Colorization of Grayscale Images

Bardia Beigi, Vamsi Chitters, Fabian Frank

{bardia, vamsikc, fabfrank}@stanford.edu

CS230
Deep Learning
Winter 2018

INTRODUCTION

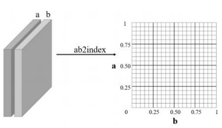
- **Motivation**: Enable colorization of grayscale photographs
- **Image colorization**: Hallucinate colors such that the output picture seems natural to the human eye
- **Models**: (1) Regression (2) Classification (3) Classification with Color Rebalancing

Model Input: Grayscale Image

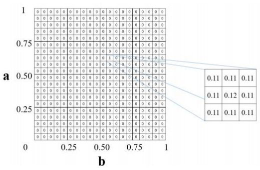
Model Output: Colorized Image

DATA

- **Datasets**: CIFAR-10, ImageNet
- **Specifications**: 60K 32x32 RGB, 10 classes (CIFAR), 1.2M 32x32 RGB, 1000 classes (ImageNet)
- **Generating Model Inputs**:
  - RGB → LAB Colorspace → L channel
- **Evaluation Metrics**:
  - Quantitative: Accuracy
  - Qualitative: User Survey
- **Pre-processing**:
  - (1) Convert ab-plane [0,1]x[0,1] to a 20x20 grid (cell size: 0.05) and output pixel color labels as indices in the grid [0,399].
  - (2) Smooth out label per pixel in the ab-plane to incorporate immediate neighbors (3x3 window) without affecting loss



(1) ab channel plane conversion



(2) Label smoothing

MODELS

➢ **Baseline: Regression**

Regression model with L2 loss between the label and predicted ab-plane color **values**

$$\mathrm{L}_2(\widehat{\mathbf{Y}}, \mathbf{Y}) = \frac{1}{2}\sum_{h,w}\|\mathbf{Y}_{h,w} - \widehat{\mathbf{Y}}_{h,w}\|_2^2$$

➢ **Classification**

Classification model with cross-entropy loss between smoothed label and predicted ab-plane color **bins**

$$\mathrm{L}_{cl}(\widehat{\mathbf{Z}}, \mathbf{Z}) = -\sum_{q}\mathbf{Z}_{h,w,q}\log(\widehat{\mathbf{Z}}_{h,w,q})$$

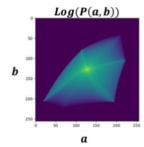
➢ **Classification w/ Color Rebalancing**

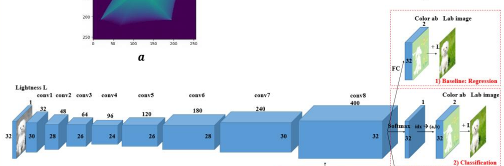
Classification model with color rarity incorporated as weights, favoring more vibrant colors

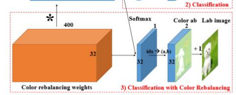
$$\mathrm{L}_{cl}(\widehat{\mathbf{Z}}, \mathbf{Z}) = -\sum_{h,w} v(\mathbf{Z}_{h,w})\sum_{q}\mathbf{Z}_{h,w,q}\log(\widehat{\mathbf{Z}}_{h,w,q})$$

$$v(\mathbf{Z}_{h,w}) = \mathbf{w}_{q^*}, \text{ where } q^* = \arg\max_q \mathbf{Z}_{h,w,q}$$

$$\mathbf{w} \propto \left((1-\lambda)\widetilde{\mathbf{p}} + \frac{\lambda}{Q}\right)^{-1}, \quad \mathbb{E}[\mathbf{w}] = \sum_q \widetilde{\mathbf{p}}_q \mathbf{w}_q = 1$$



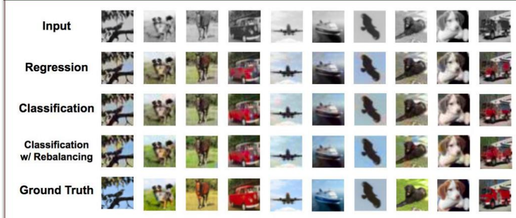




Hyperparameters:
LR: 10^{-3}

Other model specs:
Adam optimizer, ReLU, batch-norm, no max pooling

RESULTS AND ANALYSIS



Accuracy:

- Regression: % of correctly predicted color values in the ab-plane (1/65536 random chance)
- Classification: % of correctly predicted color bins in the ab-plane (1/400 random chance)

Survey Results:

- % of synthesized images picked as real

Model[1] Train size: 45,000, Test Size: 10,000	Quantitative		Qualitative
	Train	Validation	Survey Results
Regression	0.01%	0.01%	34.6%
Classification	**23.0%**	**22.6%**	57.7%
Classification with Rebalancing	19.2%	17.7%	**69.2%**

[1] Experimented with various hyperparameters, including learning rate, choice of optimization algorithm and convolutional filter size etc.

Error Analysis:

- Transitions between colors are not fully seamless
- Performs poorly on the frog class – due to lack of contrast

DISCUSSION AND FUTURE WORK

- Regression (tends to predict in the unsaturated region) performs worse than classification
- Color rebalancing proves very effective in generating realistic images
  - Handles inherent skew towards low ab values in natural images, due to background (e.g. cloud, dirt)
- Less colorful images were considered to be ‘synthesized’ in the user survey
- Continue conditional-GAN experiments to generate more realistic images
- Further scope for research:
  - Image segmentation combined with prior knowledge of color distribution (i.e. sky is blue)
  - Incorporate cluster-based techniques (i.e. K-means)
  - Hierarchical models to learn at different granularities

REFER-ENCES

- Z. Cheng, Q. Yang, and B. Sheng. Deep colorization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 415–423, 2015.
- P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. *arXiv preprint*, 2017.
- A. Krizhevsky, V. Nair, and G. Hinton. The cifar-10 dataset. online: *http://www. cs. toronto. edu/kriz/cifar. html*, 2014.
- A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- R. Zhang, P. Isola, and A. A. Efros. Colorful image colorization. In *European Conference on Computer Vision*, pages 649–666. Springer, 2016.