Imam Mohammad Ibn Saud Islamic University

College of Computer and Information Sciences

Computer Science Department

First semester 1446 H - 2024 G

# Colorization

## Project's field: Computer Vision

{Abdullah Abdulrahman Alahiadeb, 442019225@sm.imamu.edu.sa}

**Problem Statement:**

The problem is to train a deep learning model to colorize grayscale images, using the ImageNet dataset. It is interesting because it involves converting black-and-white images into realistic color versions, which requires the model to understand the context and content of images. This can have applications in restoring old photographs, enhancing medical images, or even in creative fields like art and film.

**Challenges:**

One challenge is creating a model that can effectively learn the mapping between grayscale and colored images. Since there are many possible color combinations for any given grayscale image, it is hard for the model to predict accurate and natural colors.

Additionally, training such a model requires significant compute power, as deep learning models are computationally intensive, especially with large datasets like ILSVRC2017_DET [1], which is 55GB.

Properly preprocessing the data and ensuring that the model receives input in the right format is another technical challenge.

**Dataset:**

I am using the ILSVRC2017_DET dataset for this project. This dataset is a subset of ImageNet, containing a variety of images. I have acquired it by requesting an application to the ImageNet website with my university Email address. The dataset is used as a standard benchmark for many computer vision tasks, and it offers a diverse set of images that help in training a robust colorization model.

**Methodology:**

I am using a deep learning approach for image colorization that leverages a modified ResNet-50[2] as the feature extractor, the first convolutional layer of ResNet-50 is adapted to take in grayscale images instead of RGB images, and I utilize its pretrained weights to encode grayscale features. The decoder then takes the encoded features and learns to predict the color components (ab channels) in the LAB color space.

The idea is to freeze the parameters of the encoder to retain the general feature extraction capabilities of the pretrained ResNet-50, which can help speed up training and improve performance, given limited computational resources.

**Evaluation:**

The primary evaluation metric will be the Mean Squared Error (MSE) loss between the predicted and actual ab channels, which measures how close the predicted colors are to the true colors. The latest loss is 0.01495 and its continuing to decrease as I further train the model

I will also monitor both training and validation losses across epochs to ensure that the model is learning effectively and to avoid overfitting. A lower validation loss would indicate that the model is generalizing well to unseen images.

References:

[1] ILSVRC2017_DET, "ImageNet Large Scale Visual Recognition Challenge 2017 Downloads," Available: https://image-net.org/challenges/LSVRC/2017/2017-downloads.php. [Accessed: 11-Oct-2024].

[2] ResNet-50, "torchvision.models.resnet50," PyTorch, Available: https://pytorch.org/vision/main/models/generated/torchvision.models.resnet50.html#torchvision.models.resnet50. [Accessed: 11-Oct-2024].