Image colorization using Unet

Alexandru Stanciu, Mentor: Elena Burceanu
alexandru-gabriel.stanciu@my.fmi.unibuc.ro

Department of Mathematics and Computer Science - University of Bucharest

Introduction

- We are interested in a computationally efficient[1] method of colorizing grayscale images. The colorization problem is viewed as a classification problem over the YUV color space where for every color pixel of the U and V channels there exists a color class. Techniques such as *rebalancing* and *taking an annealed mean*[3] are used to steer the model into producing more visually pleasing results (read greater color variation and higher intensity colors).

Classifying color pixels

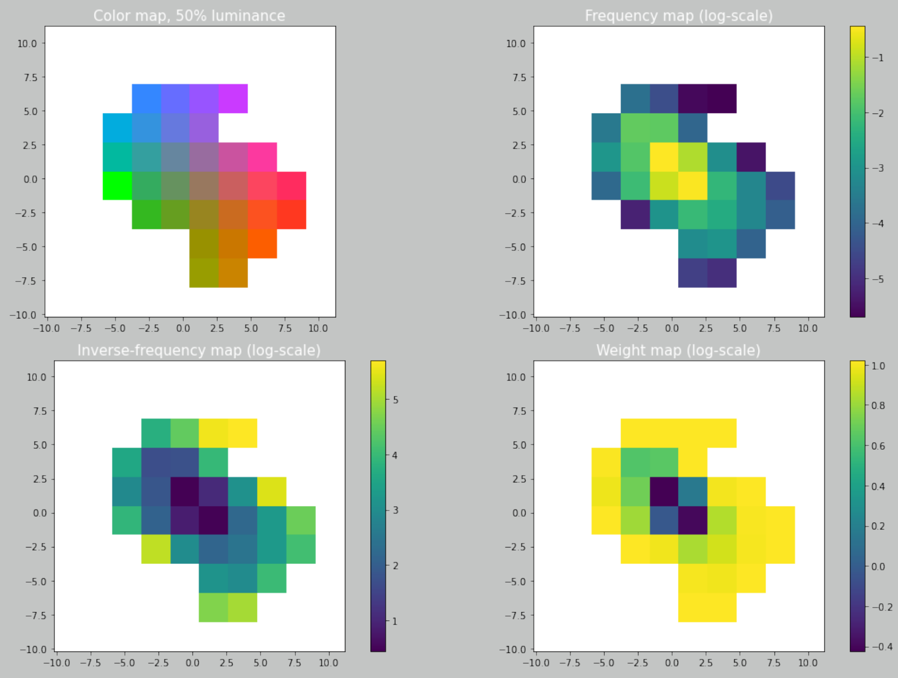
- The YUV color space allows us to separate light information (read brightness or *luminance*) in the Y channel from color information (*chrominance*) in the U and V channels respectively.
- Following the same approach as [3] and [1] we:
  - build a UV pixel distribution from a set of images, in this case a small subset of [2]
  - discretize the color space into the 32 most frequent bins
  - use multinomial Cross Entropy loss with *weight rebalancing* to encourage the prediction of rare color classes:

$$w_p = \left((1-\lambda)\hat{P}(b) + \frac{\lambda}{n}\right)^{-1}$$

  where p is pixel b is its closest bin, n is the number of bins and $\lambda \in (0,\ 1)$
  - use an *annealed mean* to extract the color value y of a UV pixel from our distribution:

$$y = f_T(z) = \frac{exp(log(z)/T)}{\sum_i exp(log(z_i)/T)}$$

  where z is the $n - dim$ probability vector for a pixel and $T \in (0, 1]$



Discretization for 32 bins, $\lambda = 0.3$

Dataset

- Two datasets were used as follows:
  - 10k unique images from Flickr, sized between (500px, 800px) queried by tags including: 'landscape', 'beach', 'mountains', 'nature', 'sunset', 'sunrise'
  - 10k unique images from 220 classes of the SUN[2] dataset with 10 images per class
  - a small number of the SUN images were used to build the initial color distribution and learn the weights
  - training and testing were done with images from only one dataset at a time, both using the SUN weights
- Datasets were split in this manner:
  - 90%*train* − 10% *test*
  - 90% *remaining train* − 10% *validation*
- The data was augmented by resizing the images to 256px on the short edge, taking 2 random crops, one flipped random crop and one random crop with Gaussian noise, all across the long edge resulting in a 5x dilation.

Model

- A bottleneck architecture based on a Unet [1] was used for its relatively low computational toll and good reported results. NOTE: I used Layer Normalization as opposed to Batch Normalization.

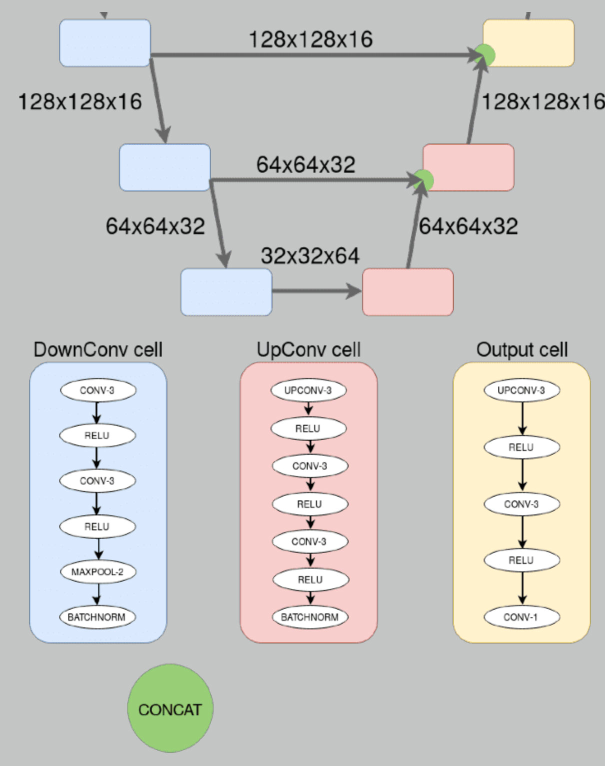


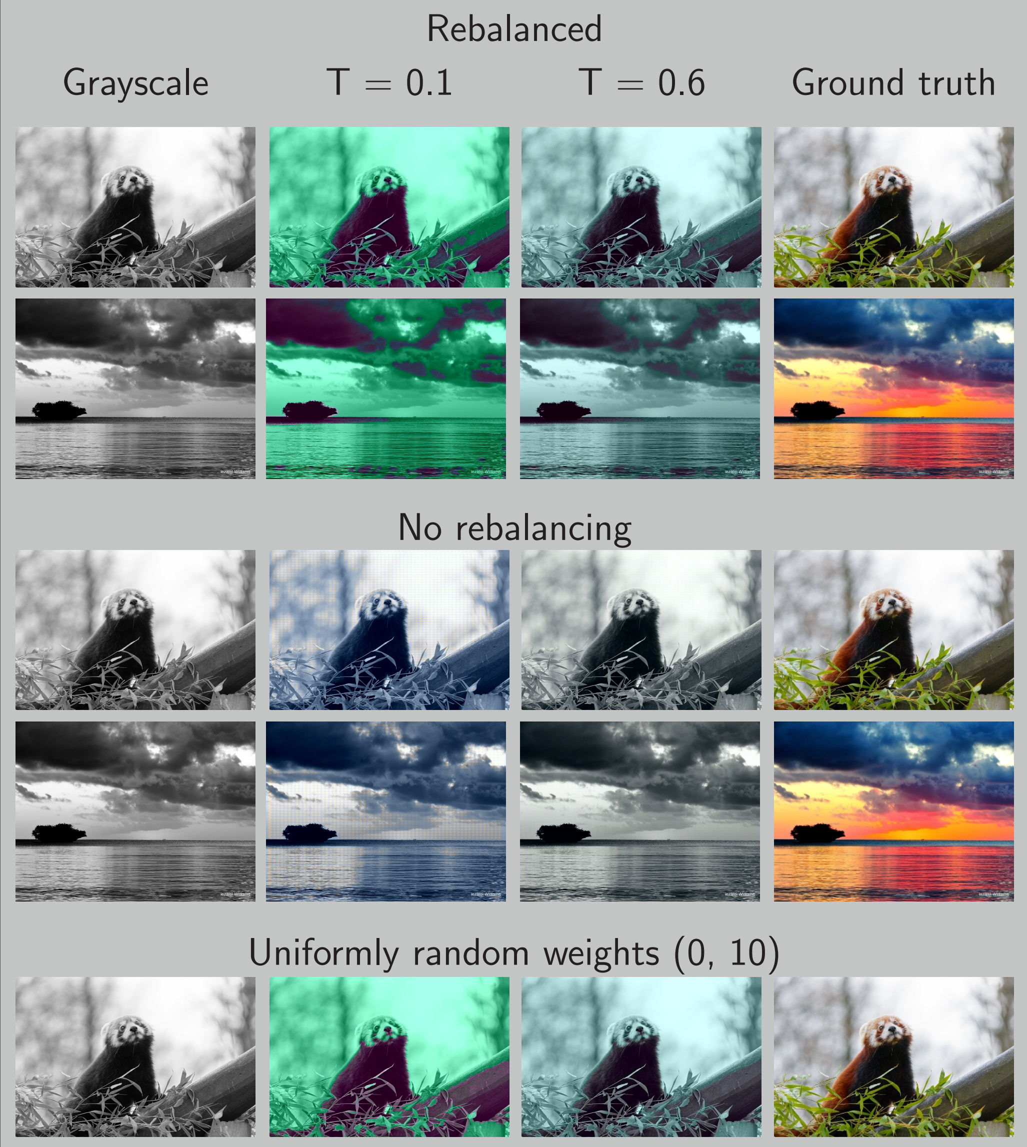
Figure: ColorUnet architecture from [1]

- A 256x256x1 input is repeatedly downsampled by the *DownConv* cells, then upsampled by the *UpConv* cells with channel wise concatenation in between. Finally the input is upsampled by the *Output* cell and the final output has shape 256x256x*n* where *n* is the number of classes.

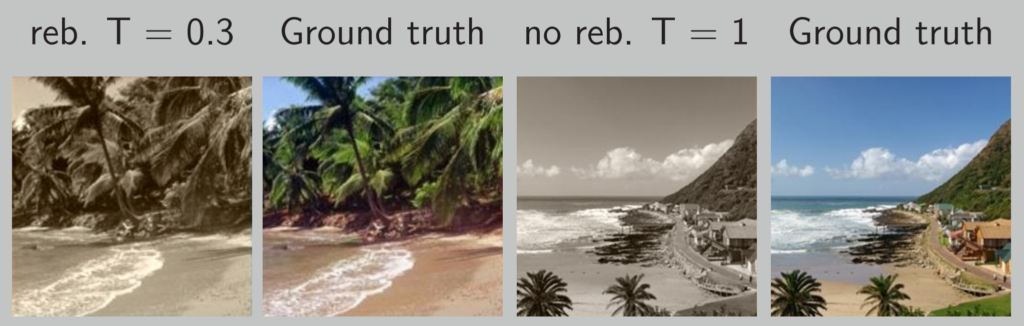
Experiments

- The model was trained using the Adam optimizer with a learning rate of 10^{-4}, batch size of 48 and the target number of epochs was set between 10 and 25 with early stopping if no improvement to the *validation loss* was present after 3 epochs.
- Experiments were done with various dataset sizes and rebalancing weight configurations but all with 32 classes.
- Attempt to diagnose the issues by running experiments on a reduced subset of SUN with 2 categories: 'beach', 'ocean' using either 32 or 6 color classes. The predictions 'looked' similar for the TRAIN, VALIDATION and TEST sets so we just display the results for the TRAIN set.

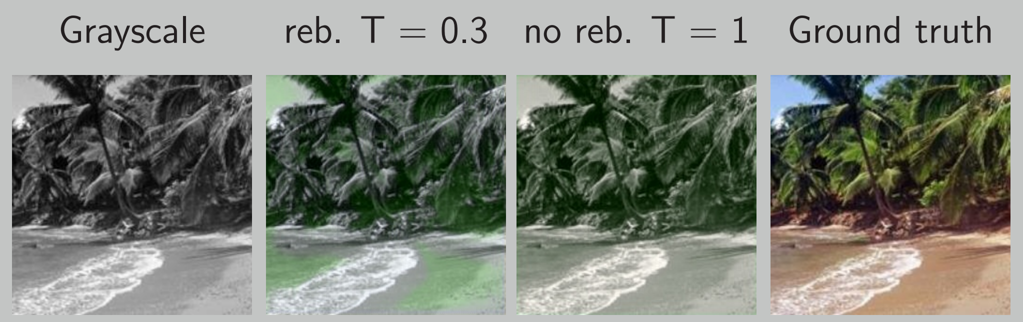
Flickr 13k images, 10 epochs



SUN 2 categories, 32 bins



SUN 2 categories, 6 bins



Observations

- Although rebalancing does seem to have an impact, the model has a hard time finding the proper colorings and resorts to applying at most 2 colors to the entire image. Greater rebalancing weights do produce more vibrant images as expected, although with no more than 2 colors. This may be caused by a lack of diversity in the rebalancing weight training images.
- Increasing the size of the dataset did not yield any visible benefits, sometimes resulting in even more muted images. This may have been caused by too large a diversity in the training dataset, especially for the Flickr dataset possibly due to image tags not being reliable. However results were similar for SUN which has validated hand-tagged images.
- Using a reduced number of color classes on a small dataset did not yield any qualitative improvements as well.

Conclusions

- The model showed some promising results initially but was not able to improve much further with either more training epochs or larger dataset sizes and appears to be predicting the mean color of the binned values, despite tweaking the T parameter.

References

[1] V. Billaut, M. de Rochemonteix, and M. Thibault.
Colorunet: A convolutional classification approach to colorization, 2018.

[2] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba.
SUN database: Large-scale scene recognition from abbey to zoo.
2010.

[3] R. Zhang, P. Isola, and A. A. Efros.
Colorful image colorization, 2016.

Apr. 2020