Grayscale Image Colorization

Ioannis Ioannidis
Dimitrios Patiniotis Spyropoulos

University of Piraeus
NCSR Demokritos
Athens

22-06-2022

Outline

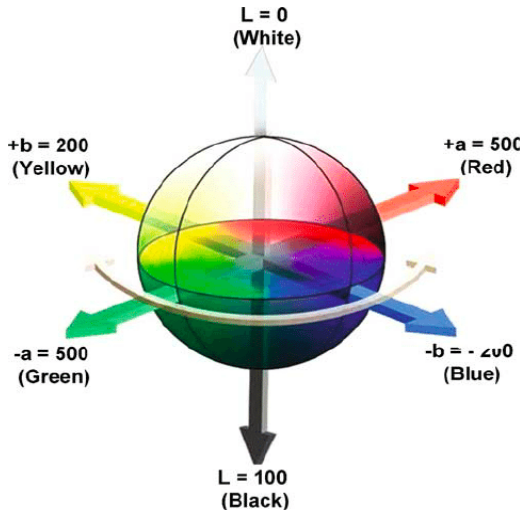
- Introduction
- Dataset Description
- Theoretical Background
- Model's Architecture
- Results
- Conclusions

Introduction

- Hallucinating the colors of a grayscale image seems daunting at first sight.
- However, in many cases the semantics of the scene and its surface texture provide ample cues for many regions in each image.
- Our goal for this project is to produce a plausible colorization that could potentially fool a human observer.

Dataset Description

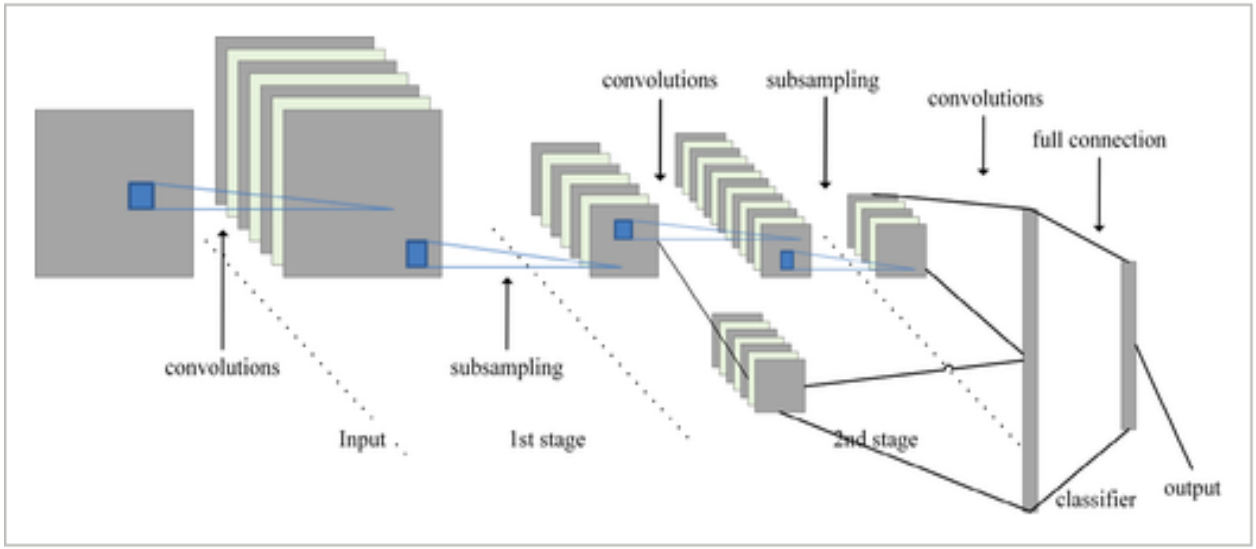
- We used a dataset consisting of 25000 images available at https://www.kaggle.com/datasets/shravankumar9892/image-
- Dataset structure is based on the LAB color space image format.



- Basically what we try to do is get the L channel as input and try to predict the respective a and b components.

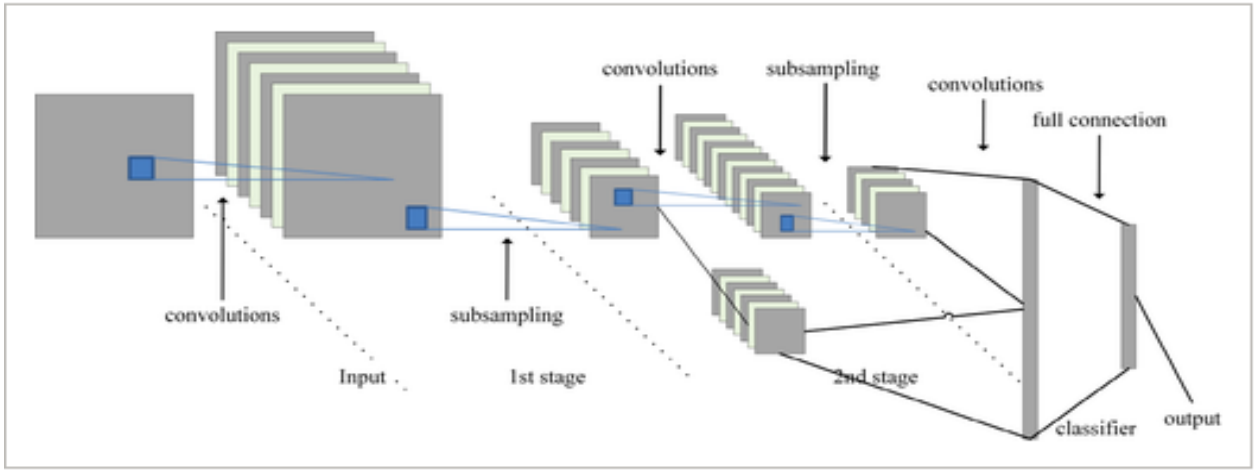
Theoretical Background - CNN

- Convolutional Neural Networks (CNNs) is a state of the art pattern recognition method in computer vision.
- Unlike traditional neural networks, which work with one-dimensional feature vectors, a CNN takes a two-dimensional image and consequentially processes it with convolutional layers.



Theoretical Background - CNN

- A CNN is composed of input and output layers and multiple hidden layers, which can be divided into a convolution layer, a pooling layer, a rectified linear unit layer, and a fully connected layer.
- Each convolutional layer consists of a set of trainable filters and computes dot productions between these filters and layer input to obtain an activation map.
- These filters are also known as kernels and allow detecting the same features in different locations.



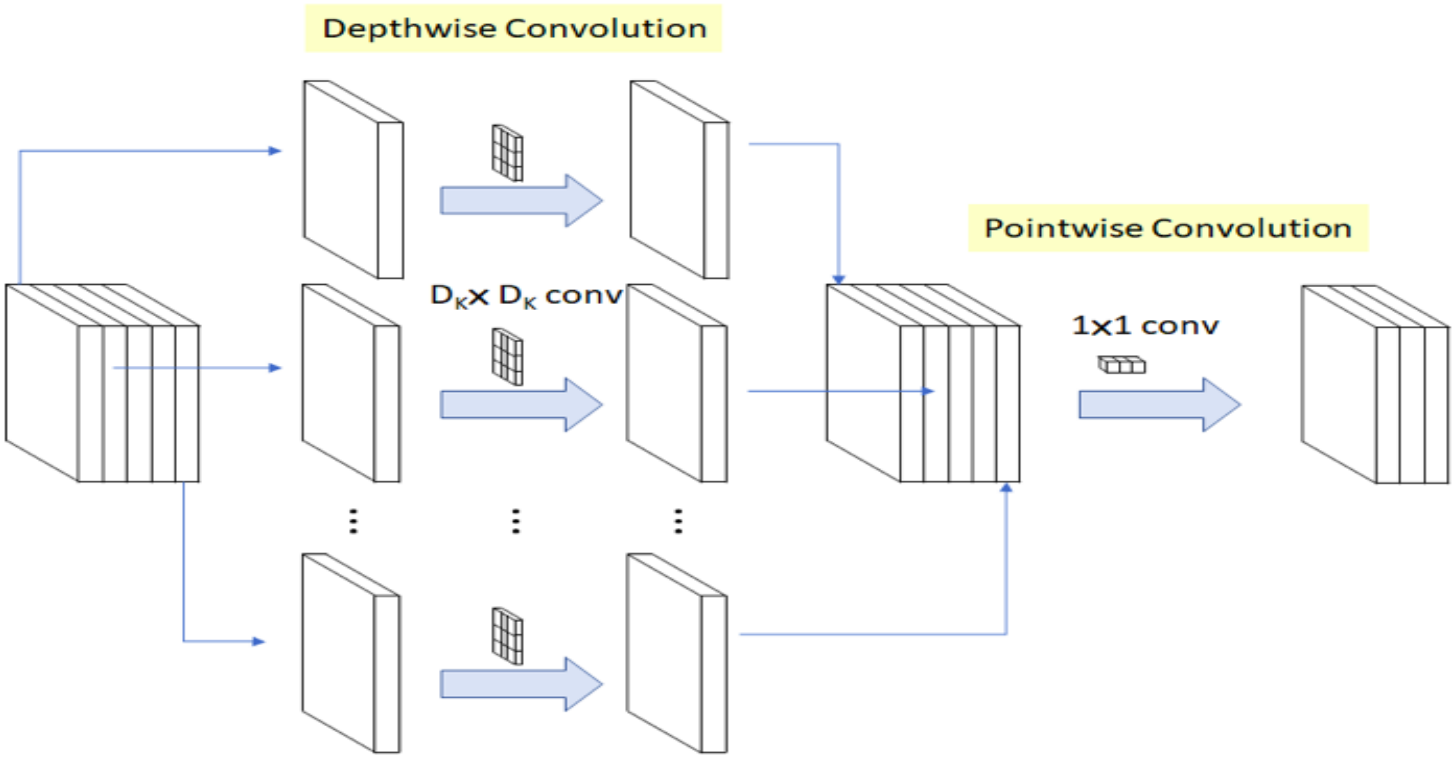
Theoretical Background - MobileNet

- MobileNet is an efficient and portable CNN architecture that is used in real world applications.
- MobileNets primarily use depthwise seperable convolutions in place of the standard convolutions used in earlier architectures to build lighter models.
- A standard MobileNet has 4.2 million parameters which can be further reduced by tuning the width multiplier hyperparameter appropriately.
- The size of the input image is $224 \times 224 \times 3$.

Theoretical Background - MobileNet

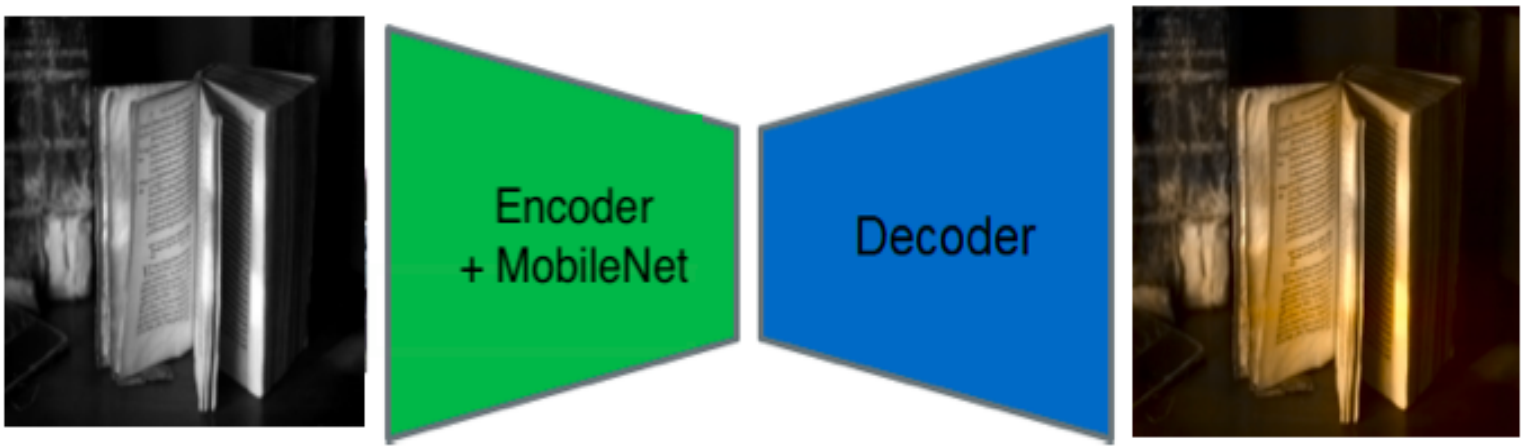
A depthwise seperable convolution is made from two operations:

1. depthwise convolution
2. pointwise convolution



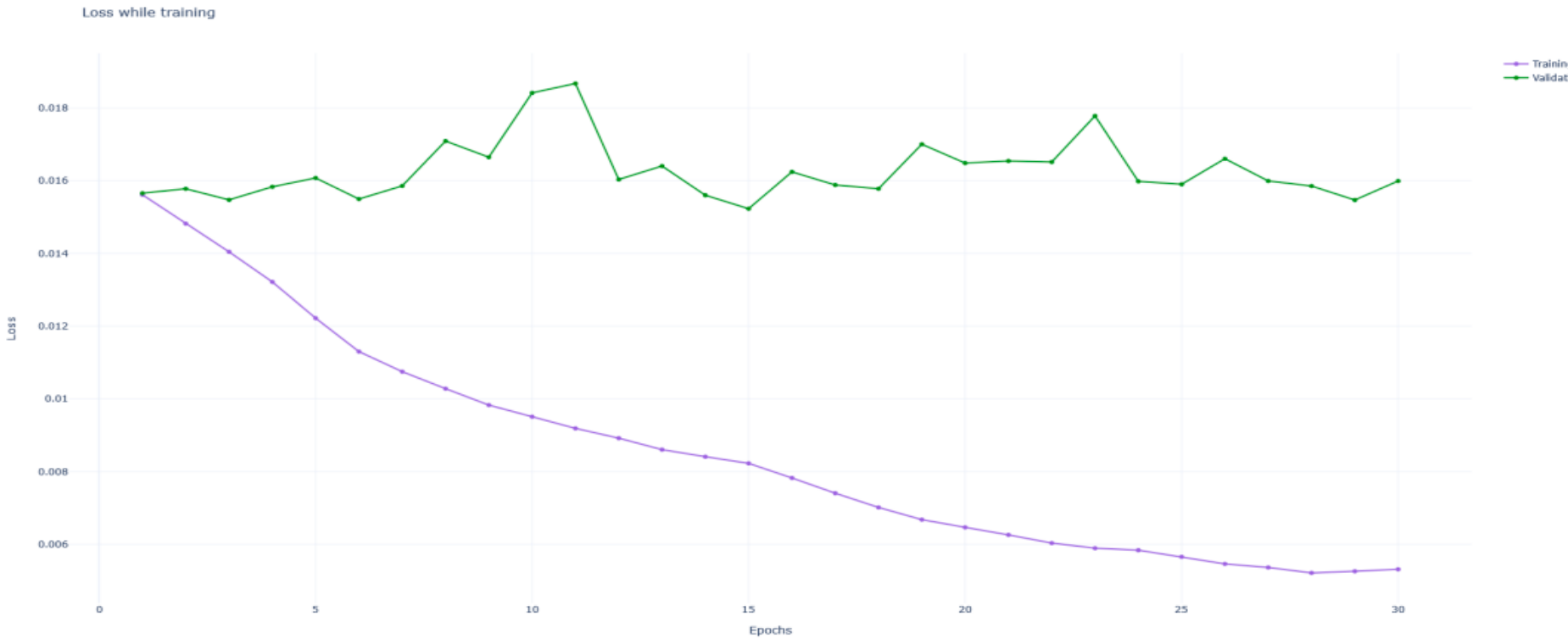
Model's Architecture

- Input grayscale image pass through the parallel branches of *Encoder* and *MobileNet*.
- Features extracted are combines in the *Fusion* layer.
- Fusion's output finally pass through *Decoder*'s convolution which attempt to recover LAB's a and b components.



Results

- Despite the gradual and normal fall of the error in training set as was expected, the validation's one remained on the same levels.



Results

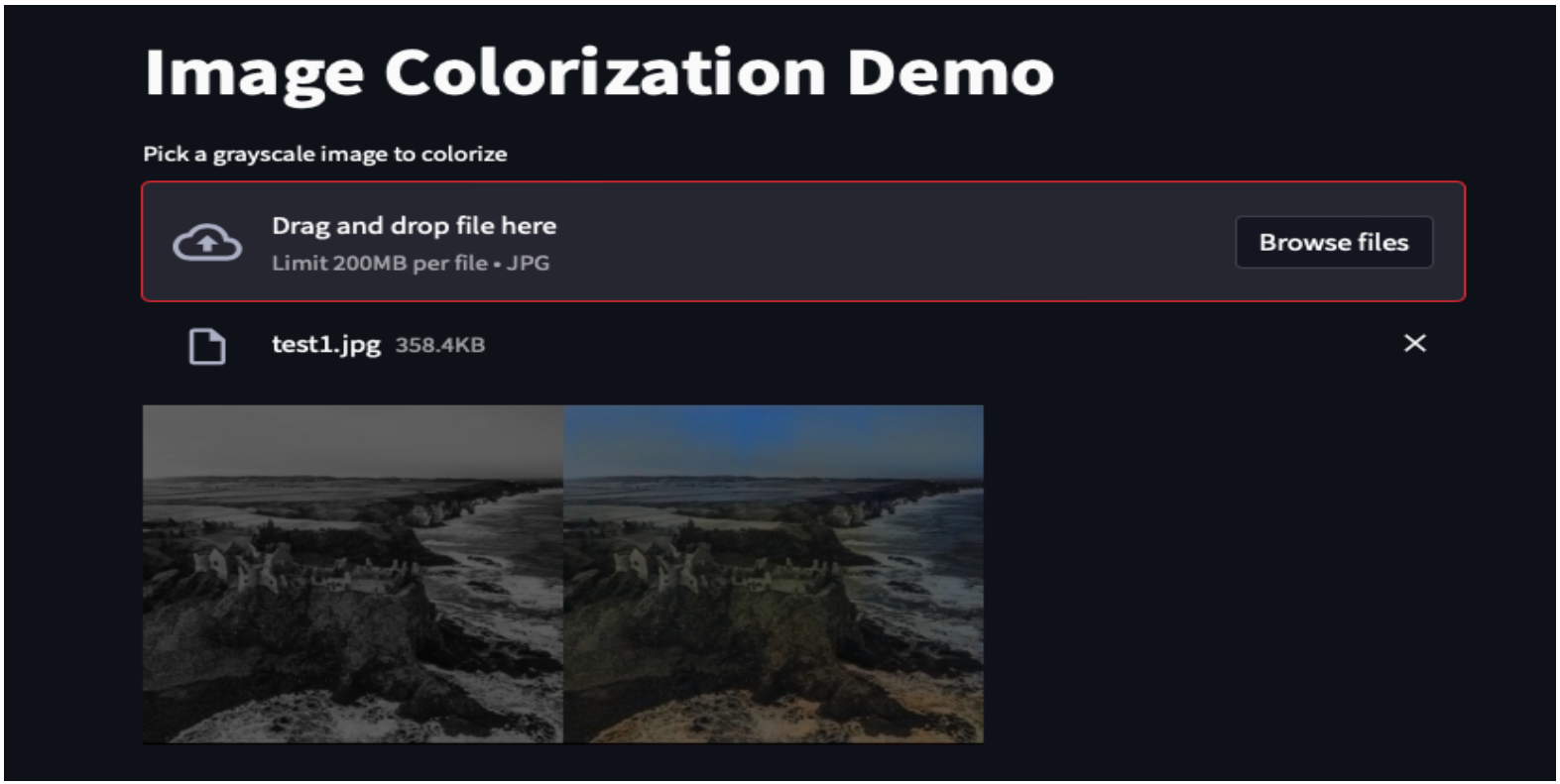
- However the model proved to efficiently predict a part of the colorized photo.

Results

- However the model proved to efficiently predict a part of the colorized photo.

Results

- The model is deployed using Streamlit for a live demo.
- You can use it by clicking https://www.colo-rize.eu



Conclusion

- Image colorization is a challenging task.
- Our final model returns plausible representation of the colorized image using a bespoke architecture, which was our initial goal.
- However, there are many ways to improve the final outcome for getting more realistic images.