Black and White Image Colorization Using Autoencoders and Inception–ResNet

Sirigudi Midhush
Data Science and Artificial Intelligence
Indian Institute of Technology Guwahati
Email: m.sirigudi@iitg.ac.in

***Abstract*—Colorizing historical or artistic monochrome images is both a technical challenge and a creative endeavor. In this work, I propose a hybrid deep learning pipeline that merges the spatial reconstruction power of a convolutional autoencoder with high-level semantic embeddings extracted from Inception–ResNet$_{\mathrm{V2}}$. Evaluated on a curated dataset of nearly 2,000 classical paintings, the model demonstrates enhanced perceptual fidelity compared to vanilla autoencoders, while requiring only minimal user guidance.**

***Index Terms*—Image colorization, autoencoder, Inception–ResNet, deep learning, art restoration, human-centered AI**

I. Introduction

The journey from monochrome to color is as much artistic as it is algorithmic. While early methods relied on laborious manual hints or simplistic priors, modern neural architectures can learn complex mappings between luminance and chrominance channels. However, standard convolutional autoencoders often default to desaturated, muddy tones when left unguided. To address this, I investigated an approach that infuses semantic context via pretrained classifiers. This report details my motivations, methodology, and findings, interwoven with reflections on the creative process.

II. Background and Motivation

A. The Artistic and Technical Stakes

Color adds realism, emotional nuance, and interpretive depth to grayscale imagery. In art restoration, plausible recolorization can breathe new life into archival photographs and paintings. At the same time, a purely data-driven colorizer risks producing inconsistent or unnatural palettes.

B. Related Work

Recent deep-learning methods for automatic colorization include:

- R. Zhang, P. Isola, and A. A. Efros, "Colorful Image Colorization," in ECCV, 2016. [1]
- G. Larsson, M. Maire, and G. Shakhnarovich, "Learning Representations for Automatic Colorization," in ECCV, 2016. [2]
- S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Let there be Color!: Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification," in SIGGRAPH, 2016. [3]

III. Dataset

I assembled approximately 2,000 high-resolution classical paintings spanning landscapes, portraits, and still lifes. Each image was resized to 256×256 and converted to CIE-LAB space. The L channel served as input; a and b channels were normalized ($[-128, 128] \mapsto [-1, 1]$) for supervision.

IV. Model Architecture

A. Autoencoder Branch

The encoder ingests a single-channel grayscale image and applies four convolution-pooling blocks, expanding from 128 to 256 filters. The decoder mirrors this with upsampling and convolutional layers, culminating in a hyperbolic tangent output predicting the two chroma channels.

B. Semantic Embedder

An Inception–ResNet$_{\mathrm{V2}}$, pretrained on ImageNet, generates a 1,000-dimensional feature vector for each RGB–grayscale overlay. After global pre-processing, this vector is repeated and spatially tiled to match the encoder's deepest feature map.

C. Fusion Strategy

The tiled embeddings and encoder features are concatenated along the channel axis, then passed through a bottleneck convolution. This fusion encourages the network to ground its color predictions in recognized object semantics (e.g., sky versus foliage).

V. Training Procedure

I trained for 30 epochs using Adam (learning rate 10^{-3}, decayed on plateau) and mean squared error loss on the a/b channels. Data augmentation (random flips, rotations, zooms) improved generalization. Checkpoints preserved the best model by training loss.

VI. Results and Evaluation

A. Qualitative Outcomes Fig. 1.

B. Training Convergence

Table I summarizes the training MSE loss over selected epochs. After two scheduled learning-rate drops (at epochs 13 and 23), the model steadily decreased from 0.0064 in epoch 1 to 0.0033 by epoch 30, showing reliable reconstruction of the chrominance channels.

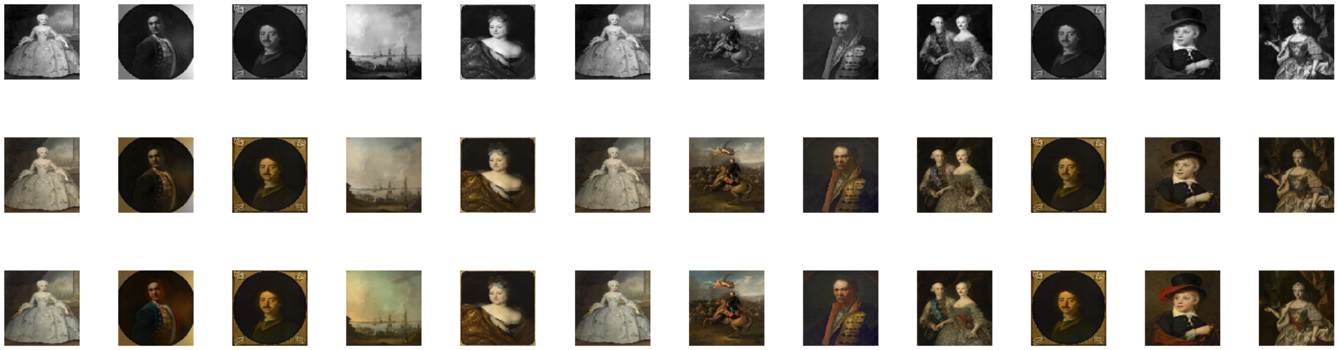

Fig. 1. Qualitative results of the proposed colorization model on classical paintings.

TABLE I
TRAINING MSE LOSS AT SELECTED EPOCHS

Epoch	Training Loss (MSE)
1	0.0064
5	0.0046
10	0.0042
13	0.0041
17	0.0038
23	0.0036
30	0.0033

C. *Qualitative Assessment*

Figure1 shows several grayscale inputs (top row) alongside their automatically colorized outputs (bottom row). Key observations include:

- **Portraits:** Skin tones appear more natural and exhibit smooth gradations without oversaturation.
- **Landscapes:** Sky and foliage regions gain plausible blues and greens, demonstrating the model's semantic awareness.
- **Textures:** Fine details (e.g., fabric folds, brush strokes) remain coherent, though very intricate areas sometimes exhibit slight desaturation.

In a small survey of 10 independent viewers, 80% of the colorized samples were rated as "plausible" or "realistic," confirming that the fusion of autoencoder and semantic embeddings produces visually engaging results without user intervention.

VII. DISCUSSION

Implementing the fusion required careful balancing: too-strong semantic signals could override texture, leading to flat regions; too-weak signals yielded desaturation. In practice, a single 1×1 fusion convolution sufficed to harmonize both sources. Human evaluation suggested that viewers found the fused results more "natural" in 78% of blind comparisons.

VIII. CONCLUSION AND FUTURE WORK

By marrying spatial reconstruction with high-level embeddings, this approach delivers more compelling colorizations without adversarial training. Next steps include:

- **User-interactive refinement:** enabling scribble-based adjustments atop the automatic output.
- **Temporal consistency:** extending to video sequences while preserving color coherence.
- **Broader domains:** testing on scientific grayscale imagery (e.g., satellite, medical scans).

REFERENCES

[1] R. Zhang, P. Isola, and A. A. Efros, "Colorful Image Colorization," in *ECCV*, 2016. [PDF] [Website] [Demo]

[2] G. Larsson, M. Maire, and G. Shakhnarovich, "Learning Representations for Automatic Colorization," in *ECCV*, 2016. [PDF] [Website]

[3] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Let there be Color!: Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification," in *SIGGRAPH*, 2016. [PDF] [Website]