# Image colorization with Deep Learning
AML Project

Michele Conti 1599133

Marco Aurelio Sterpa 1152419

**Abstract**

In this report, we explain the main aspects and features of a convolutional neural-network based system[1] that attacks the image colorization problem. Then, we compare classification and regression based models to see which one shines against the other.

Image colorization is an intrinsic multimodal problem, in the sense that several solutions are plausible for a single input. For this reason, we pose our task as a multi-label classification one, and exploit class-rebalancing during training to produce more realistic colorizations.

## 1 Introduction

Given a grayscale image, the task to find a *plausible* colorization has been previously tackled using a classic regression approach, where the goal is to minimize the error between an estimate and the ground truth. This approach tends to result in desaturated images due to the conservative estimates produced by loss functions derived from the Euclidean distance. This motivated the work of [5] to use a different loss function that takes into account the multimodal nature of image colorization, where often there is a set of plausible colors (e.g., hair could be black, blonde, etc.).

This is the structure of what follows: in 2 we outline the models architecture and formally define the loss functions, in 3 we provide experimental results on the performance of different models on the datasets used. Finally, in 4 we present the results of our best models and the limits it retains.

## 2 Approach

We train two different CNNs sharing a similar architecture up to the last layer, both taking as input a grayscale image and outputting (1) a probability distribution of possible colors and (2) a specific color in the ab space.

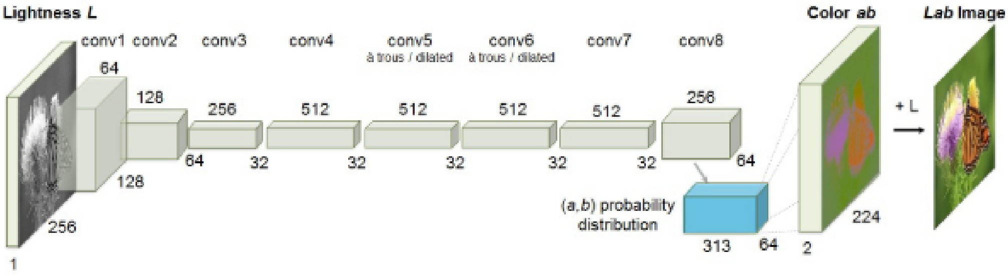


Figure 1: Network architecture. Each `conv` layer refers to a block of 2 or 3 repeated `conv` and `ReLU` layers, follower by a `BatchNorm` and a `Dropout(0.1)`

Images are considered in the Lab space, which was initially designed so that the euclidean distance between coordinates of colors would approximate the human perception of colors distance. The L channel is the luminosity of the image, i.e. the grayscale, while the ab channels represent the colors.

Formally, we are given in input a lightness channel $\mathbf{X} \in \mathbb{R}^{H \times W \times 1}$ and the goal is to learn an estimate $\hat{\mathbf{Y}}$ of the true $\mathbf{Y} \in \mathbb{R}^{H \times W \times 2}$ ab channels, where H, W are the dimensions of the image.

[1] https://github.com/mikcnt/aml-project

To compare its results, we train the model with both a regression and a classification loss. For the regression model, we use the usual Euclidean loss $L_2(\cdot)$, defined as:

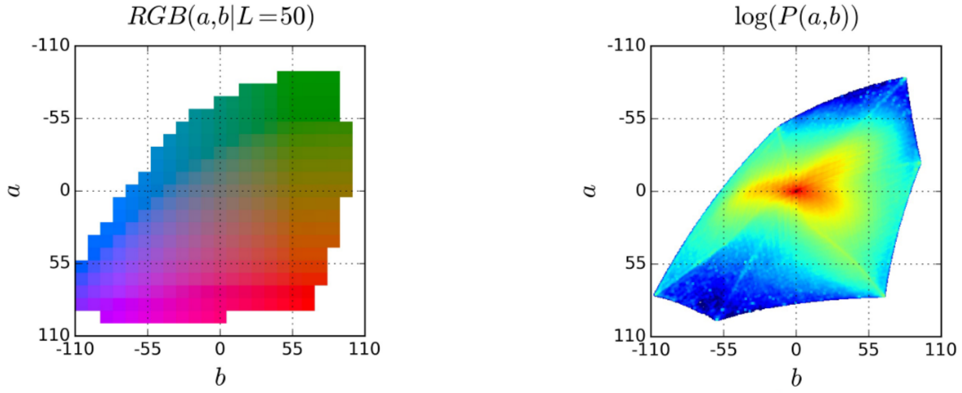
$$L_2(\hat{\mathbf{Y}}, \mathbf{Y}) = \frac{1}{2} \sum_{h,w} \left\| \mathbf{Y}_{h,w} - \hat{\mathbf{Y}}_{h,w} \right\|_2^2 \tag{1}$$

This loss presents two main problems: (1) if an object can take on a set of distinct ab values the result is the mean which favors desaturated values. This happens because the model will try to avoid high penalizations, that are usually given when trying to infer strong colorizations. Then (2) if this set is non-convex, the result will be out of the set, hence implausible.

On the other hand, the classification model is trained with a cross entropy loss function $L_{cl}(\cdot)$, defined as:

$$L_{cl}(\hat{\mathbf{Z}}, \mathbf{Z}) = -\sum_{h,w} v(\mathbf{Z}_{h,w}) \sum_{q} \mathbf{Z}_{h,w,q} \log(\hat{\mathbf{Z}}_{h,w,q}) \tag{2}$$

where $v(\cdot)$ is a weighting term used to rebalance the loss based on color-class rarity, since low ab values are much more frequent in nature (backgrounds such as grass, clouds, etc.) and $q$ ranges over the 313 bins in the quantized $ab$ space.



(a) Quantized ab color space with a grid size of 10. A total of 313 $ab$ pairs are in gamut.

(b) Empirical probability distribution of $ab$ values in log scale.

Figure 2: Quantized ab color space.

To obtain a point estimate in the ab space, the color distribution is interpolated by re-adjusting the temperature ($T = 0.38$ as in the original work) of the softmax distribution and then taking the mean:

$$\mathcal{H}(\mathbf{Z}_{h,w}) = \mathbb{E}[f_T(\mathbf{Z}_{h,w})], \qquad f_T(\mathbf{z}) = \frac{\exp(\log(\mathbf{z})/T)}{\sum_q \exp(\log(\mathbf{z}_q)/T)} \tag{3}$$

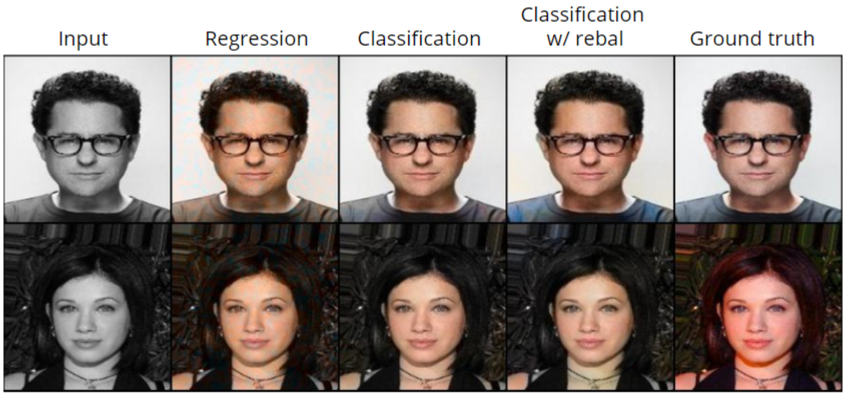
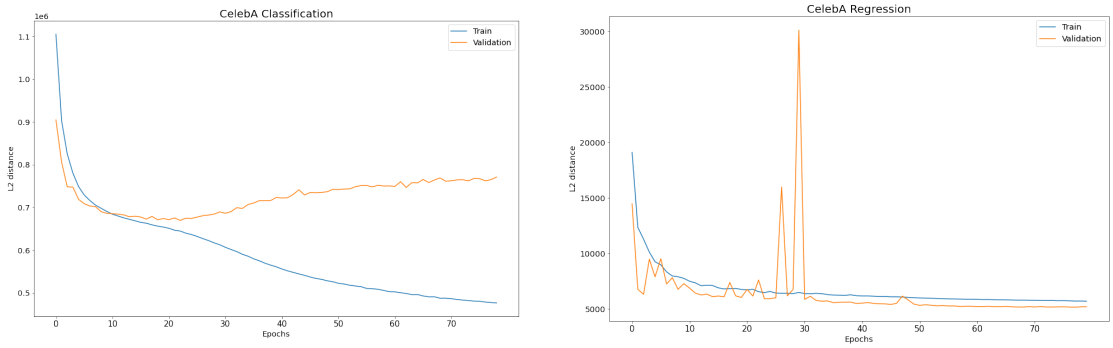


Figure 3: Colorization of the models on successful instances

# 3 Experiments

We train both networks on the CelebA datasets [3] reduced to 7.2k images and a Fruits dataset manually created downloading 1.2k images, using a 60/20/20 split for train/val/test sets. Each model has been trained for 100 epochs, we notice in the classification model losses tend to diverge around epoch 30 despite a continuous improvement in colorization output qualitatively evaluated.



(a) Loss divergence on classification with rebalancing.

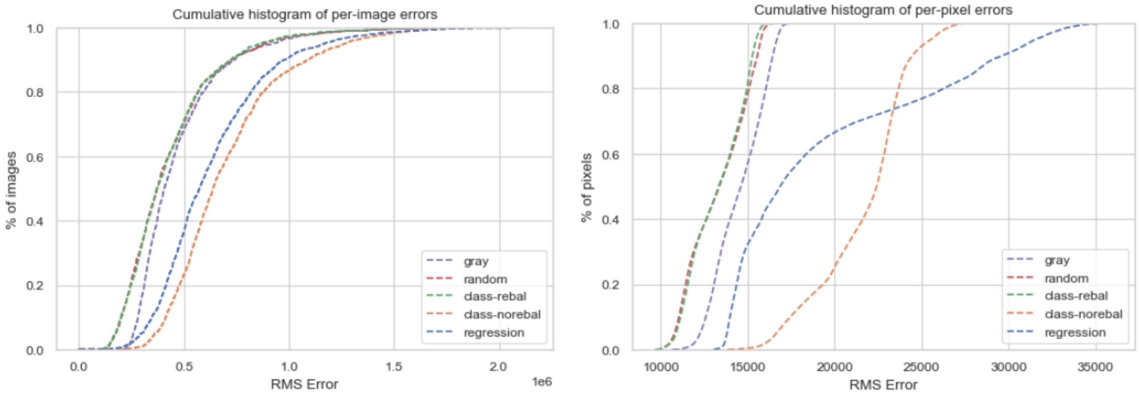
(b) Non-divergent behaviour of L2 loss.

Figure 4: Training and validation loss for different models.

In line with [5], we use two measures of accuracy:

1. We take the L2 distance of pixels in the ab space and then integrate the area under the curve defined by the percentage of pixels below different threshold values.
2. We look at the accuracy of a VGG pretrained classifier on the color estimates produced by the model, for this we needed to use the Fruits dataset because we couldn't apply this test on CelebA.

For the distances obtained in (1) we also plot the cumulative histograms of errors per image and pixel on the CelebA datasets, we observe that a random colorization produces good results due to the nature of the images where a large portion of the pixels are used for skin and hair.



(a) Cumulative histogram error per image.

(b) Cumulative histogram error per pixel.

Figure 5: Cumulative histogram errors.

We test the accuracy against different estimates:

1. **Gray**: Colors every pixel gray, i.e. $(a, b) = 0$.
2. **Random**: Colors obtained by picking a random image from the training set.
3. **L2**: Regression model with L2 loss.
4. **Class**: Classification model without class rebalancing.
5. **Class+rebal**: Classification model with class rebalancing.

| **Dataset** | **CelebA** | **Fruits** | |
|---|---|---|---|
| **Method** | **AuC (%)** | **AuC (%)** | **VGG Acc** |
| Ground Truth | 100 | 100 | 44.0 |
| Gray | 88.8 | 76.9 | 33.2 |
| Random | 87.4 | 74.1 | 28.2 |
| L2 | 91.5 | 76.9 | 31.7 |
| Class | 92.2 | **84.5** | 35.9 |
| Class+rebal | **92.3** | 84.3 | **39.8** |

Table 1: Accuracy results for Area under the Curve and VGG classification.

# 4 Conclusion

Overall, the outputs of the classification with rebalancing model outperform the other approaches both from a performance standpoint according to the measures used and from a visual qualitative assessment.

Finally, we observe that the images produced are susceptible to some form of failure, namely:

1. **Color biases**: Plausible but less occurrent colors are ignored.
2. **Color deficiency**: Colors appear too desaturated or too uniform across the image
3. **Inconsistent coloring**: Spatially inconsiste colors such as shift or artifacts
4. **Edge pollution**: Incosistent color around the edge of the image
5. **Color bleeding**: Object's colors spill over its intended boundary

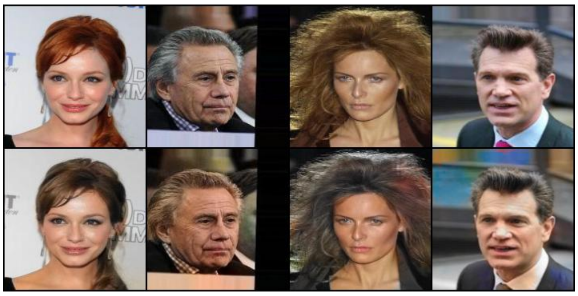


Figure 6: Examples of colorization failure. From left to right: color bias, color deficiency, inconsistent coloring, color pollution/edge bleeding.

# References

[1] S. Guadarrama, R. Dahl, D. Bieber, M. Norouzi, J. Shlens, and K. Murphy. Pixcolor: Pixel recursive colorization. *arXiv*, pages 1–17, 2017.

[2] C. Li and M. Wand. Combining markov random fields and convolutional neural networks for image synthesis, 2016.

[3] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.

[4] S. Pahal and P. Sehrawat. Image colorization with deep convolutional neural networks. In G. S. Hura, A. K. Singh, and L. Siong Hoe, editors, *Advances in Communication and Computational Technology*, pages 45–56, Singapore, 2021. Springer Singapore.

[5] R. Zhang, P. Isola, and A. A. Efros. Colorful image colorization. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9907 LNCS:649–666, 2016.