Automated Image Colorization Using Deep Learning

Hanzhao Lin
hanzhao@stanford.edu

Rafael Ferreira
rcf2132@stanford.edu

Introduction

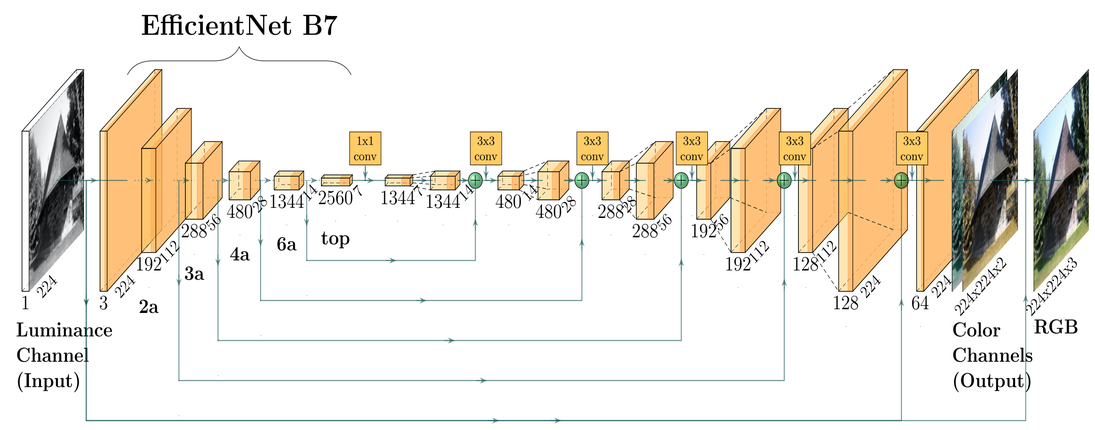
Before humans had devices to capture colored media, the only available option was to record grayscale images and videos. Even with the advance of technology, it's still challenging to automatically colorize an image due to its uncertainty. In this work, we apply neural network, specifically a pre-trained EfficientNet [1], to implement a system automatically producing a plausible colorization for a given grayscale image. We evaluate the outcome of the model in both subjective and quantitative ways.

Dataset

We utilize the small version of Places dataset [2] for our work. This dataset contains about 2 million images of 256×256 resolution comprising about 400 unique scene categories. This Places dataset makes a good balance in terms of the image diversity and data size, which allows us to make considerable progress under the constraint of limited time and computing resources. The dataset is also close to the real-world setting so that model can achieve an acceptable generalization performance and be capable of colorizing blank-and-white photos which we collect from real life.

Model architecture

Inspired by Dahl [4], our approach utilizes outputs from 5 intermediate activation layers in a pre-trained EfficientNet B7. The output of 7×7 top activation layer of EfficientNet is first passed through a 1×1 Conv2D block resulting 7×7×1344 output matrix. To fuse it with other activation layers, the output is up-scaled and added up to the output of the previous activation layer of EfficientNet, followed by another Conv2D block to resize its dimension. This process is repeated 5 times so that the original 224×224 image is recovered. Each Conv2D block consists of a 3×3 Conv2D layer, a Batch Normalization layer and a ReLU activation layer. Finally, to output color channels in CIE Lab color space, the output block is defined as a 3×3 Conv2D layer with 2 output channels and a tanh activation layer mapping values to [-1, 1]. The input luminance channel is combined with model outputs to recover the output colors back into RGB color space.



Objective functions

To achieve minimal color shifting between ground truth and prediction, the objective is defined to minimize mean squared error (MSE) for 2 output color channels in CIE Lab color space. The loss function can be formularized as,

$$\mathcal{L}(Y^{(i)}, \hat{Y}^{(i)}) = \frac{1}{2hw} \sum_{h,w} ||Y^{(i)}_{h,w} - \hat{Y}^{(i)}_{h,w}||_2^2$$

We later observed that using the objective function above led to a tendency of desaturated colors. Inspired by Zhang et al. [3], we analyzed the training set and confirmed that desaturated colors appear much more frequently than vivid colors in the real world as Figure (a). To encourage producing saturated colors, we introduce function v as shown in Figure (b) to weight each pixel based on their color rarity, and adjusted loss function accordingly to be,

$$\mathcal{L}(Y^{(i)}, \hat{Y}^{(i)}) = \frac{1}{2hw} \sum_{h,w} v(Y^{(i)}_{h,w}) ||Y^{(i)}_{h,w} - \hat{Y}^{(i)}_{h,w}||_2^2$$

$$v(a,b) \propto \frac{1}{(1-\lambda) \times \tilde{p}(a,b) + \lambda \div 4}, \quad s.t. \quad \mathbb{E}[v(a,b)] = \sum_{a,b} \tilde{p}(a,b) v(a,b) = 1$$

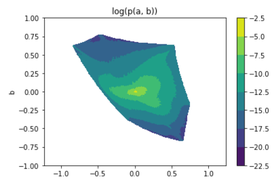


Figure (a)

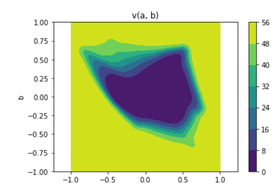
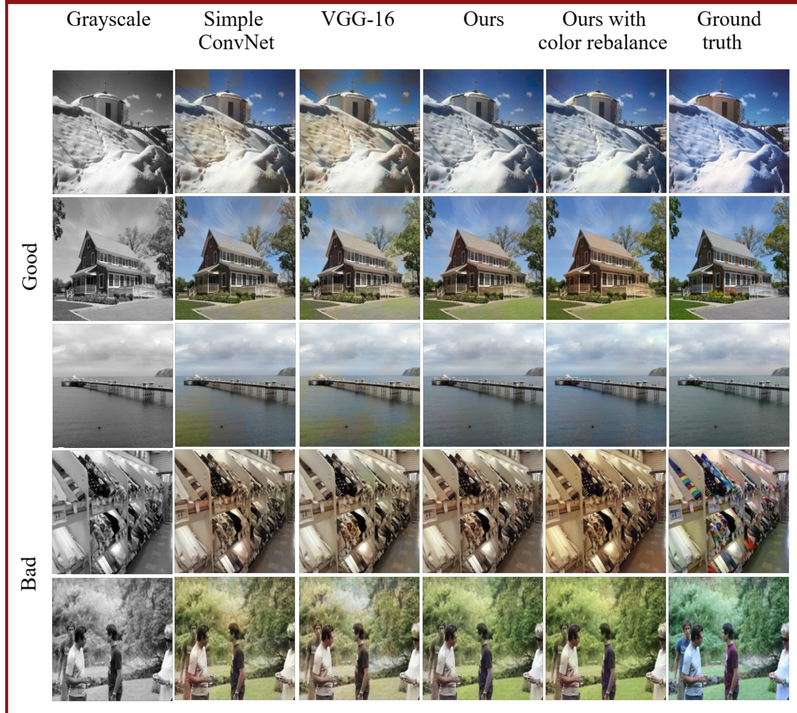


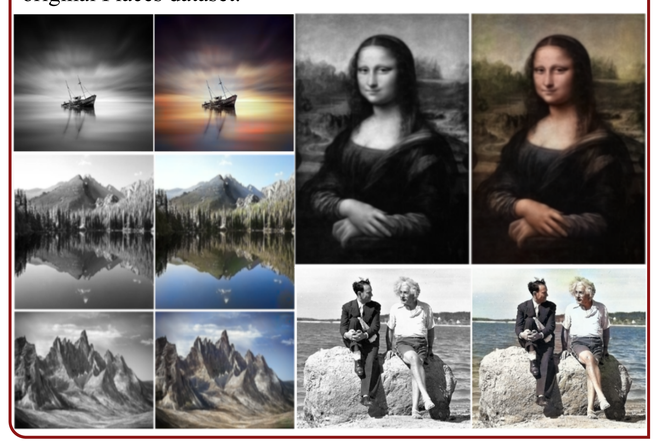
Figure (b)

Experiment result



Experiment result (continued)

We collected some grayscale images from the real life and the model has shown its capability of generalization outside of the original Places dataset.



Conclusion

Our approach automatically colorizes images with perceptual improvements from baseline approaches. But it doesn't make considerable progress on our quantitative metrics, even the model produces some images that are very close to the real-world photos. We believe better metrics as well as better training objective are necessary to improve the performance on this task. Also, training on more diversified datasets like ImageNet with may also help.

Reference

[1] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In International Conference on Machine Learning, pages 6105–6114. PMLR, 2019.

[2] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. Places: A 10 million image database for scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017.

[3] Richard Zhang, Phillip Isola, and Alexei A. Efros. Colorful image colorization. CoRR, abs/1603.08511, 2016. URL http://arxiv.org/abs/1603.08511.

[4] Ryan Dahl. Automatic colorization. URL https://tinyclouds.org/colorize/.