Towards More Vibrant Colorful Image Colorization

Ahan Mukhopadhyay*
A13491660
amukhopa@ucsd.edu

Kolin Guo*
A14090849
ruguo@ucsd.edu

Kyle Lisenbee*
A53298980
klisenbe@eng.ucsd.edu

Ulyana Tkachenko*
A13978517
utkachen@ucsd.edu

1 Introduction

The task of colorizing a grayscale image is both an unconstrained optimization problem over the human-centric believability of such a colorization and an optimization problem over the images potentially unknown ground truth. On the surface, hallucinating a plausible colored version of a grayscale image seems daunting, since two of the three dimensions of the image data has been lost. However, if done accurately and efficiently, such a task can have deeper application across a multitude of domains from areas like computer vision and graphics to object detection and classification. Moreover, it can be used as a powerful pretext task for self-supervised feature learning (an autoencoder that outputs two color channels given one lightness channel input). For our project we intend to build on the success of the CNN based model proposed in [1] by injecting reasonable constraints to encourage more vibrant colorful images and optimizing over the loss functions dual.

1.1 Task Assignment

This is a group project so each member is expected to participate in weekly discussion, mathematical deduction, and implementation equally. However, currently the tentative task assignments are: **Ahan Mukhopadhyay** and **Ulyana Tkachenko** works on the mathmetical formulation and deduction of the primal and dual problem and the KKT conditions; **Kolin Guo** and **Kyle Lisenbee** works on code implementation and result analysis.

1.2 Previous Works

1.3 Intended Contributions

1.4 Paper Organization

2 Statement of the Problem

2.1 Primal Problem

Given an input lightness channel $\mathbf{X} \in \mathbb{R}^{H \times M \times 1}$, our objective is to learn a mapping $\hat{\mathbf{Y}} = \mathcal{F}(\mathbf{X}; \mathbf{\Theta})$ to the two associated color channels *ab* of the CIE *Lab* color space $\mathbf{Y} \in \mathbb{R}^{H \times W \times 2}$, where H, W are image dimensions and Θ is the parameters of the mapping function $\mathcal{F}(\cdot)$. We choose the CIE *Lab* color space because the distances in this space model perceptual color distances of human vision.

*The authors are alphabetically ordered by first names. All authors belong to the Department of Computer Science and Engineering.

A straightforward objective function, as used in [1,2], is the Euclidean loss $L_2(\cdot,\cdot)$ between predicted and groundtruth colors:

$$L_2(\hat{\mathbf{Y}},\mathbf{Y}) = \frac{1}{2}\sum_{h,w} \|\mathbf{Y}_{h,w} - \hat{\mathbf{Y}}_{h,w}\|_2^2.$$

Our primal problem is focused on minimizing the objective function. However, this loss tends to produce a desaturated color because the optimal value to the loss function will be the mean of the ab values. To encourage more vibrant color, we will incorporate a term on the ab values to induce more vibrant color by constraining the hue of ab from the lightness-hue-chroma (CIE Lhc) colorspace [3] or the excitation purity ratio of the color saturation value [4]. Also, we could add a regularization term on the mapping function weight Θ.

In the paper [1], the authors propose to learn the mapping $\hat{\mathbf{Z}} = \mathcal{G}(\mathbf{X};\mathbf{\Theta})$ as a multinomial classification problem using a multinomial cross entropy loss $L_{cl}(\cdot,\cdot)$ on the quantized ab output space:

$$L_{cl}(\hat{\mathbf{Z}},\mathbf{Z}) = -\sum_{h,w} v(\mathbf{Z}_{h,w}) \sum_{q} \mathbf{Z}_{h,w,q} \log(\hat{\mathbf{Z}}_{h,w,q}),$$

where $v(\cdot)$ is a weighting term that is used to rebalance the loss based on the color-class rarity in the ab space. The predicted ab values $\hat{\mathbf{Y}}$ is then generated from $\hat{\mathbf{Z}}$ with function $\hat{\mathbf{Y}} = \mathcal{H}(\hat{\mathbf{Z}})$ which is the *annealed-mean* of $\hat{\mathbf{Z}}$, and the groundtruth $\mathbf{Z}$ is obtained with $\mathbf{Z} = \mathcal{H}_{gt}^{-1}(\mathbf{Y})$ which is a soft-encoding scheme of the quantized ab space using a Gaussian kernel. Similarly, we could add a regularization term on the mapping function weight Θ.

2.2 Dual Problem

2.3 KKT Conditions

3 Intended Approaches

4 Conjectured Results

To evaluate our color diversity constraint and the added regularization terms, the paper [1] uses mainly three different evaluation methods: (1) Perceptual realism by using Amazon Mechanical Turk (AMT) to let human participants classify between *fake* colorized images and real groundtruth color images (*i.e.*, a "colorization Turing test"); (2) Semantic interpretability (VGG classification) by inputing the generated colorized images into a VGG-16 network pretrained on color images and perform image classification on ImageNet; (3) Raw accuracy (AuC) of the cumulative error distribution over the predicted values in ab space.

The first evaluation method is impractical given that this is a class project. The second method represents the plausibility of the generated colorized images while the third method simply measures the raw pixel prediction accuracy. Given the fact that the paper [1] already experiments with a network trained with L2 regression loss, fine-tuned from the full classification with rebalancing color weights in the ab space and results in slightly decreased rebalanced raw AuC accuracy (66.2 vs. 67.3) and slightly increased VGG classification accuracy (56.5 vs. 56.0), we conjecture that the color-diversity constrained L2 regression will have slightly higher rebalanced raw accuracy and slightly lower VGG classification accuracy.

References

[1] R. Zhang, P. Isola, and A. A. Efros, "Colorful image colorization," in *ECCV*, 2016.

[2] Z. Cheng, Q. Yang, and B. Sheng, "Deep colorization," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 415–423.

[3] A. Ford, A. Roberts, and C. Poynton, "Colour space conversions," pp. 8–10, 1998.

[4] G. Finlayson and M. Drew, "Constrained least-squares regression in color spaces," *J. Electronic Imaging*, vol. 6, pp. 484–493, 1997.