Combining Convolution Neural Networks and Markov Random Fields for Image Colorization

Luke Melas-Kyriazi and George Han

CS281 Final Project

Overview

In our project, we construct and train a deep convolutional neural network for automatic image colorization, the task of generating a colored image from a grayscale image input.

- We reproduce and build upon recent CNN-based colorization work by Iizuka et al, Zhang et al, and Larsson et al.
- We experiment with various architectures, loss functions, and training datasets with the aim of producing natural, vibrant colorizations.
- We incorporate a Markov Random Field-based model for inferring a final colored image from pixel-level color distributions.

Introduction

Precisely, given the brightness channel of image (a $1 \times H \times W$ input), our objective is to infer the corresponding chrominance and hue channels (a $2 \times H \times W$ output). Image colorization poses a significant challenge to traditional computer vision techniques because a single black-and-white image may have a multitude of plausible colorizations.

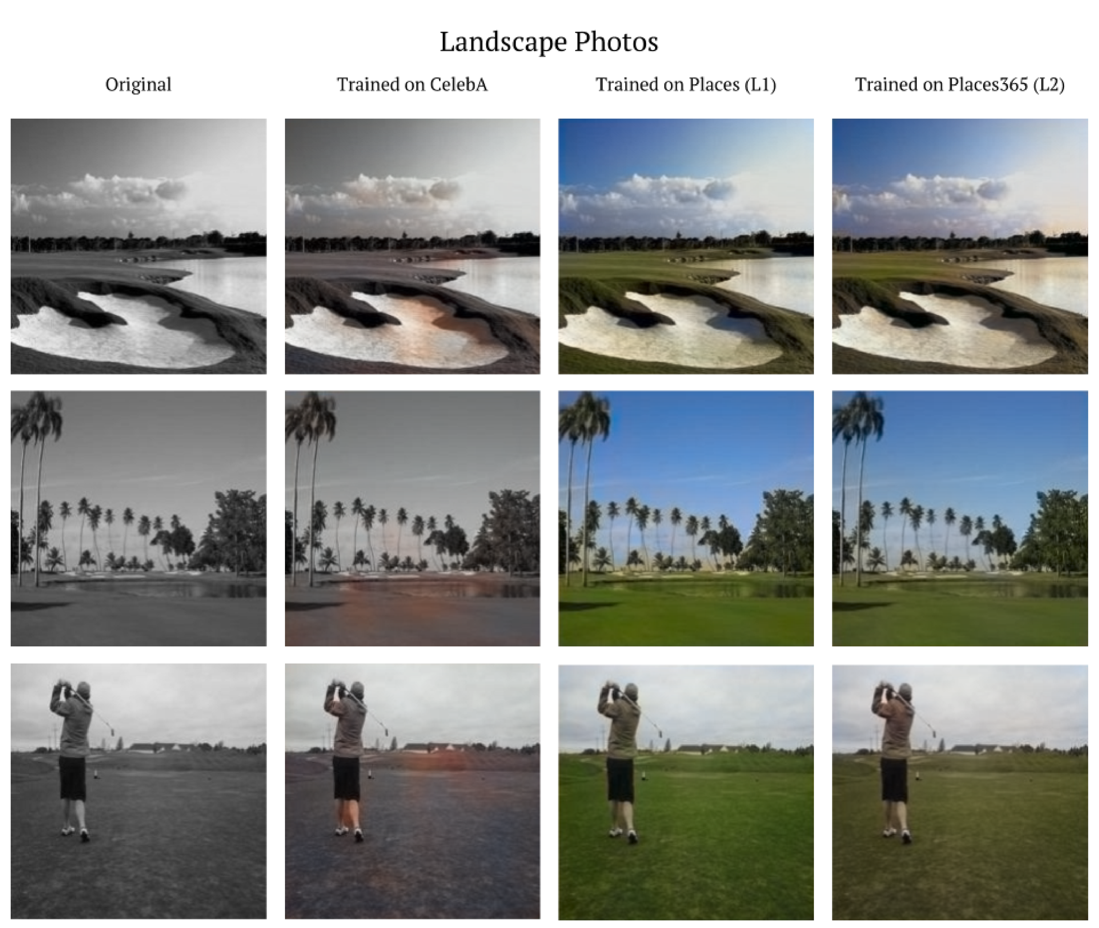


Figure 1: Colorized landscapes from the Places365 Dataset

Neural Network Model

First, we built a convolutional network to directly predict the color channels of an input grayscale input image. The backbone of our network is the **ResNet-18 classification network**, which we retrained on grayscale images ('ResNet-18-Gray'). Our model is based upon the network proposed by Iizuka et al., but takes advantage of the feature-extracting abilities of the ResNet classifier rather than training feature extractors from scratch.

Model Details

Our colorization network is an end-to-end network that can be broken down into 4 parts:

1. Classification network for greyscale images
2. Fusion layer for combining mid-level and global features (a learned linear combination of features)
3. Colorization network for computing pixel-by-pixel color (chroma and hue) distributions
4. Markov random field with mean field variation inference for computing the final colorization

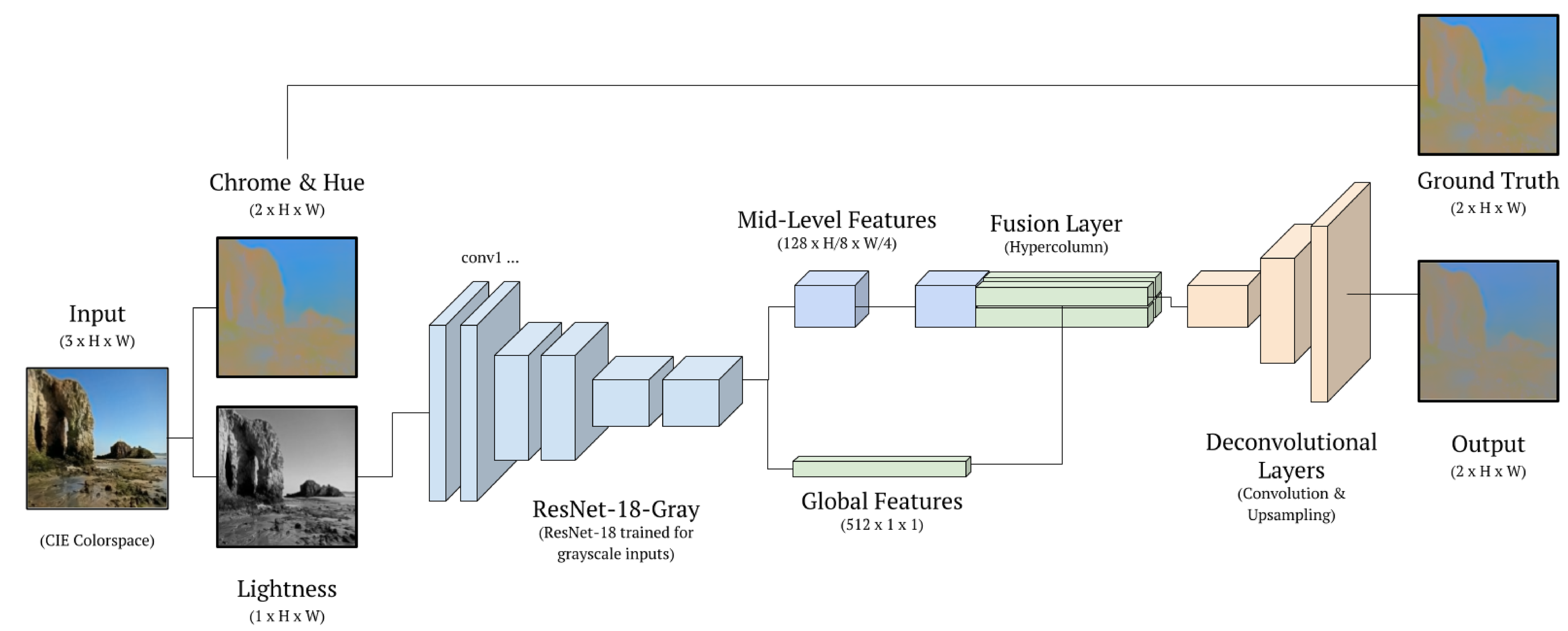
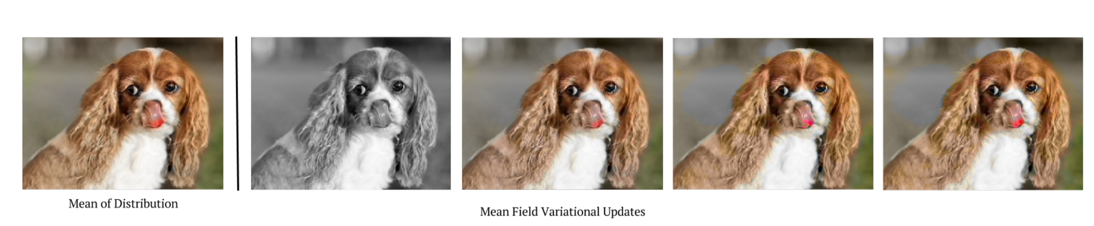


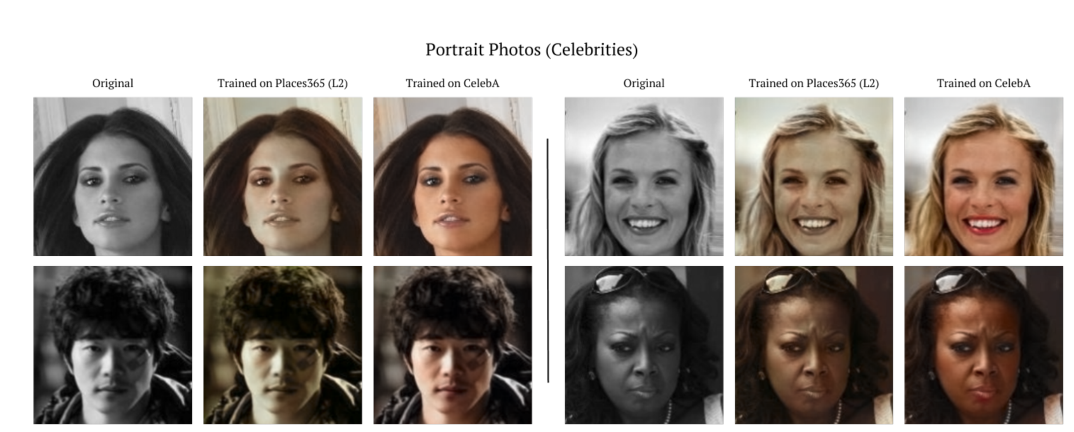
Figure 2: Network architecture (first model)

CNN + MRF Model



Second, we modified our network to output a distribution of color values for each pixel in the image rather than a single color value. We infer the final color values from this distribution using a Markov Random Field. Unary potentials in the model enforce that each pixel's colorization aligns with the networks' color distribution output, while binary potentials enforce that nearby pixels have similar colorizations.

Datasets and Training



We trained our first model separately on three datasets: Places365 (indoor and outdoor scenes), ImageNet (objects), and CelebA (celebrity faces). For most images, training on Places365 produces the most accurate colorizations, but for close-up portraits and skin tones, CelebA produces the best results.

Loss Function

Regression loss functions such as mean squared error and mean absolute error lead models to desaturated colorizations. We address this problem in our second model by predicting a color distribution for each pixel in the input image.

Historical Photos & Future Work

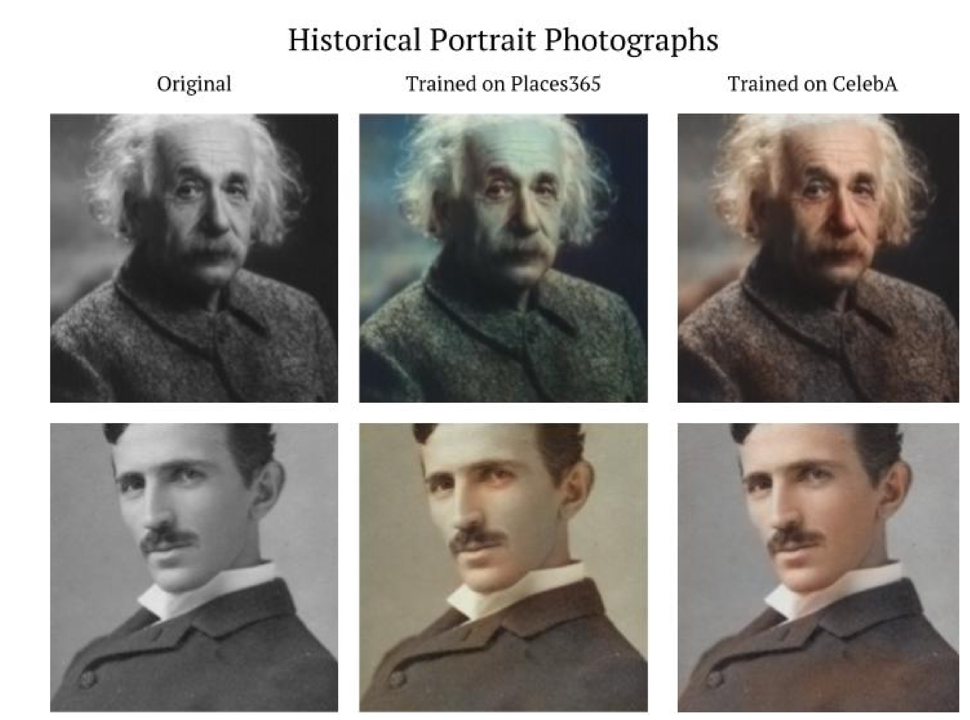


Figure 3: Historical Portrait Photos

We may extend upon this project in a number of different directions:

- Train a single network to perform well on both portraits (CelebA) and landscapes (Places365)
- Incorporate local user input into the model (as proposed by Zhang et al., 2017)
- Use our colorization network for self-supervised learning (as proposed by Larsson et al., 2017).

Acknowledgements

We would like to thank Professor Sasha Rush and teaching fellows Mark Goldstein, Ratchit Singh, and Yuntian Deng for their assistance with this project.

Contact Information

- Email: lmelaskyriazi@college.harvard.edu
- Email: hanz@college.harvard.edu