Image Colorization Using U-Net-Based CNN

Indiana University Indianapolis

Group-7
Aliza Islam
Mohammed Hossain

Dr. Sunandan Chakraborty, Ph.D
Luddy School of Informatics, Computing, and Engineering
Indiana University Indianapolis

December 17, 2024

Introduction

Image colorization, the process of converting grayscale images to full color, is increasingly important in fields such as historical image restoration, medical imaging, entertainment, and artistic style transfer. Achieving effective colorization depends on several key factors:

- **Realistic Color Generation**: The colors added must be appropriate for the objects depicted, adhering to their natural appearance.
- **Structural Detail Preservation**: To preserve the original structure, it is crucial to retain clarity in the image's edges, textures, and patterns.
- **Adaptability to Color Styles**: The colorization model should flexibly match the colors from a reference image to ensure consistency across various applications.

While traditional convolutional neural networks (CNNs) have significantly automated the colorization process, they struggle with complex textures, and diverse color styles, and often introduce unwanted artifacts. To overcome these challenges, we propose a novel hybrid model that combines:

- **U-Net for Feature Extraction**: This component extracts hierarchical semantic features from grayscale images, providing a solid foundation for accurate colorization.
- **Convolutional Sparse Coding (CSC)**: CSC improves the representation of local textures and enhances detail by promoting sparsity.
- **Neural Style Transfer (NST)** dynamically transfers color styles from reference images to grayscale images, increasing the adaptability and richness of the colorized images.

Our approach integrates the strengths of CNNs, CSC, and NST to produce colorized images that are not only vibrant and appealing but also true to the original structure of the pictures.

Problem Description

The primary challenge in image colorization is accurately predicting RGB colors for each pixel in a grayscale image, which involves addressing several key issues:

- **Color Realism**: The model must ensure that the colors it predicts are realistic and appropriate for the given context, such as ensuring bananas are colored yellow rather than an unrealistic color like red.
- **Structural Consistency**: The model needs to preserve fine details such as edges and textures, which are crucial for maintaining the authenticity of the image.
- **Style Adaptability**: The system should be capable of adapting to different color styles from various reference images, enhancing its usability across different scenarios.

Despite the progress made by existing CNN-based models, they often blur details, lack flexibility in handling various color styles, and can create artifacts in complex areas of images. Our integrated solution enhances feature extraction with U-Net, improves detail preservation through sparse coding, and increases style flexibility with advanced neural style transfer. This comprehensive approach is designed to produce vibrant, context-aware, and detailed colorized images, establishing a new benchmark in the field of image colorization.

Description of the Data

We utilized the Natural-Color Dataset (NCD), curated by Anwar et al. (2020) in our Project.

- Source: NCD GitHub Repository.
- Description: This dataset is richly diverse, comprising thousands of high-resolution RGB images spanning various categories including fruits, plants, and everyday objects. The extensive variety of images is designed to enhance the model's ability to generalize across different visual contexts, making it exceptionally robust for tasks requiring detailed colorization.

Methodology:

To prepare the dataset for training, we performed the following steps:

1. Grayscale Conversion:
   - Each RGB image was converted to a single-channel grayscale image using OpenCV. This serves as the model input.
2. Normalization:
   - Pixel values were scaled to the range [0, 1] to ensure consistent input for the neural network.
3. Resolution Standardization:
   - Images were resized to 256×256 pixels to maintain uniform dimensions across the dataset.

The dataset used for this project consists of two parts:

- Grayscale Images Part: Contains 20 subcategories of images, each representing different objects like fruits and vegetables.
- Color Images Part: Corresponding RGB versions of the grayscale images, providing the ground truth for model training.

The following categories were included in the dataset:

- Categories: Carrot, CapsicumGreen, Peach, Corn, Orange, ChilliGreen, Broccoli, Tomato, Brinjal, Banana, Pomegranate, Strawberry, Potato, Lemon, LadyFinger, Apple, Cherry, Plum, Cucumber, and Pear.

The dimensions of the dataset after preprocessing are as follows:

- Grayscale Images: (696, 256, 256, 1) — 696 grayscale images of size 256×256 pixels with a single channel.
- Color Images: (696, 256, 256, 3) — Corresponding RGB color images with 3 channels (Red, Green, Blue).
- Categories: (696,) — Labels for each image, representing their categories.

We divided the dataset into training and validation sets to ensure effective model training and evaluation:

Training Set:

- Size: 556 images (80% of the dataset).
- Purpose: Used to optimize the model parameters.

Validation Set:

- Size: 140 images (20% of the dataset).
- Purpose: Evaluate the model's performance during training to prevent overfitting.

This dataset provides sufficient diversity across different categories, making it a robust input for training the colorization model.

U-Net Architecture for Image Colorization

The U-Net architecture is a well-established deep learning model originally designed for biomedical image segmentation. Its encoder-decoder structure with skip connections makes it highly effective for image-to-image translation tasks, including image colorization.

The architecture consists of two main parts:

- Encoder (Contracting Path): Captures contextual information and compresses the input image into a smaller, feature-rich representation.
- Decoder (Expanding Path): Reconstructs the image to its original dimensions while incorporating spatial information from the encoder using skip connections.

Layer-by-Layer Description

The U-Net architecture in this project is summarized below:

Input Layer

- Shape: (256, 256, 1)
- Accepts grayscale input images.

Encoder Path

The encoder progressively reduces the image dimensions while extracting features using convolutional layers and MaxPooling2D operations. Each convolutional block includes:

- Two Conv2D Layers: Each with a kernel size of (3×3), ReLU activation, and same padding to preserve dimensions.
- MaxPooling2D Layer: Reduces the spatial resolution by half.

Decoder Path

The decoder reconstructs the original image using UpSampling2D and concatenates feature maps from the encoder to preserve spatial details. Each block consists of:

- UpSampling2D: Doubles the spatial dimensions.
- Conv2D Layer: Reduces the number of filters and refines features.
- Concatenate: Merges encoder and decoder features for better reconstruction.

Output Layer

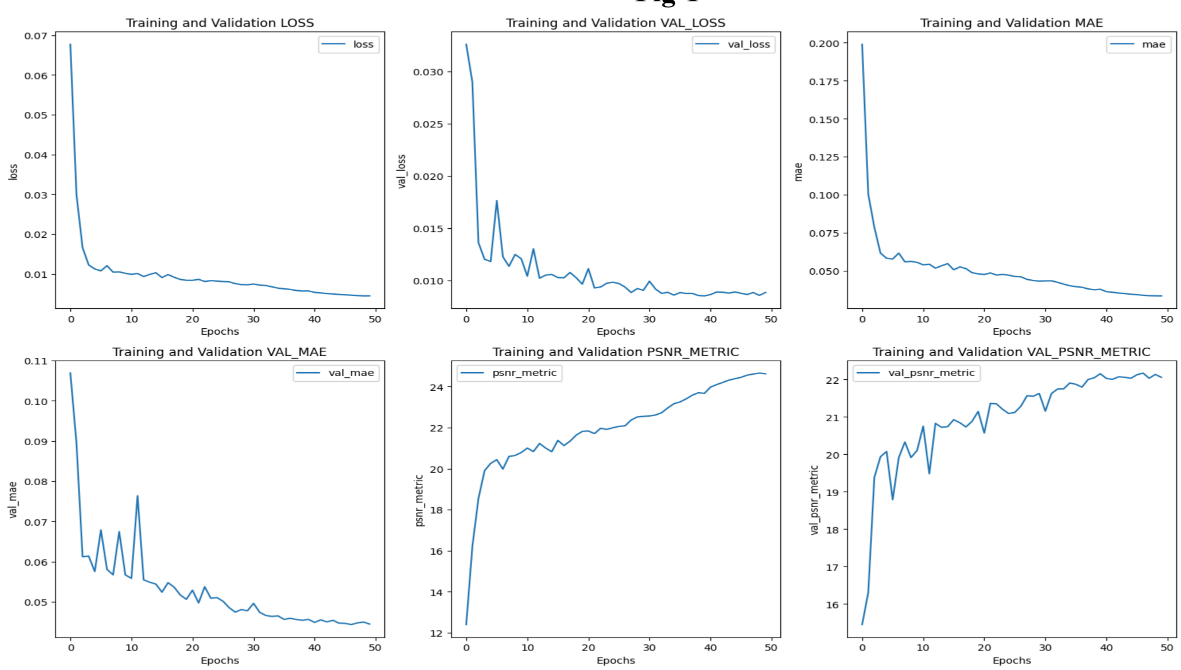
- Shape: (256, 256, 3)
- Activation: Sigmoid
- Purpose: Produces pixel-wise RGB color predictions scaled to [0, 1].

Training Configuration

The model was trained with the following setup:

- Loss Function: Mean Squared Error (MSE).
- Optimizer: Adam optimizer with an initial learning rate of 0.001.
- Batch Size: 16
- Epochs: 50 with early stopping to prevent overfitting.

Fig-1



The graphs in Fig-1 show the evolution of loss and accuracy metrics over 50 training epochs. A consistent decrease in loss and improvements in MAE and PSNR metrics demonstrate the model's enhanced ability to colorize images accurately. This validation process confirms the model's effectiveness and reliability in producing high-quality outputs. We closely monitored each training phase to prevent overfitting and ensure the model's robust generalization to new data.

Evaluation Metrics

- Mean Absolute Error (MAE): Pixel-wise color prediction accuracy.
- Peak Signal-to-Noise Ratio (PSNR): Measures the fidelity of the output image.
- Structural Similarity Index (SSIM): Ensures perceptual similarity between the ground truth and the generated output.

Results

Our model showed consistent improvement over epochs. Final evaluation metrics:

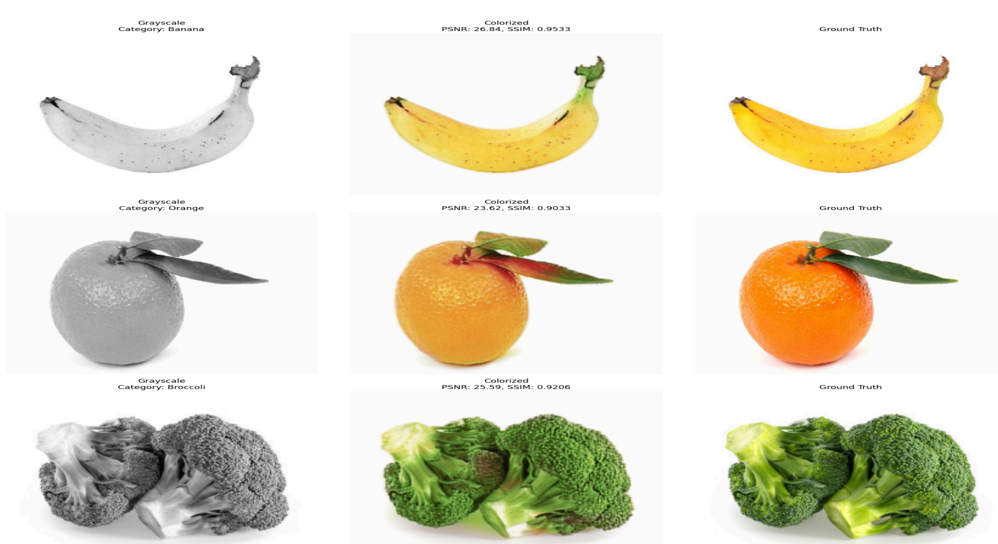
Metric	Training Set	Validation Set
MSE	0.0045	0.0086
PSNR (dB)	24.6	22.1
SSIM	0.92	0.89

Colorized outputs closely match the ground truth:

Grayscale Input	Generated Output	Ground Truth
Input	Output	GT

Following the detailed metrics in the tables, Fig-2 effectively demonstrates the model's capability to colorize a variety of objects. This section visually compares grayscale inputs with their colorized outputs alongside the actual ground truth images, clearly showing the model's accuracy in achieving natural color reproduction. The results illustrate not only technical accuracy but also the model's effective generalization across different items, confirming its practical utility in real-world applications.

Fig-2



Discussion

The U-Net architecture has proven effective in our image colorization project, especially highlighted by the improvements in MAE and PSNR metrics over the 50 epochs of training. While it excels in color accuracy and image fidelity, the use of Neural Style Transfer (NST) introduces significant challenges. Notably, NST can sometimes distort object-specific colors, leading to inaccuracies that deviate from the original hues. To counteract this issue, we employed linear color transfer techniques, enhancing the model's ability to maintain true color representation.

Also, the computational demand of the model scales significantly with its complexity. This has necessitated the implementation of optimization strategies such as network architecture refinement and model pruning. These steps are crucial to managing the computational resources effectively without compromising the quality of the colorization. Moving forward, continuous improvements and adaptations will be essential to address these limitations, ensuring that the model remains both effective and efficient in practical applications.

Future Work: User Studies

In response to the processor's inquiry during our presentation, we acknowledge the necessity of incorporating user studies into our future work. These studies will involve participants evaluating the realism and aesthetic appeal of our colorized images, providing invaluable insights into user perception and satisfaction. By comparing user feedback with our objective metrics, we can refine our model to better meet user expectations and enhance the practical application of our technology in real-world scenarios. This step will be crucial for validating and improving the user experience and ensuring the model's adaptability to diverse user needs.

References

- Saeed-Anwar. (n.d.). GitHub - saeed-anwar/ColorSurvey: This repository is for "Image Colorization: A Survey and Dataset" paper. GitHub. https://github.com/saeed-anwar/ColorSurvey

- Nathanael, O. T., Prasetyo, S. Y., & Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia. (2024). Color and Attention for U: Modified Multi Attention U-Net for a Better Image Colorization. In INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION (Vol. 8, Issue 3, pp. 1453–1459) [Journal-article]. https://www.joiv.org/index.php/joiv

- Wu, Y., Wang, X., Li, Y., Zhang, H., Zhao, X., Shan, Y., & Applied Research Center (ARC), Tencent PCG. (2021). Towards Vivid and Diverse Image Colorization with Generative Color Prior. In Inputs [Journal-article]

- Vitoria, P., Raad, L., Ballester, C., Department of Information and Communication Technologies, & University Pompeu Fabra, Barcelona, Spain. (2021). ChromaGAN: Adversarial Picture Colorization with Semantic Class Distribution [Journal-article]

- Huang, S., Jin, X., Jiang, Q., & Liu, L. (2022). Deep learning for image colorization: Current and future prospects. Engineering Applications of Artificial Intelligence, 114, 105006. https://doi.org/10.1016/j.engappai.2022.105006

Appendix : Contributions from Each Member

Team Member	Contribution
Mohammed Hossain	Data preparation, U-Net integration, loss function design.
Aliza Islam	NST implementation, model training, evaluation, and documentation.