Image Colorization and Classification

Peng Li[1], Zhefan Wang[2]

[1] Department of Civil and Environmental Engineering, Stanford University
[2] Department of Electrical Engineering, Stanford University

Introduction

Tasks

- Build an automatic colorization system that takes in grayscale images and outputs colorized images
- Duplicate the VGG network to test if the synthesized images can potentially improve classification accuracy

Motivations

- Image colorization has very useful applications such as historical image/video reconstruction
- Realistic colorization can potentially improve the classification accuracy for grayscale images due to extra information provided by colors

Challenges

- Colorization task has multi-modal nature (i.e.. objects can take multiple plausible colorization). Eg. It is reasonable to predict an apple in a grayscale image to be either green, yellow or red
- Colorization task is not to produce the original RGB images, but to produce plausible colored images

Related Work

I. R. Zhang, P. Isola, and A. A. Efros. Colorful image colorization. ECCV, 2016.
II. K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015

Data

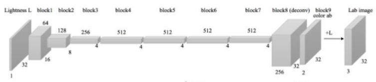
- CIFAR 10 Dataset
- Colorization:
  - 50,000 training images, 5,000 dev images, 5,000 test images, image size 32×323
  - The CIFAR RGB images are converted into Lab color space
- VGG:
  - 10,000 training images, 1000 dev images, 1000 test images, image resized from 32×32×3 to 224×224×3

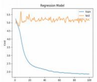
Models

Framework

Grayscale Image → Colorization → Colored Image → VGG → Predict Class

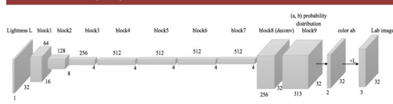
A (I). Colorization Regression Model

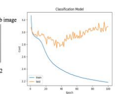




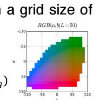
Loss Function $L_2 = (Y, \hat{Y}) = \frac{1}{2}\sum_{h,w} \|Y_{h,w} - \hat{Y}_{h,w}\|_2^2 + \lambda \sum_{W} \|W\|_2^2$

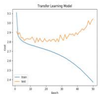
A (II). Colorization Classification Model





- Model colorization as a classification problem where the CNN predicts each pixel's ab value from 313 classes. (Quantized *ab* color space with a grid size of 10. A total of 313 ab pairs are in gamut.)
- Also used VGG to do transfer learning

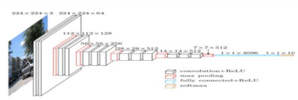


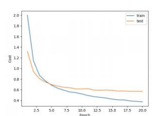


Loss Function $L = (Y, \hat{Y}) = -\sum_{q} Y_{h,w,q} \log(\hat{Y}_{h,w,q})$

Annealed Mean Technique $\mathcal{H}(\mathbf{Z}_{h,w}) = \mathbb{E}[f_T(\mathbf{Z}_{h,w})], \quad f_T(\mathbf{z}) = \frac{\exp(\log(\mathbf{z})/T)}{\sum_q \exp(\log(\mathbf{z}_q)/T)}$

B. VGG Classification Model

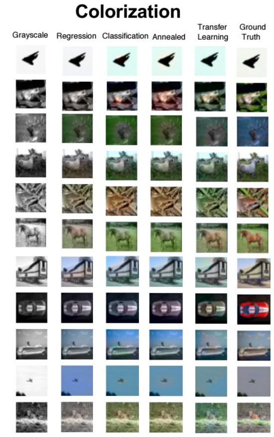




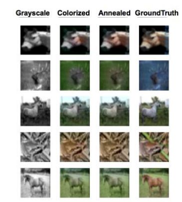
Loss Function $L = (Y, \hat{Y}) = -\sum_{Y} Y \cdot \log \hat{Y}$

Results

Colorization



Classification



Truth	Grayscale	Colorized	Annealed	Ground Truth
horse	airplane (25%)	dog (23%)	dog (70%)	horse (37%)
deer	cat (24%)	cat (35%)	frog (57%)	deer (67%)
horse	dog (23%)	horse (95%)	horse (90%)	horse (99%)
frog	airplane (25%)	frog (78%)	frog (87%)	frog (91%)
horse	cat (20%)	horse (84%)	horse (90%)	horse (99%)

Conclusions

- The regression model has the worst performance among all the models.
  - L2 loss is not robust in handling the multimodal nature of colorization.
  - It favors grayish colorization. If an object can take on a set of distinct ab values, the optimal solution to the L2 loss will be the mean of the set.
- Both the classification model and the transfer learning model perform decently well
  - Both models treat the problem as multinomial classification.
- Annealed-mean technique interpolates the predicted distribution to produces both vibrant and spatially consistent colorization results.
  - T = 0.89 works the best in this case.
- Colorization in general boosts object classification accuracy and confidence.

Future Work

- Implement the rebalancing method introduced in paper I to further improve the vibrancy of the colored images
- Train on larger data set such as ImageNet to achieve better generalization