Question 3: Assignment 5: CS 663, Fall 2024

Amitesh Shekhar
IIT Bombay
22b0014@iitb.ac.in

Anupam Rawat
IIT Bombay
22b3982@iitb.ac.in

Toshan Achintya Golla
IIT Bombay
22b2234@iitb.ac.in

November 07, 2024

1. This is a fun exercise where you are officially allowed to do a google search and find out a research paper which works on an image restoration problem which is <u>different</u> from all the ones we have seen in class. In your report, you should clearly state the problem, write the title, venue and publication year of the research paper, and mention the cost function that is optimized in the research paper in order to solve the problem. In the cost function, you should mention the meaning of all variables. For your reference, here is a list of restoration problems we have encountered in class: image denoising, image deblurring, image inpainting, reflection removal, stitching together images of torn pieces of paper (you saw this one in the midsem), notch filters for removal of periodic interference patterns in images. You are <u>not</u> allowed to mention any of these. [15 points]

Soln:

Image Restoration Problem: Colorization of Grayscale Images

Paper Details:
- Title: "Colorful Image Colorization"
- Authors: Richard Zhang, Phillip Isola, and Alexei A. Efros
- Year: 2016

Problem: Colorization of grayscale images is a task where the goal is to automatically add color to black-and-white images. It is an image restoration problem that deals with filling in missing color information from grayscale images, a problem that is particularly useful in areas such as old photograph restoration, film production, and computer vision.

Initial Cost Function:
In the paper, the authors initially present a deep learning-based approach to colorize grayscale images. They use a convolutional neural network (**CNN**) to predict the color channels (e.g., red, green, blue) from the grayscale input image. The key idea is to learn the mapping between grayscale images and their color versions using a large dataset of paired images. The cost function that is optimized in this method is the mean squared error (MSE) between the predicted color image and the ground truth color image. The cost function can be expressed as follows:

$$\mathcal{L} = \sum_{i=1}^{N} \left\| \mathbf{I}_{pred}^{(i)} - \mathbf{I}_{gt}^{(i)} \right\|^2$$

where:

- $\mathcal{L}$ is the total loss over all pixels.

- N is the total number of pixels in the image.

- $\mathbf{I}_{pred}^{(i)}$ is the predicted color for pixel i, output by the neural network.

- $\mathbf{I}_{gt}^{(i)}$ is the ground truth color for pixel i. How is ground truth color determined? During training, the color images (RGB) are converted to grayscale using a standard method, typically by applying a linear transformation that combines the RGB channels into a single intensity value (luminance). A common formula used is:

$$I_{\text{gray}} = 0.2989 \cdot R + 0.5870 \cdot G + 0.1140 \cdot B$$

This gives a grayscale image that approximates how humans perceive brightness. The ground truth color values for a given pixel i in the training images are the original RGB color values from the paired color image (before it was converted to grayscale). In other words, for each grayscale image in the training set, the corresponding full-color image (RGB) provides the ground truth color for each pixel.

- The norm $\|\cdot\|$ represents the L2 (Euclidean) distance between the predicted and ground truth color values.

The loss function encourages the network to minimize the difference between the predicted colors and the actual colors in the training data, effectively learning how to colorize grayscale images.

However, there are some issues with minimizing the loss function mentioned above because the problem is inherently ambiguous and multimodal, as there can be multiple plausible colorizations for a given grayscale image. Standard Euclidean loss functions tend to produce desaturated results, so this paper approaches the problem using a multinomial classification framework to achieve realistic colorizations.

Better Cost Function:

The paper then proposes to optimize the *multinomial cross entropy loss*, L_{cls}, to handle the multimodal nature of colorization. The cost function is defined as:

$$L_{\mathrm{cls}}(\hat{\mathbf{Z}}, \mathbf{Z}) = -\sum_{h,w} v(\mathbf{Z}_{h,w}) \sum_{q} \mathbf{Z}_{h,w,q} \log\left(\hat{\mathbf{Z}}_{h,w,q}\right)$$

where:

- $\hat{\mathbf{Z}} \in [0,1]^{H \times W \times Q}$: The predicted probability distribution over Q (Note that Q denotes the number of bins into which the ab color space has been quantized), where H and W represent the image dimensions.

- $\mathbf{Z} \in [0,1]^{H \times W \times Q}$: The ground truth color distribution, converted from the true color values using a soft-encoding scheme.

- $v(\mathbf{Z}_{h,w})$: A weighting term that balances the loss based on color-class rarity, so that less common colors receive more weight.

- q: Index over the space Q.

This loss function encourages the model to assign higher probabilities to plausible colors, avoiding the desaturated effect common in colorization tasks using Euclidean loss.