A Project Report Submitted

for Machine Learning (UML501)

by

Batch 3 COE 2

102103035 Ankit Bhardwaj

102103053 Aaryan Gupta

Image Colorization Using Conditional GAN

Submitted to

Dr. Anjula Mehto

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY (A DEEMED TO BE UNIVERSITY), PATIALA, PUNJAB

INDIA

August-December 2023

Introduction

In recent years, the field of computer vision has witnessed remarkable advancements, with image colorization being one of the fascinating areas of research. Image colorization refers to the process of automatically adding color to grayscale images, recreating the visual experience of the scene. This task, though seemingly simple to the human eye, poses a significant challenge for machines due to the ambiguity and complexity involved.

One popular approach to tackle this problem is by utilizing Generative Adversarial Networks (GANs), which have demonstrated remarkable success in various image generation tasks. The combination of GANs with conditional image generation has paved the way for exciting advancements in image colorization.

Conditional GANs (cGANs) are an extension of GANs that take into account additional input information, known as conditioning variables, to generate more targeted and controlled outputs. In the context of image colorization, the conditioning variable could be a grayscale image, serving as the input, and the corresponding colored image, serving as the output.

The project will involve training the Conditional GAN on a data set of colored images. We will employ a generator network that takes a LAB image as input and generates a colorized version, while a discriminator network will assess the quality and realism of the generated images. Through a process of adversarial training, the generator network will learn to produce colorized images that are indistinguishable from real ones, fooling the discriminator network.

Problem Statement

The task of image colorization has been a long-standing challenge in the field of computer vision. While humans effortlessly perceive and interpret colors in images, teaching machines to replicate this ability remains a complex endeavor. Traditional methods of image colorization often rely on manual intervention or simplistic algorithms that fail to produce high-quality and realistic results.

The aim of this project is to address the problem of image colorization using Conditional Generative Adversarial Networks (cGANs). By leveraging the power of deep learning and adversarial training, we seek to develop a model capable of automatically adding color to grayscale images, producing visually coherent and realistic colorized outputs.

The primary challenges in this project include:

1. Ambiguity in grayscale-to-color mapping: Grayscale images lack explicit color information, making it challenging for machines to accurately infer the appropriate color distribution. The model must learn to understand the underlying relationships between grayscale and colored images to generate plausible and visually appealing colored images.

2. Preservation of image semantics: An effective image colorization model should not only maintain the overall structure and semantics of the original image but also capture the intrinsic relationships between objects and their corresponding colors. Preserving these details while adding color is crucial for producing visually accurate and recognizable outputs.

3. Handling diverse color distributions: Images encompass a wide range of color distributions, influenced by various factors such as lighting conditions, object textures, and artistic styles. The model should be capable of adapting to these diverse color distributions and generating colored images that are consistent with the given input.

4. Generating realistic and visually coherent colored images: The colorized outputs should appear natural and visually appealing to human observers. The model needs to learn to produce realistic color textures, gradients, and shading, ensuring that the colorized images look convincing and coherent.

5. Data set availability and quality: Acquiring a large-scale data set of grayscale and corresponding colored images is essential for training the model effectively. The data set should encompass a diverse range of images, covering various scenes, objects, and color distributions. Ensuring the quality and accuracy of the data set is crucial to avoid biased or misleading training.

Data Set

We utilized an external data set containing RGB images of various scenes and places. The data set consists of a total of 2791 RGB images and was obtained from a third-party source while conducting research on different methods for image colorization.

The data set used in this project is not our own, and credit goes to the original owner who employed it to train a Generative Adversarial Network (GAN). We accessed the data set through the following link: [Source Link]

To implement the image colorization using Conditional GANs, we specifically adopted the pix2pix method, as described extensively in the paper titled "Image-to-Image Translation with Conditional Adversarial Networks" by Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2016) [Source Link].

This data set offers a diverse collection of RGB images, enabling our model to learn the mapping between grayscale and colored images effectively. By leveraging this data set, we aimed to train our Conditional GAN model to generate visually appealing and realistic colorized outputs for grayscale images.

Please note that the data set is not our original creation, and all credit goes to the respective owner. We have used this data set solely for academic purposes and to explore the application of Conditional GANs for image colorization in our college project.

Methodology

1. **Data Collection**

In this project, a data set containing RGB images of various scenes and places was used. It consists of 2791 RGB images.

2. **Data Preprocessing**

The data set is divided into a training set consisting of 2500 images and a validation set containing 291 images. Our target images within the data set are in RGB format, while we generate input images for each by converting them into LAB format. To ensure consistency, both the RGB and LAB images are resized and their pixel values are normalized.

3. **Model Architecture**

Pix2Pix is a Generative Adversarial Network, or GAN, model designed for general purpose image-to-image translation. The approach was presented by Phillip Isola, et al. in their 2016 paper titled "Image-to-Image Translation with Conditional Adversarial Networks" and presented at CVPR in 2017.

The GAN architecture is comprised of a generator model for outputting new plausible synthetic images and a discriminator model that classifies images as real (from the data set) or fake (generated). The discriminator model is updated directly, whereas the generator model is updated via the discriminator model. As such, the two models are trained simultaneously in an adversarial process where the generator seeks to better fool the discriminator and the discriminator seeks to better identify the counterfeit images.

The Pix2Pix model is a type of conditional GAN, or cGAN, where the generation of the output image is conditional on an input, in this case, a source image. The discriminator is provided both with a source image and the target image and must determine whether the target is a plausible transformation of the source image.

Again, the discriminator model is updated directly, and the generator model is updated via the discriminator model, although the loss function is updated. The generator is trained via adversarial loss, which encourages the generator to generate plausible images in the target domain. The generator is also updated via L1 loss measured between the generated image and the expected output image. This additional loss encourages the generator model to create plausible translations of the source image.

Generator Architecture:

The encoder-decoder architecture consists of:

encoder:

C64-C128-C256-C512-C512-C512-C512-C512

decoder:

CD512-CD512-CD512-C512-C256-C128-C64

After the last layer in the decoder, a convolution is applied to map to the number of output channels (3 in general, except in colorization, where it is 2), followed by a Tanh function. As an exception to the above notation, BatchNorm is not applied to the first C64 layer in the encoder. All ReLUs in the encoder are leaky, with slope 0.2, while ReLUs in the decoder are not leaky.

The U-Net architecture is identical except with skip connections between each layer i in the encoder and layer $n - i$ in the decoder, where n is the total number of layers. The skip connections concatenate activations from layer i to layer $n - i$. This changes the number of channels in the decoder:

U-Net decoder:

CD512-CD1024-CD1024-C1024-C1024-C512-C256-C128

Discriminator Architecture:

The 70×70 discriminator architecture is:

C64-C128-C256-C512

After the last layer, a convolution is applied to map to a 1-dimensional output, followed by a Sigmoid function. As an exception to the above notation, BatchNorm is not applied to the first C64 layer. All ReLUs are leaky, with slope 0.2.

4. **Training**

The Conditional GAN model is trained using the prepared LAB and RGB images. The training process involves optimizing the generator and discriminator networks through an adversarial training scheme. The generator aims to generate colorized images that can fool the discriminator, while the discriminator learns to distinguish between real and generated colorized images. 2500 images trained for 6 epochs and batch size 4.

5. Evaluation

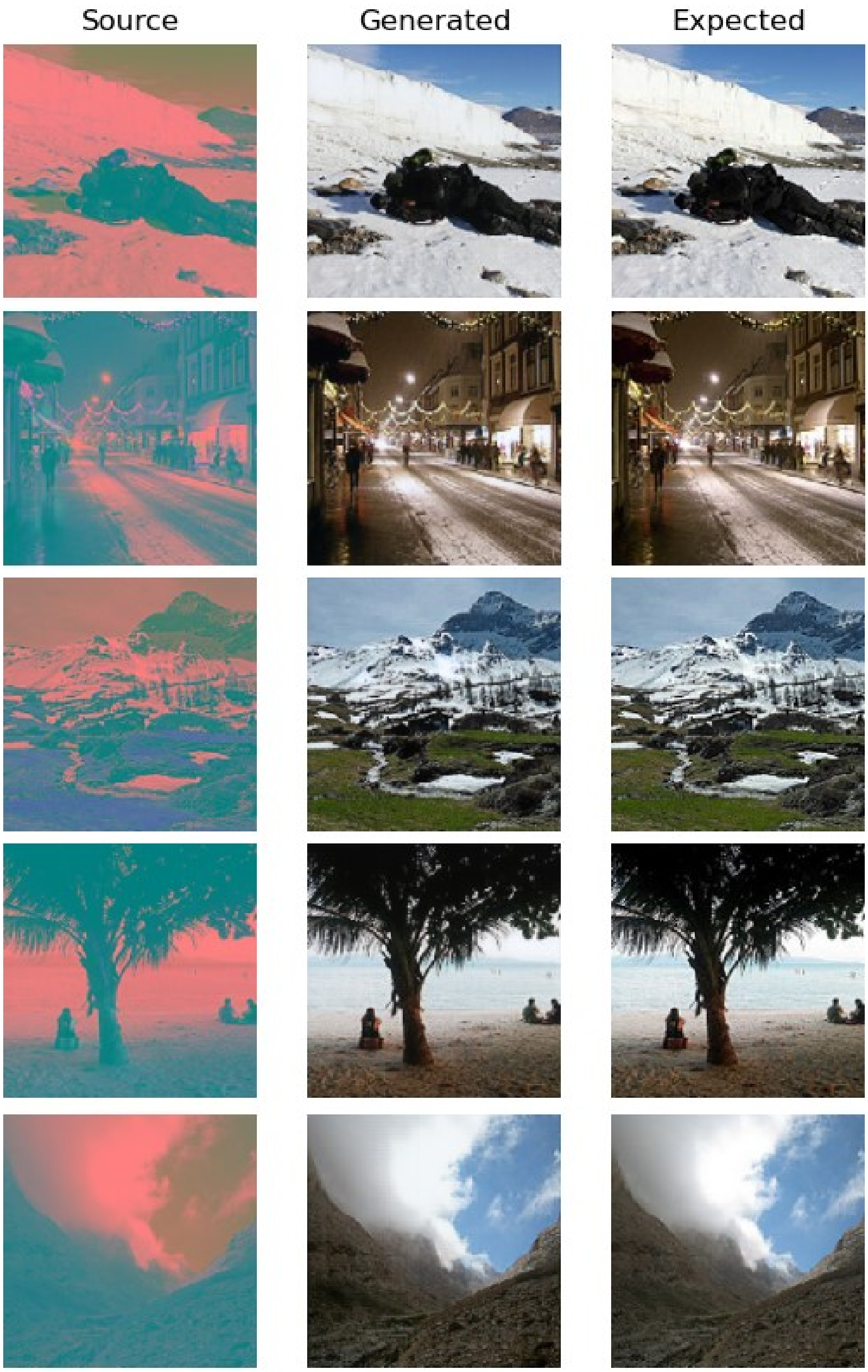
The trained model is evaluated using a combination of visual inspection and quantitative metrics. Visual inspection involves visually examining the generated colorized images to assess their quality and similarity to the ground truth color images. This subjective evaluation provides insights into the overall performance of the model.

In addition to visual inspection, quantitative evaluation metrics are utilized to provide objective measures of the model's performance. Two commonly used metrics are the peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM). We have used SSIM, which assesses the structural similarity between the generated images and the ground truth images, taking into account luminance, contrast, and structural information.

By employing both visual inspection and quantitative metrics such as SSIM, a comprehensive evaluation of the trained model's performance is obtained, enabling a thorough assessment of the effectiveness and accuracy of the image colorization process.

SSIM gives an average accuracy of 0.949.

Final Outcome



Average Accuracy is 0.949.

GitHub Link: https://github.com/Barbaaryan/MLProject