Image Colorization

The Project is build for Computer Vision Course at IIT Jodhpur

1st Harsh Parashar
M22AIE210
Indian Institute of Technology
m22aie210@iitj.ac.in

2nd Prateek Singhal
M22AIE215
Indian Institute of Technology
m22aie215@iitj.ac.in

3rd Amit Panwar
M22AIE202
Indian Institute of Technology
m22aie202@iitj.ac.in

***Abstract*—This document includes the implementation, experiment and result of academic project of the work based on the paper "Colorful Image Colorization" [1] for course requirement of Computer vision. The code is available at https://github.com/parasharharsh16/Computer-Vison-Image-Colorization**

I. Introduction

The challenge of image colorization often yields unexpected and desaturated outcomes, primarily stemming from the loss of information in grayscale images, which only contain luminance data (L). To address this issue, the proposed approach involves predicting the 'a' and 'b' components of an image and combining them with the luminance information to produce LAB images. The advantageous aspect of color prediction is the abundance of training data, as any colored photograph can serve as a training example by simply using its L channel as input and its ab channels as the supervisory signal. The primary loss function employed in this method is the L2 loss (MSE), supplemented by a class rebalancing loss. This additional loss function involves segmenting the image into 313 probability distributions, and subsequently, each pixel's probability distribution is considered. To streamline computations, the ab space of the Lab color space is discretized into 313 bins, as illustrated in Figure 2. This discretization allows us to simplify the process, as instead of determining the 'a' and 'b' values for each pixel individually, we only need to identify a bin number ranging from 0 to 312. An alternative perspective on the problem is to recognize that we already possess the L channel, which spans values from 0 to 255, and our objective is to ascertain the ab channel, which operates within the range of 0 to 312. Consequently, the color prediction task is transformed into a multinomial classification challenge.

II. Dataset

We will be using combination of following dataset to include diversity in images and train a generalize image colorization model. This approach aims to ensure the training of a robust and generalized model capable of accurately colorizing a wide range of images. The dataset comprises the following components:

- **Sport Images**:This dataset consists of approximately 14.4k sports images, each sized 224×224 pixels, covering a spectrum of 100 different sports.
- **Landscape Images**: Included in this dataset are approximately 7.1k landscape images, each sized 150×150 pixels. These images depict various natural elements such as trees, buildings, mountains, glaciers, and more.
- **Pascal VOC 2012**: This dataset encompasses approximately 31k images, with our focus being on the images categorized within the Train set, totaling around 17k images. The images in this dataset vary in size, ranging approximately from 250×500 to 500×500 pixels.

From each of the mentioned datasets, we have extracted a representative subset to construct our working dataset, resulting in approximately 11k samples. This selection process ensures diversity in image content and size, facilitating the training of a versatile image colorization model.

A. Data Preprocessing

In our image colorization project, we've created a custom dataset to prepare our images for colorization. First, we load images from file paths in the RGB format. Next, we can resize images to a specified size using Pillow's resizing capabilities. Now converting images from RGB to LAB color space. This breaks down each image into its lightness (L) and color (a and b) components. After separating the channels, we focus on the lightness channel (L) while keeping the color channels (a and b) together. Finally, the pre-processed data, containing the lightness and color channels, are ready to be fed into our deep learning model for colorization. Where L is setup as input and ab as target for the model. The image has been resized to dimensions of (256, 256).

III. Model Architecture

This paper presents the architecture of the ECCVGenerator, a neural network designed for grayscale image colorization tasks. The ECCVGenerator utilizes a series of convolutional and transpose convolutional layers to extract hierarchical features from input images and predict chrominance channels (a and b) based on the luminance component. The model incorporates dilation convolutional layers to increase the receptive field effectively without significantly increasing the number of parameters. Additionally, the output layer employs softmax activation to produce a probability distribution over color bins in the chrominance space. The network is capable of transforming grayscale images into colorized versions with high accuracy.

Layer	Configuration
Input Layer	1-channel grayscale image (L)
Feature Extraction	Conv2d(1, 64, 3x3, stride=1, padding=1) + ReLU, Conv2d(64, 64, 3x3, stride=2, padding=1) + ReLU, ...
Dilation Convolution	Conv2d(512, 512, 3x3, dilation=2) + ReLU, ...
Decoder	TransposeConv2d(512, 256, 4x4, stride=2, padding=1) + ReLU, Conv2d(256, 256, 3x3, stride=1, padding=1) + ReLU, ...
Output Layer	Conv2d(256, 313, 1x1) + Softmax

TABLE I
ECCVGENERATOR ARCHITECTURE

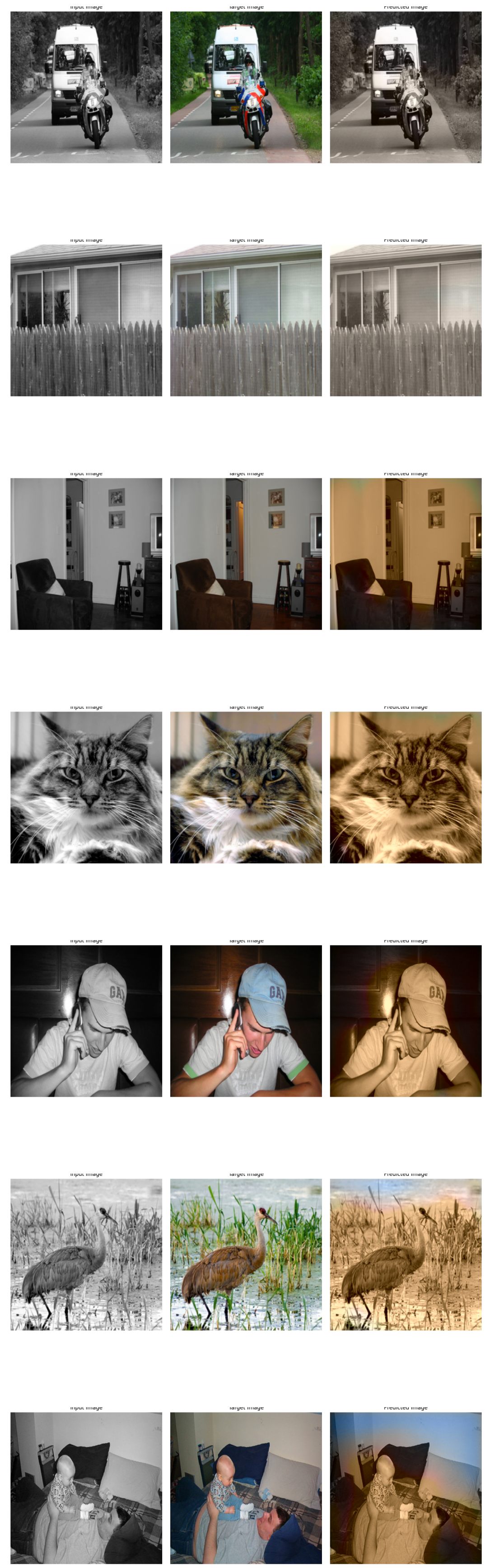
IV. MODEL TRAINING

The ECCVGenerator architecture, presented in this paper, was trained using an L2 loss function on a subset of the dataset. Initially, we trained the model on a small fraction, approximately 1%, of the total dataset consisting of around 55,000 images. However, during the initial training phase, we encountered challenges as the colorization results appeared to lack color variation. This was attributed to the fact that the loss calculation was performed on the entire LAB image rather than solely on the AB channel.

To address this issue, we experimented with focusing the loss calculation and backpropagation solely on the AB channel. This adjustment resulted in saturated grayscale images with hints of hue-like colors, marking a notable improvement over the initial outcomes. Further optimization was pursued by adjusting the learning rate, as fluctuations in loss were observed during training. By fine-tuning this parameter and increasing the size of the training dataset, we began to observe more refined colorization results compared to the previous iterations. Additionally, attempts were made to apply class rebalancing techniques to the existing data. However, this endeavor was hindered by the absence of an appropriate weight approximation method as described in the literature and the proposed solutions. Future research will aim to address this limitation and explore more advanced techniques for improving model performance and colorization quality.

V. RESULTS

The performance of the ECCVGenerator model was evaluated based on its ability to accurately colorize grayscale images. A selection of sample images from the test dataset along with their corresponding colorized outputs are presented below Upon visual inspection, it can be observed that the ECCVGenerator model effectively infuses vibrant and realistic colors into the grayscale input images. However, it is not very accurate due to less training data and no class rebalancing.



VI.

REFERENCES

[1] arXiv:1603.08511v5 [cs.CV] 5 Oct 2016