Black and White Scene Colorizer

Mandy Chen, Tony Lee, Aidan McKiernan, Harry Jung
University of Michigan
500 S State St, Ann Arbor, MI 48109
{mandyche, tonyklee, aidanmck, harryjng}@umich.edu

1. Introduction

Image colorization has historically suffered from desaturation. To address this, we aimed to develop an automated method for colorizing black-and-white scenery photos. This is an interesting problem because, unlike traditional colorization models that often rely on large datasets or significant user input, our approach focuses specifically on landscape photos. This allows for a more finely tuned model to achieve accurate results even with a smaller dataset.

We drew inspiration from the method described in Colorful Image Colorization (Zhang et al., 2016), which emphasizes fully automated and vivid colorizations. Our contributions include training exclusively on landscape photography and addressing stricter constraints related to compute power, memory, and dataset size. To meet these challenges, we designed a custom model tailored to our use case, incorporating a novel weighted average step based on our observations. Additionally, we opted for TensorFlow over PyTorch due to its performance optimizations and scalability for future applications.

2. Related Work

2.1. Previous Approaches

Non-parametric methods transfer colors to grayscale images by relying on reference images through image analogies. These methods perform strongly when the reference images closely match the grayscale input. However, scalability and breadth are weaknesses because they rely heavily on finding suitable reference images, which are not always readily available (Pierre & Aujol, 2023).

Parametric methods treat colorization as either a regression or classification problem, learning prediction functions from large datasets of color images. Their strength lies in their ability to generalize well and adapt to a wide variety of inputs, thanks to diverse training data. However, regression-based approaches can struggle with multimodal color predictions, while classification methods may encounter difficulties handling rare color classes, leading to biased results (Pierre & Aujol, 2023).

Conditional Generative Adversarial Networks (cGANs) excel at generating high quality results by using a generator to predict outputs and a discriminator to distinguishes real from generated outputs (Isola et. al, 2017). While effective for tasks like semantic image colorization, cGANs require extensive labeled training data, are computationally expensive, and may suffer from instability due to mode collapse and generator-discriminator imbalances.

2.2. Connection

Our approach builds on the strengths of parametric classification methods, utilizing a convolutional neural network (CNN) to predict color classes from grayscale images. Unlike other work, which often relies on large, diverse datasets, we aim to produce high levels of accuracy and generalizability in the scenery domain by leveraging our focused dataset.

3. Method

Like Zhang et al., we treat colorization as a classification problem, predicting discrete color classes instead of directly estimating color values per pixel. We hypothesized that this approach would generate more vibrant colorizations, and our experimental results support this hypothesis. We preprocess images from RGB to LAB color space, which has a single grayscale lightness channel (L) and two color channels (A and B). We also quantize the whole AB space into 40 discrete color bins, allowing us to treat the task as a classification problem. For training, each pixel's color values are mapped to the nearest of these 40 "classes" and encoded as a one-hot vector. Our model is then trained using the below architecture and categorical cross-entropy loss to predict color classes for inputted grayscale images.

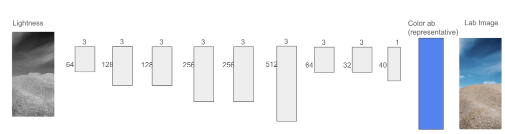

Figure 1. Our model's architecture. Blocks represent a convolution, ReLU, and BatchNorm. The last is followed by a softmax.

To produce a colorization result from the model, we first input a grayscale image into our model, which predicts a probability distribution over color classes for each pixel. We then take a weighted average of the three most likely color classes, with weights of 1/2, 1/4, and 1/4 respectively to get a final AB color prediction. We made this design decision since we found that it helped mitigate unnatural "blobs" where a single color class was predicted for a large area. Finally, we concatenate the original grayscale channel (L) with the predicted AB channels and convert the resulting LAB image back to the RGB color space.

4. Experiments

4.1. Data

We used the public domain Scene Classification data set, consisting of roughly 25 thousand unique photos of natural and urban scenery. The data set is available to be downloaded at https://www.kaggle.com/datasets/nitishabharathi/scene-classification.

4.2. Hyperparameters

Our hyperparameters are described in Table 1.

Hyperparameter	**Value**
Number of CNN Layers	8
Kernel Size	3×3
Learning Rate	10^{-4}
Optimizer	Adam
Batch Size	50
Epochs	25
Loss Function	Categorical Cross-Entropy
Training Dataset Size	1000 images
Evaluation Metrics	Accuracy

Table 1. Key hyperparameters used in the experiments.

4.3. Challenges and Ablation Study

To better understand the factors influencing performance, we conducted an ablation study. Reducing the number of layers in the model resulted in blotchy outputs, likely due to insufficient capacity for recognizing complex patterns. Increasing the number of color bins caused patches of blue to dominate, likely due to the sparse data distribution leading to overfitting.

Figure 2. Original vs removed layers vs increased number of bins.

Additionally, we observed a bias toward green and blue tones, even in cityscape images. This is likely due to the overrepresentation of forest, sky, and water scenes in the dataset, which skewed the model's predictions.

Figure 3. Nature colorization vs building colorizations with green.

4.4. Evaluation

To evaluate the believability of our colorized images, we conducted a survey to compare them against baseline ground truth images and sample results from the Zhang et al. paper. A total of 31 respondents rated a series of images on a scale from 1 to 5, with 5 indicating the most believable.

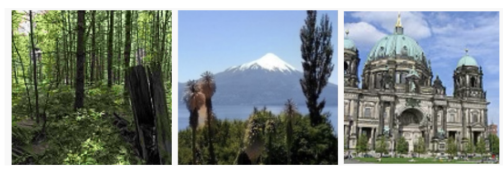

Figure 4. One of our colorizations, a Zhang et al. colorization, and a real image, respectively, that participants were asked to rate.

The survey results below show that while our model's colorized images scored the lowest, they still achieved a reasonable level of realism.

Image Type	**Average Score**
Real Images	4.06
Our Colorized Images	3.89
Colorized Images (Zhang et al.)	4.60

Table 2. Comparison of survey results.

5. Conclusions

Our method is a highly reproducible and fairly accurate method to produce plausibly colored photos. The greatest strengths lie in its ability to generate vibrant images with only a small training set. However, its weakness lies in its limitation to specific classes of images as we saw with the poor transfer from nature to city images. In the future, interesting work could be done by attempting a similar method with a more sophisticated network or on non-scenery datasets like interiors or cityscapes exclusively. Future work could also be done optimizing inference time and memory usage, potentially with methods that specifically work to optimize memory allocation algorithms and the use of memory augmented networks (Yoo et. al, 2019).

References

[1] R. Zhang, P. Isola, and A. A. Efros, Colorful Image Colorization, arXiv preprint arXiv:1603.08511v5, 2016.

[2] J. Yoo, Y. Uh, S. Chun, B. Kang, and J. Ha, Photorealistic Style Transfer via Wavelet Transforms, arXiv preprint arXiv:1906.11888, 2019.

[3] M. He, D. Chen, J. Liao, P. V. Sander, and L. Yuan, Deep Exemplar-based Colorization, IEEE Transactions on Image Processing, vol. 27, no. 1, pp. 424-435, 2018.

[4] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, Image-to-Image Translation with Conditional Adversarial Networks, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[5] F. Pierre and J.F. Aujol, Recent Approaches for Image Colorization, , Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging, 2023.