IMAGE COLOURIZATION VIA CONVOLUTIONAL NEURAL NETWORKS AND DEEP LEARNING

Youssef Fikry
Student# 1005678901
youssef.fikry@mail.utoronto.ca

Harkirpa Kaur
Student# 1011242479
harkirpa.kaur@mail.utoronto.ca

Peter Leong
Student# 1005678901
peter.leong@mail.utoronto.ca

Thulasi Thavarajah
Student# 10115358424
t.thavarajah@mail.utoronto.ca

ABSTRACT

This project addresses the challenge of automated colourization for 256×256 grayscale images using a dataset of 12,600 image pairs, balanced across human subjects, animals, and natural scenery. We frame colourization as a supervised learning problem in the CIELAB colour space, where a model predicts chrominance channels (a^*, b^*) from the luminance channel (L^*). A shallow convolutional neural network (CNN) provides the baseline performance, while our primary solution employs a deeper convolutional encoder-decoder architecture. This design captures high-level semantic features and spatial context, addressing limitations of shallow networks in perceptual realism. All source code, datasets, and results are publicly available here. —-Total Pages: 3

1 BRIEF PROJECT DESCRIPTION

The invention of photography provided the technology to capture a moment in time. However, for most of photographic history, the process of obtaining coloured images eluded photographers (Science & Museum, 2020). As a result, much of historic photography is grayscale; lacking the visual richness found in modern photography and is inaccessible to individuals with vision impairments (Vogelsang et al., 2024). This project aims to leverage deep learning to automate the colorization of grayscale images, thereby aiding in the revitalization of grayscale photographs. Furthermore, image colorization technology assists archivists and museums in restoring lost visual information and has applications in the media, medical and geospatial industries.

Traditional image colorization methods involve manual labour that is costly and time-consuming. On the other hand, deep learning approaches automate the colorization process and supplant traditional techniques. In machine learning, colorization is defined as a model taking a black-and-white image as an input, and outputting its coloured counterpart. From a machine learning perspective, deep convolutional neural networks are best suited for this task as they can extract and learn features such as colors, patterns and shapes in images and affiliate them with object classes. These characteristics aid CNNs in excelling at object classificition tasks, as well as image colorization.

//insert image here.

2 INDIVIDUAL CONTRIBUTIONS & RESPONSIBILITIES

3 BASELINE MODEL

To establish a performance benchmark for our colorization task, we implemented a shallow convolutional neural network called `ShallowColorizationNet`. This model uses a simple encoder-bottleneck-decoder architecture and does not incorporate adversarial training. Its purpose is to test the feasibility of learning a direct mapping from grayscale luminance (L) input to the chrominance (ab) channels in the CIELAB color space.

3.1 Model Architecture

The architecture of the baseline model is as follows:

- **Encoder:** Two convolutional layers with ReLU activation followed by max pooling, reducing spatial dimensions while increasing channel depth.
- **Bottleneck:** A single convolutional layer that expands the feature representation.
- **Decoder:** A transposed convolution to restore spatial resolution and a final convolutional layer outputting two channels (for a and b), with a $\tanh$ activation to constrain values between $[-1, 1]$.

3.2 Training and Comparison Approach

The baseline model was trained using the L1 loss between predicted and ground truth ab channels. This model was compared to our primary neural network—a conditional GAN model that includes a discriminator to encourage more realistic colorizations—based on both quantitative and qualitative results.

3.3 Quantitative and Qualitative Results

Quantitative: The baseline model was trained for 5 epochs using L1 loss. The generator's loss remained stable around **1.5647**, while the discriminator's loss converged to approximately **0.7074**. These consistent values suggest that the baseline network was able to learn a basic mapping from grayscale to chrominance, but likely lacked the capacity or incentive to produce high-fidelity or diverse outputs. These values serve as reference points for evaluating improvements made by our primary GAN-based model.

Qualitative: The baseline model produced smooth but somewhat desaturated colorizations. It tended to predict average colors in ambiguous regions, resulting in low color diversity. See Figure 1 for representative examples.

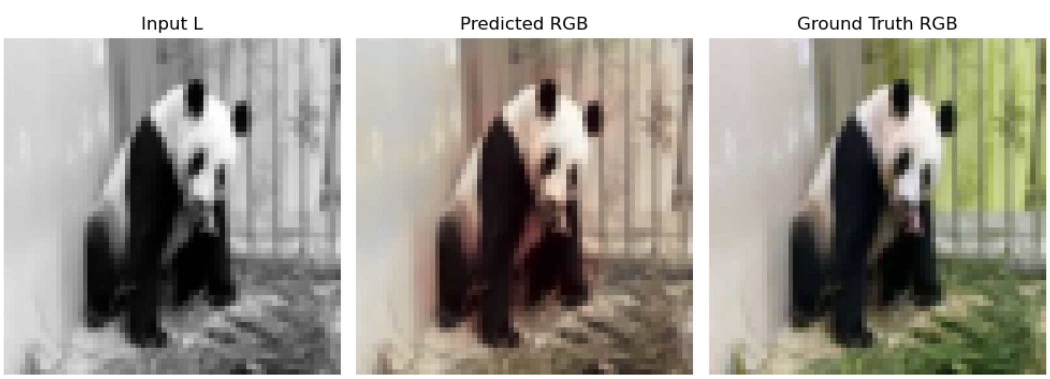

Figure 1: Sample output from the baseline model on grayscale input images.

3.4 Challenges

The primary challenge encountered during development was the tendency of the baseline model to learn overly conservative color predictions. Without a discriminator or explicit diversity loss, the network favored colorizations close to the dataset mean. Additionally, selecting the appropriate normalization strategy for LAB space and constraining output to valid chrominance ranges required careful tuning.

3.5 Feasibility Assessment

Despite its simplicity, the baseline model demonstrated that the colorization task is learnable with a low-capacity network, although the outputs lacked the vividness and fidelity achieved by our primary GAN-based model. The baseline thus served its role in confirming the viability of end-to-end colorization from grayscale input and establishing a benchmark for model improvement.

4 Data Processing

5 Baseline Model

6 Primary Model

REFERENCES

National Science and Media Museum. A short history of colour photography, Jul 2020. URL `https://www.scienceandmediamuseum.org.uk/objects-and-stories/history-colour-photography`.

Marin Vogelsang, Lukas Vogelsang, Priti Gupta, Tapan K. Gandhi, Pragya Shah, Piyush Swami, Sharon Gilad-Gutnick, Shlomit Ben-Ami, Sidney Diamond, Suma Ganesh, and Pawan Sinha. Impact of early visual experience on later usage of color cues. *Science*, 384(6698):907–912, 2024. doi: 10.1126/science.adk9587. URL `https://www.science.org/doi/abs/10.1126/science.adk9587`.