INDIAN INSTITUTE OF INFORMATION TECHNOLOGY, VADODARA

Government Engineering College, Sector-28, Gandhinagar

Gujarat - 382028

Design Project Report

ON

COLORIZATION OF GRAYSCALE IMAGES USING DEEP LEARNING

Submitted by :

Hardik Soni : 201751068

Varnika DasGupta : 201751060

Preeti Yadav : 201751039

Vaidehi Vaishnav : 201751059

under the supervision of

Dr. Jignesh Bhatt

Contents

1 ABSTRACT

Image Colorization is an exciting research topic in Computer Vision. We take a gray scale image as input and want to produce an output image as similar and realistic as the ground truth version. We used two approaches, in the first approach, we used convolutional Neural Network and the in the second one we used encoder-decoder architecture with Inception ResNet V2 (trained on 1.2 million images), to achieve our goal of colorizing gray scale images more accurately.

2 INTRODUCTION

Nowadays, colorization is done by hand using photoshop but earlier it took months and required intense research to get the perfect match of even face color. A greyscale image is taken as input and using neural network a predicted colored image is obtained as output with the aim to get most realistic output.

Figure 1: Gray scale image (Left). Target colorized image (Right)

We start by a simple gray scale image as an input (image scaled to 256x256 pixels). We then use a neural network to output a predicted colorized image. We have trained our model to output photos with realistic colors by training it on realistic images. We may not get exact output. Instead, the model should produce a colorized image so realistic that a viewer could not spot the fake when looking at a true color image and an image produced by our model.

3 LITERATURE SURVEY

The digital gray scale image would be touched up pixel by pixel. This allowed for batch work to be done, but this was still done manually. While this did not automate the process to a degree, it still required heavy user input.

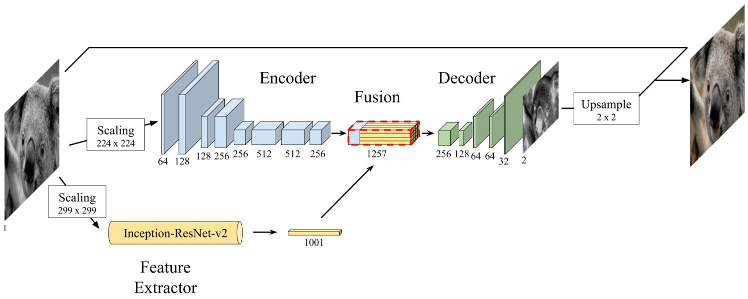
Most of our research involved the use of convolutional neural networks to colorize images. Early models use a simple mean squared error for the loss function. This leads to desaturated photos as it encourages the model to make safe predictions, which can lead to less color or browning.

4 THE PRESENT INVESTIGATION

1.Dataset
We used 80% of the available dataset in the training set. These photos are scaled to 256 x 256 pixels and include various content including faces animals, and objects. Although the input of our model is a gray scale image, our data set is all colored photos so that we have a target to optimize towards. To help with generalization, we also performed various transformations including zooms, flips, and shears to prevent overfitting. Below are examples of images used as trainig and testing dataset:

2.Model
The model is made up of four components: an encoder, a parallel classifier, a fusion layer, and a decoder. Encoder is layers of CNNs (Convolutional neural networks), while the classifier is a forward pass on the gray scale image in parallel and gets a 1000 x 1 prediction. This vector is concatenated with the output from the encoder in the fusion layer, which is then run through the decoder. Our base architecture was inspired by the the Deep Kolarization network [1] and we then experimented with deeper convolutional layers to extend the work of the paper. For this model,

we used in adam optimizer, and a mean squared error loss function. batch size of 20 images.



The model transfers learns from the Inception ResNet V2 classifier to create an embedding for each input image. We merge this embedding with the output of the encoder before feeding both as inputs to the decoder. The Inception ResNet V2 has been trained on 1.2 million images to achieve state of the art accuracy in classifying images. By transferring the learning from the classifier to the coloring network, the network can get a sense of what's in the picture.

5 RESULTS AND DISCUSSIONS

Once trained, we fed our network with some images and the results were average. However, due to the small size of our training set our network performs better when certain image features appear. For instance, natural elements such as the sea or vegetation seem to be well recognized. However, specific objects are not always well colored. The final results can be varied if the loss function is changed. We also believe our model could benefit from increased training time and a larger set of images. We run our experiment over 20 epochs on Google Colab platform.

6 CONCLUSIONS AND FUTURE WORK

In particular, our approach is able to successfully color high-level image components such as the sky, the sea or forests. Nevertheless, the performance in coloring small details is still to be improved. As we only used a reduced subset of ImageNet, only a small portion of the spectrum of possible subjects is represented, therefore, the performance on unseen images highly depends on their specific contents. To overcome this issue, our network should be trained over a larger training data set.

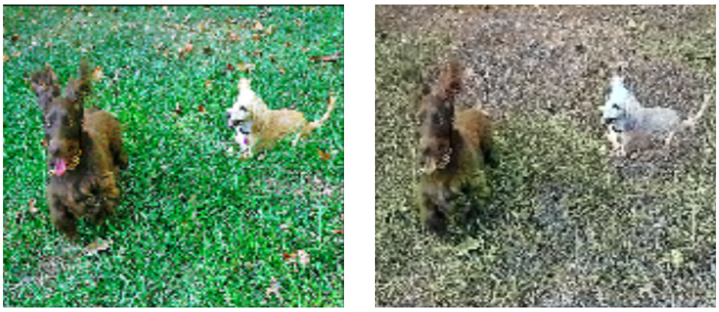


Figure 2: Ground Truth image (Left). Output image of Fusion model (Right)

Finally, it could be interesting to apply colorization techniques to video sequences, which could potentially re-master old documentaries.

7 ACKNOWLEDGMENT

We cordially thank IIIT Vadodara for this opportunity and Dr.Jignesh Bhatt for his motivation and guidance which drove our journey from root till here.

8 REFERENCES

[1] Baldassarre, Federico, Diego González Morín, and Lucas Rodés-Guirao. "Deep Koalarization: Image Colorization using CNNs and Inception-ResNet-v2." arXiv preprint arXiv:1712.03400 (2017).

[2] Rosebrock, Adrian https://www.pyimagesearch.com/2019/02/25/black-and-white-image-colorization-with-opencv-and-deep-learning/