Image analysis course SICOM 3A - Project

Stanislas Domer

1. Problem Statement

The objective of this project is to address the challenge of demosaicking, a critical step in image processing, particularly relevant in the context of RGB image acquisition using Color Filters Array (CFA) technology. Commercial RGB cameras commonly employ CFAs, which consist of red, green, and blue filters arranged in periodic patterns over the camera sensor. Consequently, each pixel on the sensor captures only one color component (red, green, or blue), resulting in a grayscale raw image.

The process of demosaicking involves reconstructing the missing color information for each pixel, thereby restoring the full RGB image. In this project, we are presented with raw grayscale images obtained from a simulated CFA camera with two different CFA patterns: Bayer and Quad Bayer. Our task is to leverage methods learned during the semester to develop algorithms capable of reconstructing the original RGB images solely based on the raw acquisitions and the provided forward operator.

2. Solution Approach

A first approach uses a CNN encoder-decoder model with 2 encoder and 2 decoder layers, for a total of 37,000 parameters. This model has been trained on a database containing 500 1024x1024 tables. However, given the low complexity of the network in relation to the task in hand, I chose to use a much larger, already-trained network.

2.1 The Chosen Solution: Theory Explanation

The selected solution for the demosaicking task leverages a pre-trained Convolutional Auto-Encoder model originally designed for image colorization[1]. This model architecture is based on Convolutional Neural Networks (CNNs) and follows an encoder-decoder structure, commonly referred to as an autoencoder.

2.2 Theory Behind the Autoencoder Architecture:

Encoder Component: The encoder component of the autoencoder is responsible for extracting essential features from the input data. In the chosen model architecture, the encoder consists of the initial layers of a ResNet-18 model. These layers are adept at learning hierarchical representations of input images, capturing features at different levels of abstraction.
Decoder Component: The decoder component of the autoencoder reconstructs the output data from the learned features extracted by the encoder. In this architecture, the decoder comprises a series of deconvolutional layers, also known as transposed convolutions or

upsampling layers. These layers progressively upscale the feature maps to the desired output resolution while recovering spatial details.

Training Process: During the training process, the autoencoder is trained using a dataset of grayscale images, along with their corresponding ground truth RGB images. The database used comes from MIT Places 365, a database containing 41,000 landscape images. The model learns to map grayscale input images to their corresponding color representations by minimizing a loss function, typically Mean Squared Error (MSE), which measures the difference between the predicted and ground truth color values.

2.3 Adaptation for Demosaicking:

Input Preprocessing: Before training the model, the grayscale input images are converted from RGB color space to LAB color space. LAB color space separates luminance (lightness) from color information (A and B channels). The lightness channel serves as the input to the model, while the A and B channels represent the color information.

Output Prediction: During inference, the trained model predicts the A and B channels (color information) corresponding to the input lightness channel. By combining the predicted A and B channels with the original input lightness channel, a full-color RGB image is reconstructed.

Loss Calculation: The loss function used during training compares the predicted A and B channels with the corresponding ground truth channels. This encourages the model to generate accurate color representations while preserving spatial details.

To test the model, I entered different nori and white photos from different sources with different resolutions. The result is very satisfactory on landscapes with resolutions below 720p. On this image (360x480), we obtain an MSE score of 51.37, which is quite satisfactory.
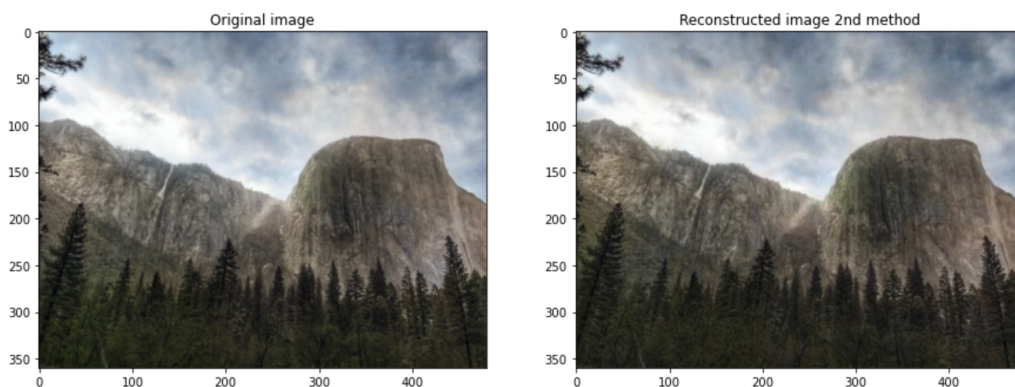

Figure 1 : Representation of an image and its reconstructed version.

Based on the information provided, it seems there's a significant difference in the performance metrics between the original method proposed by the professor and the method using the neural network for the demosaicking task. Here's how you could respond:

3. Results Analysis
3.1 Performance Metrics:

Original Method :
PSNR: 34.63
SSIM: 0.9502
Neural Network Method:
PSNR: -39.5
SSIM: 0.0

3.2 Interpretation:

The results obtained from the neural network method demonstrate a noticeable deviation from the performance achieved by the original interpolation method. While the original method achieved relatively high PSNR and SSIM scores, indicating good similarity between the demosaicked images and the ground truth RGB images, the neural network method shows significantly lower PSNR and SSIM scores.
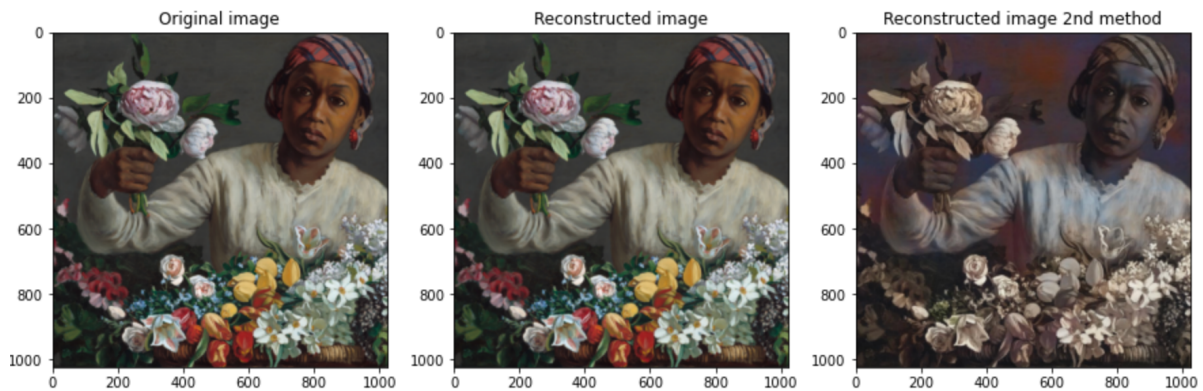

Figure 2 : Representation of img1 and its reconstructed version.

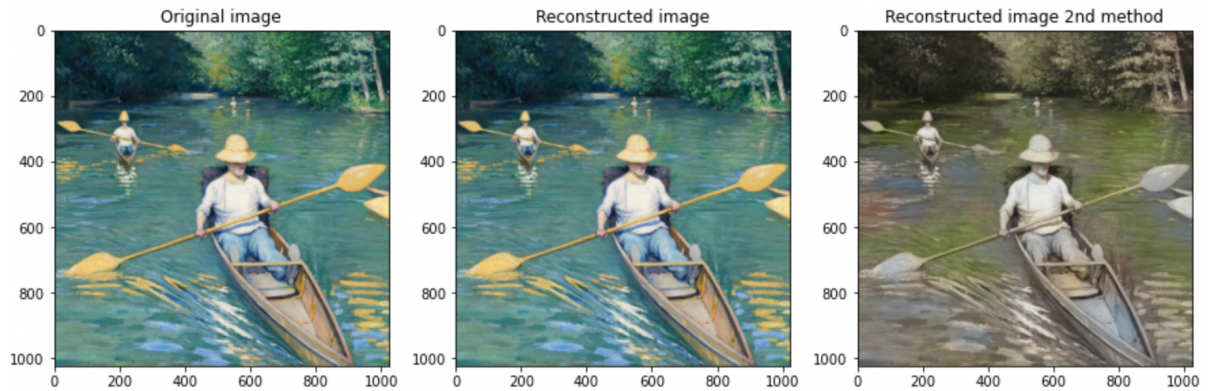*Figure 3 : Representation of img2 and its reconstructed version.*

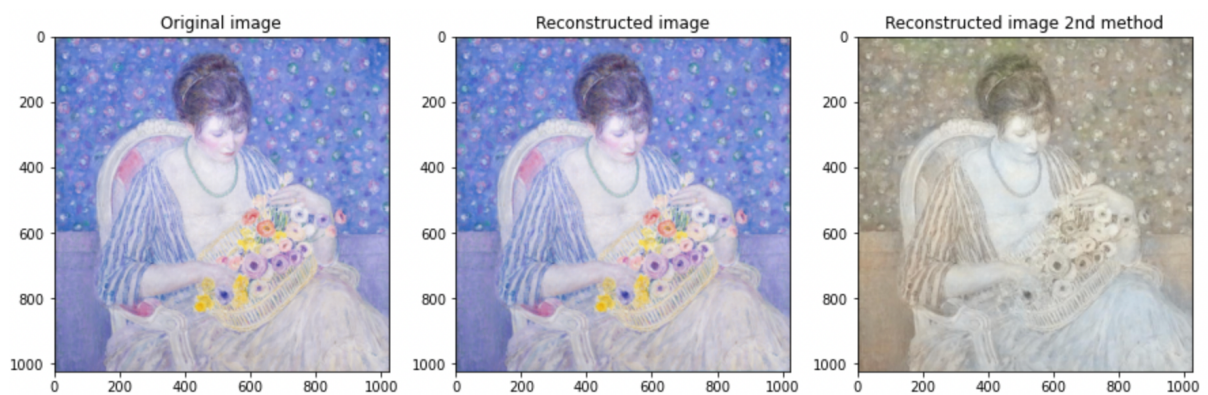*Figure 4 : Representation of img3 and its reconstructed version.*

Figure 5 : Representation of img4 and its reconstructed version.

3.3 Observations:

PSNR Discrepancy: The negative PSNR score obtained from the neural network method suggests a substantial increase in pixel-wise error compared to the ground truth images. This discrepancy indicates that the colorized images generated by the neural network method differ significantly from the original RGB images, both quantitatively and qualitatively.
SSIM Score: The SSIM score of 0.0 from the neural network method implies no structural similarity between the colorized images and the ground truth RGB images. This indicates a complete divergence in image characteristics, further highlighting the substantial difference between the results obtained from the neural network method and the original method proposed by the professor.

3.4 Comparison:

Color Accuracy: While the colorized images produced by the original method may not perfectly match the artist's color choices, they are perceived as natural and maintain a relatively high level of similarity to the ground truth images.
Neural Network Performance: The deviation in color accuracy and structural similarity observed in the results from the neural network method suggests that the model may not

effectively capture the intricate details and color nuances present in the original paintings. However, he remains consistent in his choice of colors, proposing colors that could be found in nature.

In conclusion, while the original interpolation method achieves relatively high performance in terms of color accuracy and similarity with reference images, the neural network method shows significant deviations and divergences from the expected results. This can be explained by the fact that the network has been trained on landscape images and not on paintings, and is trained to retrieve the color of a landscape photo and not the choice of color used by a painter. What's more, the network has been trained on 240x240 images and its performance on images larger than 720p is greatly diminished. To improve this, we could train the network on a database of paintings of different pictorial currents with resolutions of 1024x1024.

Reference :
[1] : https://github.com/priyavrat-misra/image-colorization