

# AI GreyScale Image Colorization

Course Code & Name: CSC301 - Seminar

Prepared by:

Khwaish Gahoi 22csu363

Paras Chawla 22csu337

Monalika Goel 22csu246

Instructor: Dr. Garima

Date: April 23­­, 2025

# THE NORTHCAP UNIVERSITY

Gurugram, India

## Index

1. Abstract-------------------------------------------------------------------------- Page 3
2. Introduction--------------------------------------------------------------------- Page 3
3. Related Work ------------------------------------------------------------------ Page 4
4. Tech Stack -----------------------------------------------------------------------Page 5
5. Implementation---------------------------------------------------------------- Page 5
6. Results ---------------------------------------------------------------------------Page 9
7. Conclusion ----------------------------------------------------------------------Page 11
8. References ----------------------------------------------------------------------Page 11

## ABSTRACT

This project investigates an AI-based approach to image colorization, aiming to transform greyscale images into visually coherent and naturally colored photographs. The study focuses on the colorization of mountain and forest imagery, domains characterized by dominant blue and green tones. Due to the computational intensity of training on the complete 25GB publicly available dataset, a curated subset comprising 10,000 images (5,000 each from mountainous and forested regions) was utilized. This strategic dataset selection was designed to facilitate model proficiency in rendering sky and vegetation hues while ensuring training efficiency. A lightweight convolutional neural network (CNN) architecture, consisting of three convolutional layers and enhanced through ReLU activations, dropout, batch normalization, and image augmentation techniques, was employed. The dataset was partitioned into 90:5:5 for training, validation, and testing, respectively. Experimental results demonstrate the model's capability to produce high-quality colorized outputs with domain-appropriate accuracy. The findings underscore the potential of domain-specific training and resource-efficient architectures in addressing the colorization problem effectively.

## INTRODUCTION

Image colorization refers to the process of inferring and adding plausible color information to greyscale images. It holds significance in a range of applications, including historical image restoration, digital media production, and satellite imagery enhancement. Traditional methods heavily relied on user guidance or hand-crafted rules, limiting scalability and adaptability. However, recent advancements in deep learning have enabled automated and data-driven approaches, allowing models to learn intricate mappings between luminance and chrominance components.

The present study addresses the colorization of natural landscape images, particularly those depicting mountainous and forested environments. These categories present unique challenges due to the presence of diverse textures and subtle color variations, predominantly in blue and green spectrums. To achieve better contextual understanding and color fidelity, a custom dataset comprising 10,000 greyscale images, equally distributed between mountain and forest scenes, was created. This dataset design was not a constraint but a deliberate decision to enhance the model’s specialization in the targeted domains.

A total of 9,000 images were used for training, while 500 images each were allocated for validation and testing. The use of a lightweight CNN architecture, designed with computational efficiency in mind, aimed to demonstrate the feasibility of high-performance colorization without reliance on large-scale computing infrastructure. Key components such as dropout, batch normalization, and data augmentation were integrated to improve generalization and robustness.

This report outlines the existing work in image colorization, the technological stack employed, the model design and implementation, experimental outcomes, and conclusions drawn from the analysis. The primary contributions include the creation of a domain-focused dataset, the development of an efficient neural architecture, and a comprehensive evaluation of performance across the target image categories.

**RELATED WORK**

Numerous studies have explored the task of image colorization, with early methods relying on user inputs or handcrafted features to guide color propagation. Levin et al. (2004) proposed an optimization-based approach that required scribble inputs, while Welsh et al. (2002) introduced a texture-based method that transferred color from a reference image to a greyscale target. Although effective for their time, these techniques were often labor-intensive and lacked generalization.

The advent of convolutional neural networks (CNNs) marked a significant advancement in colorization. Zhang et al. (2016) introduced a CNN-based method that formulated colorization as a classification task in the ab color space of CIELab, achieving realistic results with minimal human intervention. Iizuka et al. (2016) extended this by incorporating global and local features, allowing more context-aware predictions. Other approaches, such as the use of generative adversarial networks (GANs), have further improved color realism by learning to produce perceptually convincing images.

Recent developments include frameworks such as **DeOldify**, which utilizes GANs and attention mechanisms to produce high-quality colorization. DeOldify has gained recognition for its effectiveness in restoring old black-and-white photographs and videos. Another influential model, **Colorful Image Colorization** by Zhang et al., was originally implemented using the Caffe framework and laid the foundation for colorization as a classification task rather than a regression problem. These tools demonstrate how advancements in deep learning, particularly through open-source ecosystems, have accelerated progress in this domain.

In addition, domain adaptation and semantic segmentation have been investigated to improve color accuracy, with scene-specific training proving advantageous in certain applications. This principle aligns with the present study’s strategy of curating a focused dataset comprising mountainous and forest scenes.

**TECH STACK**

The technical stack for this project was selected to balance performance with resource efficiency. Python was used as the primary programming language, leveraging several key libraries for data processing, model construction, and evaluation.

* **Framework**: TensorFlow served as the deep learning framework due to its flexibility allowing rapid experimentation.
* **Data Handling**: NumPy and OpenCV were employed for image preprocessing and manipulation. Data Augmentation was also performed on the images for better understanding for the model.
* **Model Architecture**: A custom CNN model was built using TensorFlow modules, incorporating convolutional layers, ReLU activation functions, dropout for regularization, and batch normalization for stable training.
* **Training and Evaluation**: The training loop utilized cross-entropy loss in the CIELab color space, with accuracy and visual inspection as key evaluation metrics. Matplotlib and Seaborn were used to visualize training metrics and output comparisons.
* **Hardware and Environment**: The computational environment was managed using Google Colab and Jupyter Notebooks for ease of experimentation and visualization.
* **Web Application**: Stream lit used to develop the web application for deploying the model.

This combination of tools enabled the effective development, training, and evaluation of a robust yet lightweight image colorization model tailored to natural landscapes.

**IMPLEMENTATION**

1. **A. Dataset Construction and Domain Focus**

To facilitate efficient training and to ensure domain relevance, a focused dataset comprising 10,000 high-resolution images was assembled. This dataset includes 5,000 images representing mountainous terrain and 5,000 depicting forested environments. These categories were selected due to their prevalence in natural imagery and their distinct color profiles, primarily comprising blue and green hues. The curated dataset was drawn from publicly available image repositories and resized to a uniform dimension to maintain consistency across inputs.

Places 365 Dataset link: [Click here](https://www.kaggle.com/datasets/benjaminkz/places365)

Subset dataset link: <https://drive.google.com/drive/folders/1SCd1MvrHPpFX5jHySAKZfVure5XGhEMX?usp=sharing>

**B. Data Preprocessing Pipeline**

Each RGB image was initially converted into the CIELab color space to decouple the luminance (L) channel from the chrominance (a and b) channels. The L channel served as the grayscale input to the model, while the a and b channels were treated as prediction targets. This transformation allowed the model to focus solely on learning color mappings, independent of brightness information.

All images underwent normalization and resizing to 256x256 pixels. Standard preprocessing also included grayscale conversion and tensor conversion to enable efficient GPU processing.

**C. Data Augmentation Techniques**

To increase model robustness and simulate a broader range of environmental conditions, the training set was augmented using the following transformations:

Random horizontal and vertical flips to enhance spatial invariance.

Random crops and rescaling to simulate camera zoom variation.

Color jittering (brightness, contrast, saturation) applied to the chrominance channels during reverse transformation to improve generalization.

1. **Architectural Design of the Convolutional Neural Network**

**A. Model Overview**

The core architecture was designed as a lightweight convolutional neural network (CNN) tailored to operate efficiently on mid-range GPUs. The model architecture comprises three convolutional layers, interleaved with batch normalization and dropout layers to mitigate overfitting and stabilize learning.

**B. Layer-wise Description**

Input Layer: Accepts a single-channel grayscale image (L channel).

Convolutional Block 1: 64 filters with a kernel size of 3x3, followed by ReLU activation, batch normalization, and 25% dropout.

Convolutional Block 2: 128 filters, similar activation and regularization sequence.

Convolutional Block 3: 256 filters, culminating in a final convolutional layer that outputs two channels corresponding to a and b chrominance values.

The final output is reshaped and combined with the input L channel to reconstruct a full-color image in the CIELab space.

1. **Training Methodology, Hyperparameter Tuning, and Optimization**
2. **Training Configuration**

The model was trained over 50 epochs using the Adam optimizer, selected for its adaptive learning capabilities and efficient convergence. An initial learning rate of 0.001 was adopted, with scheduled decay applied upon plateau detection in validation loss.

1. **Loss Function**

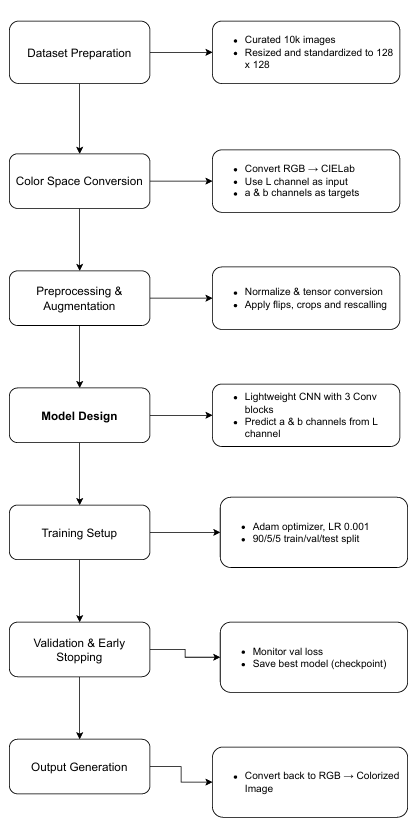
Cross-entropy loss was employed as the primary objective function, operating over quantized bins of the a and b channels. This formulation aligns with prior successful approaches that treat colorization as a classification problem in color space.

1. **Data Partitioning**

The dataset was split into training (90%), validation (5%), and test (5%) subsets. This stratification ensured representative sampling across both domains while maintaining sufficient data for reliable evaluation.

1. **Checkpointing and Early Stopping**

To avoid overfitting, model checkpoints were saved based on minimum validation loss. Early stopping was implemented to terminate training upon stagnation of performance over a defined patience interval.



**RESULTS AND EVALUATIONS**

The performance of the proposed CNN-based colorization model was evaluated through both qualitative and quantitative assessments on the test set comprising 500 greyscale images (L channel in Lab color space) across mountainous and forested categories and predicts the corresponding color information (ab channels), which are then combined to reconstruct the full color image.

### Quantitative Evaluation

The model was benchmarked using three standard metrics: **Mean Squared Error (MSE)**, **Peak Signal-to-Noise Ratio (PSNR)** and **Structural Similarity Index Metric (SSIM)** all these are calculated in the CIELab color space. These metrics provide objective measures of reconstruction fidelity in terms of chrominance prediction.

* **MSE** was computed as the average squared difference between the predicted and ground truth a/b channels. The model achieved an **MSE of 0.0075**, indicating minimal deviation in chromatic reconstruction.
* **PSNR** was calculated to evaluate the ratio between the maximum possible power of a signal and the power of corrupting noise. The model attained an average **PSNR of 29.00 dB**, which reflects high visual fidelity and low reconstruction noise.
* **SSIM was computed** for evaluating the similarity between two images. The model procured a **SSIM** of **0.8930**, indicates a good degree of similarity between two images, but not a perfect match.

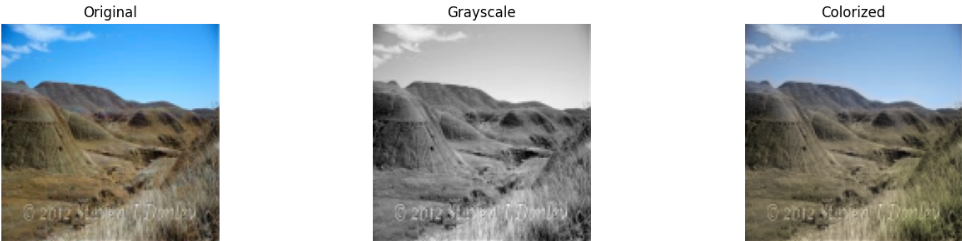
Validation loss consistently decreased over training epochs without significant fluctuations, suggesting effective generalization and absence of overfitting. Training convergence was stable, with the best model checkpoint selected based on minimum validation loss.

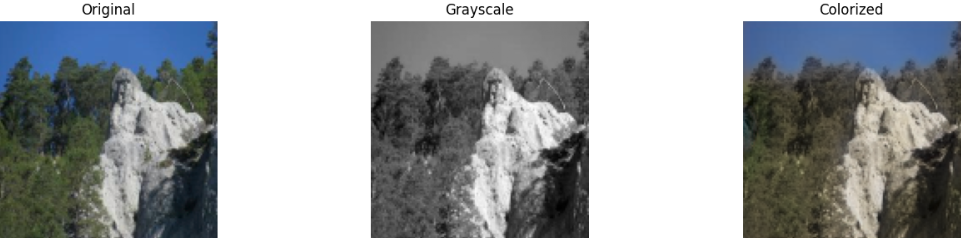
### Qualitative Evaluation

To visually assess the performance, the model's predictions on 10 unseen test images were compared with their ground truth counterparts. Each test sample includes:

* The original color image (for reference)
* The grayscale input (L channel)
* The colorized output generated by the model

The visualizations demonstrate that the model is effective in restoring natural-looking colors, especially in distinguishing between different scene types (e.g., blue sky vs. green vegetation). While there may be occasional artifacts or slightly desaturated outputs, the model captures the general color distribution and spatial coherence well.



**CONCLUSION**

This study presents a focused approach to image colorization by leveraging a lightweight convolutional neural network (CNN) trained on a domain-specific dataset comprising mountainous and forest landscapes. By constraining the training data to visually coherent environments, the model was able to learn robust mappings between luminance and chrominance components, achieving high visual fidelity in the reconstructed color images.

The key contributions of this work include: (1) the construction of a targeted dataset curated for domain-specialized learning, (2) the development of an efficient CNN architecture capable of generalizing under limited computational resources, and (3) the demonstration of strong colorization performance validated through the quantitative metrics (MSE, PSNR and SSIM) and qualitative visual assessment.

The experimental results affirm that high-quality colorization does not necessitate large-scale architectures or vast computational infrastructure. Instead, domain-specific training, coupled with architectural simplicity and regularization techniques, can yield compelling outcomes. Notably, the model was able to consistently reproduce realistic sky blues and vegetation greens, suggesting successful internalization of spatial and textural cues inherent to the dataset.

Future directions may include expanding the domain coverage to incorporate urban or aquatic environments, integrating semantic segmentation to improve spatial color coherence, and exploring transfer learning techniques to accelerate model adaptation to new visual contexts. Furthermore, augmenting the perceptual evaluation with user studies could provide a more holistic measure of realism and user satisfaction.

In summary, this work illustrates the efficacy of combining domain-focused data curation with computationally efficient deep learning architectures to solve the colorization problem, paving the way for scalable and adaptable applications in visual media restoration and enhancement.

**REFERENCES**

1. Levin, A., Lischinski, D., & Weiss, Y. (2004). Colorization using optimization. ACM Transactions on Graphics (TOG), 23(3), 689–694.
2. Welsh, T., Ashikhmin, M., & Mueller, K. (2002). Transferring color to greyscale images. ACM Transactions on Graphics (TOG), 21(3), 277–280.
3. Zhang, R., Isola, P., & Efros, A. A. (2016). Colorful image colorization. In European Conference on Computer Vision (ECCV).
4. Iizuka, S., Simo-Serra, E., & Ishikawa, H. (2016). Let there be color!: Joint end-to-end learning of global and local image priors for automatic image colorization. ACM Transactions on Graphics (TOG), 35(4), 1–11.
5. DeOldify. (2020). https://github.com/jantic/DeOldify
6. Zhang, R., Isola, P., & Efros, A. A. (2016). Colorful image colorization (Caffe implementation). https://github.com/richzhang/colorization