CORE: Color Regression for Multiple Colors Fashion Garments

Alexandre Rame*1,2, Arthur Douillard1,2, and Charles Ollion2

1Sorbonne Université, CNRS, LIP6, 4 place Jussieu, 75005 Paris
2Heuritech, 134 Rue Legendre, 75017 Paris

October 29, 2019

Abstract

Among all fashion attributes, color is challenging to detect due to its subjective perception. Existing classification approaches can not go beyond the predefined list of discrete color names. In this paper, we argue that color detection is a regression problem. Thus, we propose a new architecture, based on attention modules and in two-stages. The first stage corrects the image illumination while detecting the main discrete color name. The second stage combines a colorname-attention (dependent of the detected color) with an object-attention (dependent of the clothing category) and finally weights a spatial pooling over the image pixels’ RGB values. We further expand our work for multiple colors garments. We collect a dataset where each fashion item is labeled with a continuous color palette: we empirically show the benefits of our approach.

1. Introduction

Convolutional Neural Networks (CNNs) [26, 20] generated a lot of interest in the fashion industry. Recent datasets of fashion images [28, 49, 17] encouraged various approaches for attributes classification [7, 40], visual search [44, 19, 30, 13] and object detection [23, 34].

One of the key attributes of a fashion item is its color. However, colors are subjective properties of garments as not all humans recognize colors the same way [32]: their automatic estimation is therefore challenging. So far, most approaches consider the problem through the angle of discrete color naming: a classifier chooses amongst 11 colors of the English language [4], based on features from histograms [2, 9, 39] or from CNNs [42, 33, 10, 46, 45, 21, 24]. Recent approaches increased the number of color names up to 28 [47] or 313 [48]. We take a step forward and tackle the problem through the angle of continuous color regression to handle its inherent ambiguity and multimodality. Rather than predicting an approximate discrete color name, we aim at predicting the exact continuous RGB color value. This refined representation is necessary for many industrial applications such as precise visual search and fine-grained trend detection.

Factors such as complex garments structure and varying illumination spark discrepancies between the raw color and what we perceive: the human brain automatically focuses on the right regions to detect the color and corrects the contrast. First, our approach needs to focus on the interesting regions of the image, by reducing the impact of complex backgrounds or clutters [43, 10, 46]. Second, we need color constancy so that scene’s illumination is ideal white light [5, 3, 29, 6, 36, 15, 22, 37].

Our architecture tackles all above challenges end-to-end simultaneously: (1) a semantic segmenter learns to detect fashion garments, (2) a color constancy module learns to correct the illumination of the image, and finally (3) we predict colors’ RGBs by refining the rough discrete color names through weighted pooling over pixels.

2. Model

The network in Figure 1 is trained end-to-end on the task of continuous color regression, with neither illuminations nor clothing segmentation ground-truth annotations.

2.1. Pixel Image Correction

Our model needs to correct the bias in the pixels of the image. All the images are first contrast normalized in a preprocessing step by global histogram stretching in order to be robust across various lighting conditions. Note that this does not change the relative contributions of the three RGB channels. Second, to remove illumination color casts, we need to detect the initial illumination of the image (a scalar value per channel) and then correct each pixel in the image using the Von Kries method [41]. Following [29] and because of its capacity to extract style attributes [16], we choose the VGG16 [38] neural architecture.
2.2. Spatial Pooling over Pixels for Continuous Color Regression

To sample only the interesting regions of the image, we apply a spatial pooling over the pixels of the image weighted by two complementary attention modules: the first focuses on the clothing category, the second on the color name detected. These attention masks will be multiplied to create the combined-attention.

2.2.1 Through Object-Attention

Following the work from [46], we use the popular fully convolutional Deeplabv3 [8]. This attention network is pre-trained on the task of semantic segmentation over the clothing crops from Modanet [49] and therefore has an emphasis on the garment surface: it produces an object-attention different for each clothing category (top, coat, ...). This spatial prior is then fine-tuned and used to modulate the importance of each pixel: it identifies the regions which contain the relevant color information given a garment type.

2.2.2 Through Colorname-Attention

The previous object-attention will fail for multiple colors garments. If the apparel can take on a set of distinct and distant RGB values, the mean of this set will be predicted: it leads to unsaturated colors predictions. In Figure 1, we need to discard the white background in the shirt that would otherwise lead to predict a light pink. Thus we add a colorname-attention, that only selects pixels in the image sufficiently close to the main color.

In the mono color setup, we first need to detect the main discrete color name of the garment. The VGG16 features are followed by a fully connected layer with a softmax activation function [42, 33, 10, 46, 45]: it predicts a distribution over the 72 color names, and is trained by minimizing a categorical cross-entropy loss. If the color name with maximum score has color name $\text{RGB}_c$, given a pixel $p$ of RGB value $\text{RGB}_p$, the colorname-attention $CA_p$ for pixel $p$ is:

$$CA_p \propto \exp \left( -\frac{1}{127.5^2} \sum_{c=1}^{72} (\text{RGB}_p - \text{RGB}_c)^2 \right).$$

It sums to 1 over all pixels and its peakedness depends on $T$, the temperature of the $\text{RGB Spatial Softmax}$. $T$ is the only parameter of the colorname-attention module, which can be either a predefined hyper-parameter, either learned, either input-dependant and predicted with a fully connected layer from VGG features.

2.3. Training Overview

The LAB colorspace best models perceptual distance in colorization [48, 11]: thus, the chosen regression objective is the Euclidean distance between predicted and ground truth colors converted to LAB. Our two losses, color naming and color regression, are back propagated simultaneously.

2.4. Multiple Colors

We can easily generalize our approach for multiple RGB colors. We predict multiple color names with a sigmoid [45] followed by a combination of categorical cross entropy and binary cross entropy [34]. Multiple colors share the same object-attention but have different colorname-attention: therefore we predict different RGBs.

The challenge is then to know how many colors should be predicted. Following the work from [45], we explicitly learn the number of colors in each garment (1, 2 or 3 colors). Note that we apply class weights for handling unbalanced classes.

Selecting the color names with maximum scores would often predict several times the same major color: for example a light red and a dark red would converge towards a
Table 1. Score for RGB regressions: percentage of predictions closer to the ground truths RGBs than several thresholds (10, 20, 30, 40), according to the deltaE cie2000 distance [14, 31]. Bold highlights best score.

| Method                          | Main Color | Multiple Colors |
|---------------------------------|------------|-----------------|
| Unsupervised K-means Clustering [27] | ≤ 10 | ≤ 10 |
| Color naming [45]               | 47 72 83 89 | 19 31 37 42 |
| Object-Attention                | ✓ ✓ ✓ ✓ | ✓ ✓ ✓ ✓ |
| Colorname-Attention             | ✓ ✓ ✓ ✓ | ✓ ✓ ✓ ✓ |
| Combined-Attention              | ✓ ✓ ✓ ✓ | ✓ ✓ ✓ ✓ |
| Combined-Attention & Illumination | ✓ ✓ ✓ ✓ | ✓ ✓ ✓ ✓ |

**2.5. Discussion: Analogy with Faster R-CNN**

Our two-stages approach is highly inspired from Faster R-CNN [35]. The anchors of the Region Proposal Network are replaced by color names. Given a selected anchor (color name), the final regression adds a small continuous offset leading to a precise color quantified as RGB values. In both cases, the first classifier gives a rough estimate, refined by the second regression stage.

**3. Experiments**

**3.1. Setup**

**Dataset** As far as we know, there is no dataset for the task of continuous colors regression. Thus, we have collected a new dataset of 30,269 fashion garments: 5,363 coats, 8,166 dresses, 3,991 pants, 6,871 shoes and 5,878 tops. The validation dataset and the test dataset are composed of 2000 images, the remaining images are used for training. Each garment was labeled by a single operator with its exact RGBs colors values, not like it appears in the image, but like the operator though it would appear in ideal white light: the annotation process is therefore more complex and time-consuming than classical classification. The color names can automatically be derived by nearest neighbours in the LAB space. In case of multiple colors items, the operators were asked to tag them in decreasing order of importance.

**Implementation** Our code is in Tensorflow [1] and Keras [12]. We chose an Adam [25] optimizer with a learning rate of 0.0001 and a batch of 16 images during 50 epochs. We applied standard data augmentations methods: random cropping and translation.

**3.2. Results**

Table 1 shows our experiments done on our internal dataset for the regression task, in the main color and in the multiple colors setups.

**Baselines** We compare our model against two existing baselines. First we train an unsupervised K-means pixels clustering [27] directly on the pixel space, after a background removal done by an external semantic segmentation model trained on Modanet [49]. Colors are ranked according to the number of pixels in each cluster. The second baseline, Colorname RGB, is the direct extension of the multitagger approach from Yazici et al. [45]: it detects the color names (e.g. velvet red, dark purple, etc.) and then produces the associated RGB values (e.g. (134, 71, 71) for velvet red).

**Attention** First, fine-tuning the semantic segmentation model on the regression task (object-attention) already surpasses previous approaches in the mono color setup. The colorname-attention improves overall results in the multiple colors setup. These two attentions, when combined, are mutually reinforcing and complementary.

**Temperature** We show that the temperature $T$ value is important: bigger $T$ leads to sharp distribution and takes into account fewer pixels from the initial images that with a lower $T$. Rather than grid-searching its optimal value, it can be learned for improved results. The optimal $T$ depends on the image: therefore, best results are obtained with $T$ predicted from VGG features.

**Illumination** Finally, the color constancy improves performances. It may be more impactful with pretraining on the Color Checker Dataset [18].

**4. Conclusion**

In this paper we address the problem of detecting refined multiple color continuous values for fashion garments. We collect a unique dataset of 30,269 fashion garments and empirically show the benefits of our two-stages architecture. Finally, we hope to motivate researchers to investigate the different subjective properties of fashion garments.

**Acknowledgements** This work is supported by Heuritech, which forecasts fashion trends to produce more sustainably.
[32] T. R. Partos, S. J. Cropper, and D. Rawlings. You don’t see what i see: Individual differences in the perception of meaning from visual stimuli. PloS one, 11(3):e0150615, 2016.

[33] R. F. Rachmadi and I. Purnama. Vehicle color recognition using convolutional neural network. arXiv preprint arXiv:1510.07391, 2015.

[34] A. Rame, E. Garreau, H. Ben-Younes, and C. Ollion. Omnia faster r-cnn: Detection in the wild through dataset merging and soft distillation. arXiv preprint arXiv:1812.02611, 2018.

[35] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems, pages 91–99, 2015.

[36] W. Shi, C. C. Loy, and X. Tang. Deep specialized network for illuminant estimation. In European Conference on Computer Vision, pages 371–387. Springer, 2016.

[37] O. Sidorov. Artificial color constancy via googlenet with angular loss function. arXiv preprint arXiv:1811.08456, 2018.

[38] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556, 2014.

[39] J. Van De Weijer, C. Schmid, J. Verbeek, and D. Larlus. Learning color names for real-world applications. IEEE Transactions on Image Processing, 18(7):1512–1523, 2009.

[40] A. Veit, B. Kovacs, S. Bell, J. McAuley, K. Bala, and S. Belongie. Learning visual clothing style with heterogeneous dyadic co-occurrences. In Proceedings of the IEEE International Conference on Computer Vision, pages 4642–4650, 2015.

[41] J. von Kries. Chromatic adaptation, festschrift der albercht-ludwig-universität, 1902.

[42] Y. Wang, J. Liu, J. Wang, Y. Li, and H. Lu. Color names learning using convolutional neural networks. In 2015 IEEE International Conference on Image Processing (ICIP), pages 217–221. IEEE, 2015.

[43] W. Wieclawek and E. Pietka. Car segmentation and colour recognition. In 2014 Proceedings of the 21st International Conference Mixed Design of Integrated Circuits and Systems (MIXDES), pages 426–429. IEEE, 2014.

[44] K. Yamaguchi, M. Hadi Kiapour, and T. L. Berg. Paper doll parsing: Retrieving similar styles to parse clothing items. In Proceedings of the IEEE international conference on computer vision, pages 3519–3526, 2013.

[45] V. O. Yazici, J. van de Weijer, and A. Ramisa. Color naming for multi-color fashion items. In World Conference on Information Systems and Technologies, pages 64–73. Springer, 2018.

[46] L. Yu, Y. Cheng, and J. van de Weijer. Weakly supervised domain-specific color naming based on attention. In 2018 24th International Conference on Pattern Recognition (ICPR), pages 3019–3024. IEEE, 2018.

[47] L. Yu, L. Zhang, J. van Weijer, F. S. Khan, Y. Cheng, and C. A. Parraga. Beyond eleven color names for image understanding. Machine Vision and Applications, 29(2):361–373, 2018.

[48] R. Zhang, P. Isola, and A. A. Efros. Colorful image colorization. In European conference on computer vision, pages 649–666. Springer, 2016.

[49] S. Zheng, F. Yang, M. H. Kiapour, and R. Piramuthu. Modanet: A large-scale street fashion dataset with polygon annotations. arXiv preprint arXiv:1807.01394, 2018.