, 2007.

[5] Zhang R., Isola P., Efros A.A., *Colorful Image Colorization*, European Conference on Computer Vision (ECCV), 649–666, 2016.

[6] Su JW, Chu HK, Huang JB, *Instance-aware image Colorization*, EEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 7965–7974, 2020.

[7] Messaoud S, Forsyth D, Schwing AG, *Structural consistency and controllability for diverse colorization*, European Conference on Computer Vision (ECCV), 649–666, 2018.

[8] Seungjoo Yoo, Hyojin Bahng, Sunghyo Chung, Junsoo Lee, Jaehyuk Chang, Jaegul Choo, *Coloring with limited data: Few-shot colorization via memory augmented networks*, Computer Vision and Pattern Recognition (CVPR),pp. 11283-11292, 2019.