Image Colorization

Deliar, John, Xinzhou

Introduction

- The goal of this project is to be able to take a grayscale image of a person as input and output a realistic colored version of the Image. The figure below illustrates this goal, on the left is a grayscale image, on the right is the ground truth image.
- This project is interesting because it has many real-world applications such as restoring historic images, colorizin surveillance images, and so on.
- Image Colorization has been a research hot-spot ever since deep learning took off.
- With the rise of deep neural networks, many research scientists around the globe have been competing to create the best model for this task.

Training & Evaluation

- Training Strategy:

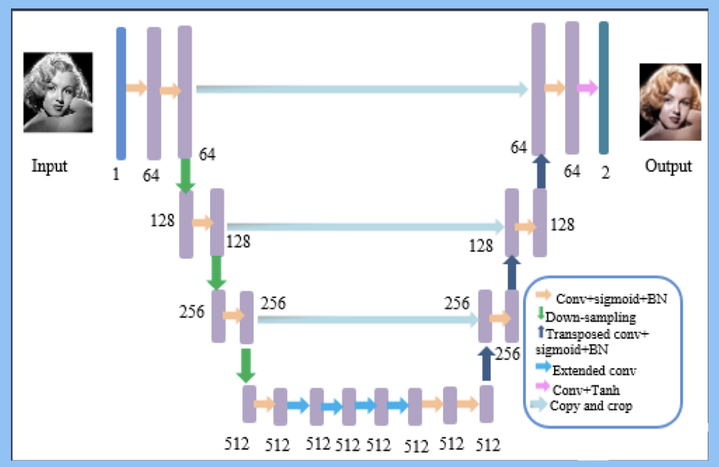
- During the training phase, a preliminary test was conducted with a dataset size of approximately 20 to verify the model's capability of learning colors. It was found that reasonable colors could only be produced when the loss value was reduced to below 1000, which was used as a critical criterion for subsequent training. Once it was confirmed that the model was capable of producing certain colors, training was initiated.
- We trained the CUnet with different sizes of dataset for find the upper limit of the model.
- We tried to train the model with dataset size of 20, 40, 100.
- With lr was setted to 0.001, after training 400-600 epochs, the loss value would be decreased to below 500.
- For training size of 200, we tried to decrease the lr from 0.001 to 0.0001, once the loss value fluctuated steadily in some range.

- Evaluation Metrics:

- Quantitative metrics: mean square error (MSE)
- Calculate the loss value show the difference between the output of the model with the ground truth of inputs.\
- Non quantitative metrics:
- human perception and subjective evaluation of the model's output if it Conforms to human color perception

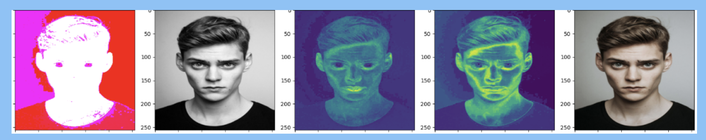
Model Architecture

- The name of the architecture is CU-net, created by Na Wang, Guo-Dong Chen, and Ying Tian.
- Based on the U-net architecture, the CU-net also utilizes downsampling, upsampling, and skip layers
- The main difference is the use of repeated extended convolutions to replace two downsampling and upsampling steps
- Extended, or dilated, convolutions, are convolutions with gaps in them. These allow for high-level features to be learned through their large receptive field, without losing feature resolution or increasing training time.
- Below is a diagram of our model



Dataset

- The dataset we used for this project was sourced from Kaggle https://www.kaggle.com/datasets/ashwingupta3012/human-faces
- It contains 7000 images of human faces.
- The images are very diverse, displaying a multitude of skin colors, eye colors, lighting, and angles. This should allow our model to generalize well to new images.
- The images are all different sizes, and must be resized to 256x256 before being passed to our network. Additionally, we convert the images to the LAB color space and extract the L channel to feed into the model. The image on the bottom demonstrates the different channels of an LAB image and the reconstruction of the image from those channels.
- Below are a few samples from the dataset.



Discussion

- After training for several hundred iterations, our model was able to perform fairly on the dataset. In particular it did well with images that did not contain any bold or bright colors. Of all the images colored by the model, roughly half of them passed as realistic images. In order for an image to pass it would have to fool us into thinking it was a real image and not outputted by the model.
- The images to the right demonstrate some of the outputs, on the left hand side is the ground truth, and on the right is the image colored by the model.
- We noticed that when the model ran on images that contained very bright colors, it would often guess incorrectly for that image. This is because we are using MSE which is a loss function that promotes more conservative guesses for each pixel.
- In order to make the image work well for brighter colors we would need to develop a loss function that does not punish predictions of bright colors as much as our current loss function.
- To improve the model even further, we can try to segment different parts of the image such as eyes, nose, and hair. Then color these segments individually. This could result in more accurate renditions.
- With this model functioning, we can expand its applications to colorizing videos, historic images, and so much more.