Image Colorization

Group 11

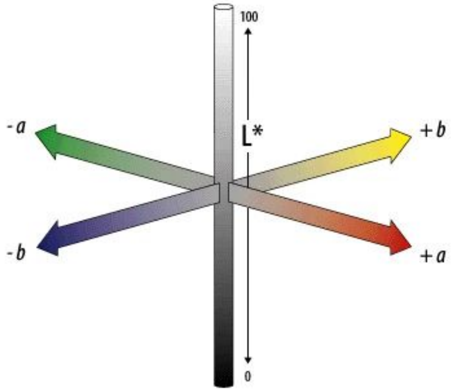
Matheus CARIÚS, Pedro MACEDO,
João OLIVEIRA, Mateus SILVA

Summary

- Problem Overview
  - Model
  - Loss functions
- Results

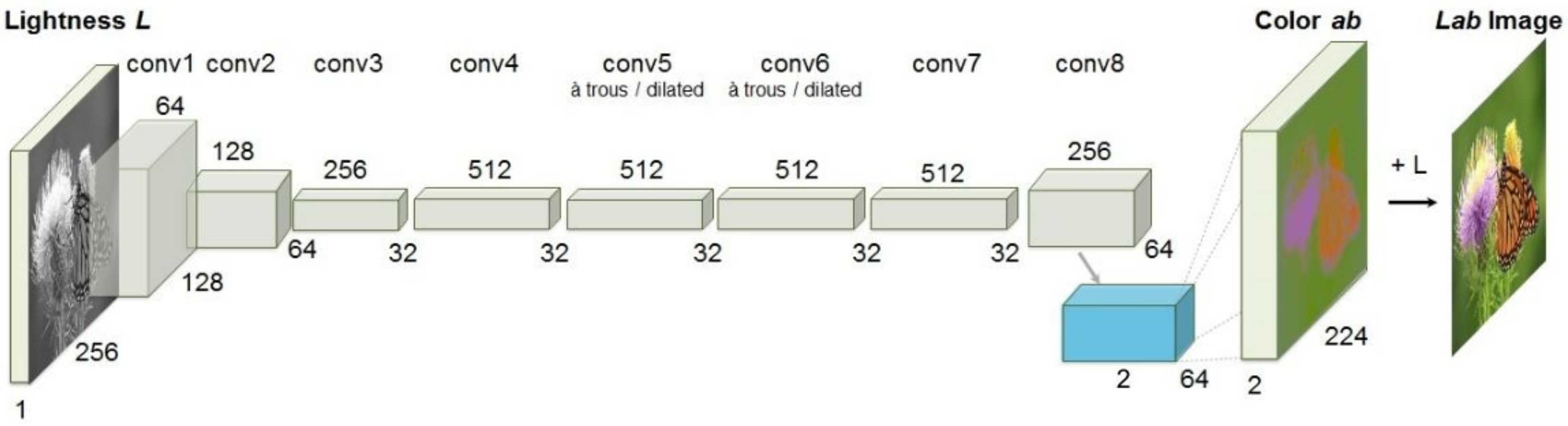
Problem Overview - Objective

- Given a grayscale photograph as input the goal is to create a plausible version of a colorized image
- Given a lightness channel **L**, the model predicts the corresponding **a** and **b** color channels of the image (CIE Lab colorspace)



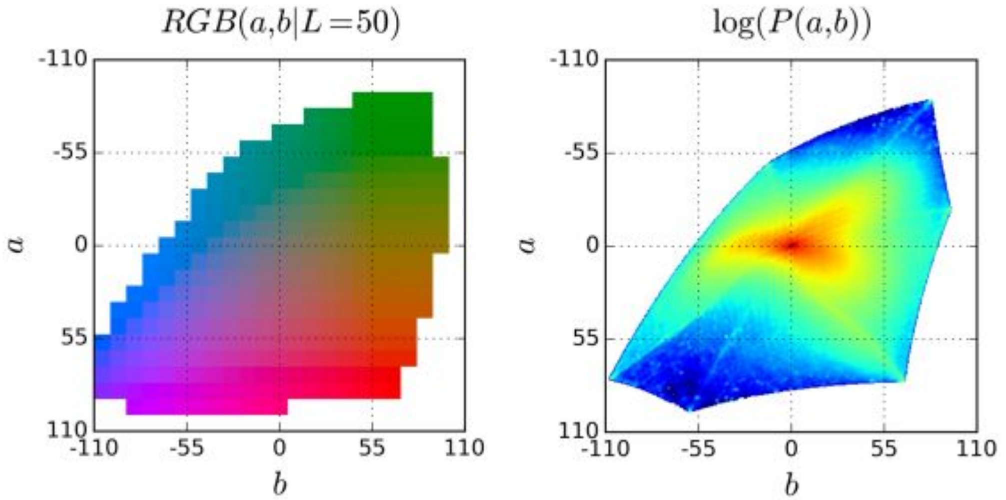
Problem Overview - Architecture

- The system is implemented as a feed-forward CNN:



Problem Overview - Loss functions

- The loss functions we used were the Multinomial Cross Entropy Loss, L2 Loss and the Focal Loss for this task.



Empirical probability distribution of ab values, shown in log scale

Problem Overview - Multinomial Cross Entropy Loss

- This approach transforms the ground truth color to a vector (**Z**) and compute the Cross Entropy with weighting factors (to rebalance the color-class rarity).

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

Problem Overview - L2 Loss function

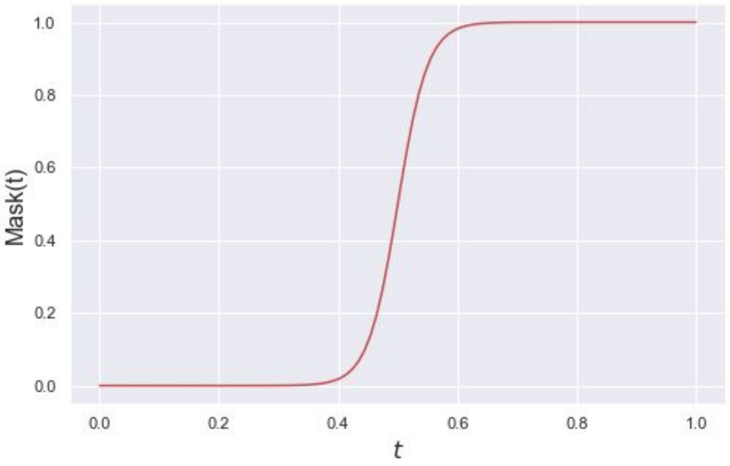
- This loss is not robust to the inherent ambiguity and multimodal nature of the colorization problem

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

In color prediction, this averaging effect favors grayish, giving sometimes desaturated results.

Problem Overview - Focal Loss

- In this case we normalized those pixels with the greatest loss, for averaging the model loss.



$$t = \frac{\text{pixel loss}}{\text{max loss}}$$

$$\text{Mask}(t) = \frac{1}{1+\alpha exp(-\beta t)}$$

Training and Testing

- Trained on the SUN objects dataset (16k images).
- Tested on both the SUN objects (16k images) and the SUN scenes (100k images) datasets.
- Trained for ~12h on 4 V100 GPUs (PyTorch's DataParallel for parallelization)
- Approximately 1 epoch each 2 minutes for L2 loss.
- Approximately 1 epoch each 120 minutes for Multinomial Cross Entropy loss.

Bad results - L2 loss

Ground truth

Predicted

Bad results - L2 loss

- If an object can take on a set of distinct ab values, the optimal solution to the Euclidean loss will be the mean of the set. In color prediction, this averaging effect favors grayish, desaturated results.

Results - L2 loss on training set

Ground truth

Predicted

Bad results - L2 loss on the scenes dataset

Ground truth

Predicted

Bad results - L2 loss on the scenes dataset

Ground truth

Predicted

Good results - L2 loss on the scenes dataset

Ground truth

Predicted

Good results - L2 loss on the scenes dataset

Ground truth

Predicted

Results - Focal Loss on the training set

Ground truth

Predicted

Results - Focal Loss on the scenes dataset

Ground truth

Predicted

Results - Focal Loss on the scenes dataset

Ground truth

Predicted

Improvements

- Pre-process all the images to use the multinomial cross-entropy loss.
- Train on a larger dataset.
- Use an exact focal loss.
- Use a block loss to try to keep colors more uniform.