Metric Learning as Convex Combinations of Local Models with Generalization Guarantees

Supplementary Material

Valentina Zantedeschi
valentina.zantedeschi@univ-st-etienne.fr

Rémi Emonet
remi.emonet@univ-st-etienne.fr

Marc Sebban
marc.sebban@univ-st-etienne.fr

Univ Lyon, UJM-Saint-Etienne, CNRS, Institut d Optique Graduate School,
Laboratoire Hubert Curien UMR 5516, F-42023, SAINT-ETIENNE, France

1. Overview

This supplementary material is organized as follows: in Section 2, we provide the proof of the generalization bound expressed in Lemma 2 of Section 4.1 of the paper; in Section 3, we present a visual comparison of C2LM and a local metric learning approach for the perceptual color distance application; Lastly, Section 4 is dedicated to image segmentation, performed using the color distance learned by means of C2LM.

2. Generalization Guarantees

We recall that $\mathcal{Z} = X \times Y$ is the set of all possible valued pairs $p = (x_1, x_2, y(x_1, x_2))$, where $(x_1, x_2) \in X = U^2$ is a pair of instances and $y(x_1, x_2)$ is the associated target value, and that we also denote $P = \{p_i\}_{i=1}^n \subset \mathcal{Z}$ the set of n training pairs. We partitioned the space X into $\mathcal{N}(\gamma_1/2, X, \|.\|_2)$ subsets and the space Y into $\mathcal{N}(\gamma_2/2, Y, |.|)$, so that any region of X (resp. Y) has a diameter smaller than γ_1 (resp. γ_2).

We also recall the definition of our optimization problem, considering that the space U has been decomposed into K regions denoted $\{R_z\}_{z=1}^K$:

$$\arg\min_W F_P(W) = \hat{R}^l + \lambda_1 D(W) + \lambda_2 S(W)$$
$$s.t.\, \forall i, j = 1, ..., K : \sum_{z=1}^{K} W_{ijz} = 1 \; and \; W_{ij} \geq 0 \tag{1}$$

where W_{ij} is the vector of non-negative weights associated to pair of regions $R_{ij} = (R_i, R_j)$,

$$\hat{R}^l = \frac{1}{n} \sum_{i,j,p \in R_{ij}} l(W, p) = \frac{1}{n} \sum_{i=1}^{K} \sum_{j=1}^{i} \sum_{p \in R_{ij}} \left| \sum_{z=1}^{K} W_{ijz} s_z(x_1, x_2) - y(x_1, x_2) \right| \tag{2}$$

is the mean loss over all training pairs, with $\{s_z(.)\}_{z=1}^K$ the set of metric functions related to the local models, and

$$D(W) = \sum_{i=1}^{K} \sum_{j=1}^{i} \left\| E_{ij}^T W_{ij} \right\|_F^2 \tag{3}$$

$$S(W) = \sum_{i=1}^{K} \sum_{j=1}^{i} \sum_{i'=1}^{K} \sum_{j'=1}^{i'} K_{iji'j'} \left\| W_{ij} - W_{i'j'} \right\|_2^2 \tag{4}$$

are the two regularizers used to avoid overfitting, and λ_1 and λ_2 are the two hyper-parameters.

In our paper, we proved that, supposing $s_z(.)\ \forall z = 1, ..., K$ to be θ_z-lipschitz w.r.t. the norm $\|.\|_2$, the previous problem is $(H, \theta\sqrt{2}\gamma_1 + \gamma_2)$-robust, with $\theta = \max_{z=1..K} \theta_z$ (see Lemma 1 of Section 4.1 of the paper).

We denote R^l the true loss $R^l = \mathbb{E}_{p\sim\mathcal{Z}} l(W, p)$ and $\hat{R}^l$ the empirical loss $\hat{R}^l = \mathbb{E}_{p\sim P} l(W, p)$ corresponding to the mean loss over all the training pairs. We can now derive a PAC generalization bound for our problem, considering the theoretical distribution of the valued pairs $p \in \mathcal{Z}$ over the regions of the partition given in the paper (see Prop. 1).

Lemma 2 As $F_P(W)$ is $(H, \theta\sqrt{2}\gamma_1 + \gamma_2)$-robust and the training set P is obtained from n IID draws according to a multinomial random variable, for any $\delta > 0$ with probability at least $1 - \delta$, we have:

$$|R^l - \hat{R}^l| \leq \theta\sqrt{2}\gamma_1 + \gamma_2 + B\sqrt{\frac{2H\ln 2 + 2\ln 1/\delta}{n}}\,, \tag{5}$$

with B the upper bound of the loss function $l(W, p) = l(W_{ij}, (x_1 \in R_i, x_2 \in R_j, y(x_1, x_2))) = |\sum_{z=1}^{K} W_{ijz} s_z(x_1, x_2) - y(x_1, x_2)|$.

Proof.

$$\begin{aligned}
&|R^l - \hat{R}^l| = \\
&= \left|\mathbb{E}_{p\sim\mathcal{Z}} l(W, p) - \frac{1}{n}\sum_{p'\sim P} l(W, p')\right| \\
&= \left|\sum_{i=1}^{K}\sum_{j=1}^{K} \mathbb{E}[l(W_{ij}, p \in R_{ij})]p(R_{ij}) - \frac{1}{n}\sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{p'\in R_{ij}} l(W_{ij}, p')\right| \\
&= \left|\sum_{i=1}^{K}\sum_{j=1}^{K} \mathbb{E}[l(W_{ij}, p \in R_{ij})]p(R_{ij}) - \sum_{i=1}^{K}\sum_{j=1}^{K} \mathbb{E}[l(W_{ij}, p \in R_{ij})]\frac{n_{ij}}{n}\right| \\
&\quad + \left|\sum_{i=1}^{K}\sum_{j=1}^{K} \mathbb{E}[l(W_{ij}, p \in R_{ij})]\frac{n_{ij}}{n} - \frac{1}{n}\sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{p'\in R_{ij}} l(W_{ij}, p')\right| \qquad (6) \\
&\leq \left|\sum_{i=1}^{K}\sum_{j=1}^{K} \mathbb{E}[l(W_{ij}, p \in R_{ij})](p(R_{ij}) - \frac{n_{ij}}{n})\right| + \left|\frac{1}{n}\sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{p,p'\in R_{ij}} \mathbb{E}[l(W_{ij}, p)) - l(W_{ij}, p')]\right| \\
&\leq \max(l(W_{ij}, p \in R_{ij}))\sum_{i=1}^{K}\sum_{j=1}^{K}\left|p(R_{ij}) - \frac{n_{ij}}{n}\right| + \frac{1}{n}\sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{p,p'\in R_{ij}} \max(l(W_{ij}, p) - l(W_{ij}, p')) \\
&\leq B\sqrt{\frac{2H\ln 2 + 2\ln(1/\delta)}{n}} + \theta\sqrt{2}\gamma_1 + \gamma_2. \qquad (7)
\end{aligned}$$

Eq. 6 is due to the triangle inequality. The first term of Eq. 7 is because B is the upper bound of the loss function and because of the Bretagnolle-Huber-Carol inequality (see Proposition 1 of the paper), and the latter is due to the robustness of the problem. □

3. Illustration of Learned Combinations

In this section, we illustrate a metric learned using C2LM and compare it with the one learned using a local metric learning approach.

In the context of the perceptual color distance, Fig.1 shows a 2D projection of the contour lines of our learned combination of metrics, drawn around an arbitrary point, in the RGB space. While the method from [3] causes a strong discontinuity at the boundaries of the cluster (because one jumps from a local metric to the global one), we can see that our combination is smoother. In addition, it is evident that, while comparing points not belonging to the same region, our metric is more accurate because our method captures better the geometric variations of the space than a global linear metric.

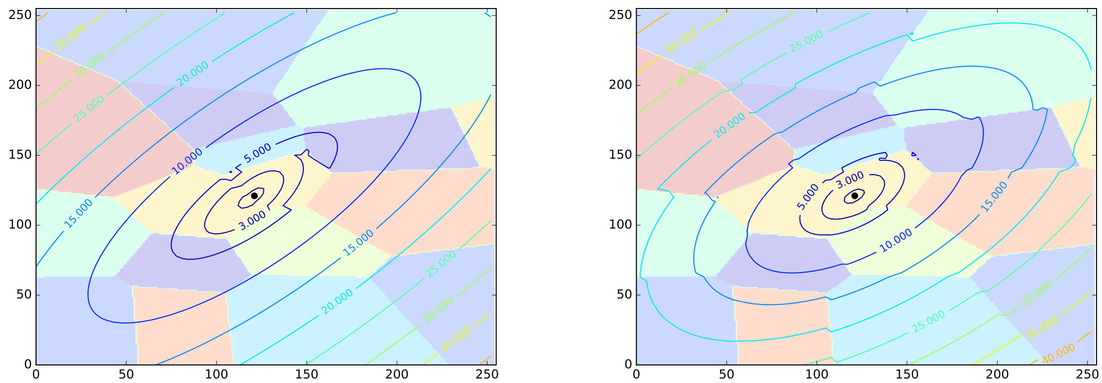

Figure 1: On the left: illustration of the discontinuity of [3]'s method; on the right: illustration of the smoothness of our learned metrics.

4. Image Segmentation

In this section, we present some examples of images segmented using the perceptual color distance learned in Section 5.1 of the paper.
The task of image segmentation consists in partitioning an image into regions according to a color distance: two adjacent pixels are assigned to the same region if their color distance is smaller than a given threshold value. It is clear, then, that the quality of the obtained segmentation relies on the quality of the used distance, this is why the latter should be as close as possible to the one perceived by a human observer.

For our experiments, among all the possible methods to perform such a task, we make use of the Color Mean-Shift algorithm presented in [1], which allows one to cluster the pixels of an image using different distance functions between colors. We compare, then, the Color Mean-Shift method using the Euclidean distance directly on the RGB components and using the perceptual color distance learned with C2LM in Section 5.1 of the paper. In Fig. 2 we show the results on some pictures extracted from the Berkeley dataset [2]: for each image, we have at our disposal a ground truth segmentation computed as the average of the segmentations provided by 30 different human subjects. Notice that, for each result of the Color Mean-Shift method, we mentioned between brackets first the total number of computed clusters and second the number of clusters larger than 150 pixels. As a matter of fact, the used method detects several small segments that a human observer could not even see. For this reason and because we cannot fix the number of segments to find, we rather compare the two distances on segmentations that have a number of significant clusters (the second value) close to the number of segments of the ground truth.

References

[1] K. Bitsakos, C. Fermüller, and Y. Aloimonos. An experimental study of color-based segmentation algorithms based on the mean-shift concept. In *Computer Vision–ECCV 2010*, pages 506–519. Springer, 2010. 3

[2] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. 8th Int'l Conf. Computer Vision*, volume 2, pages 416–423, July 2001. 3

[3] M. Perrot, A. Habrard, D. Muselet, and M. Sebban. Modeling perceptual color differences by local metric learning. In *ECCV*, pages 96–111. Springer, 2014. 2, 3

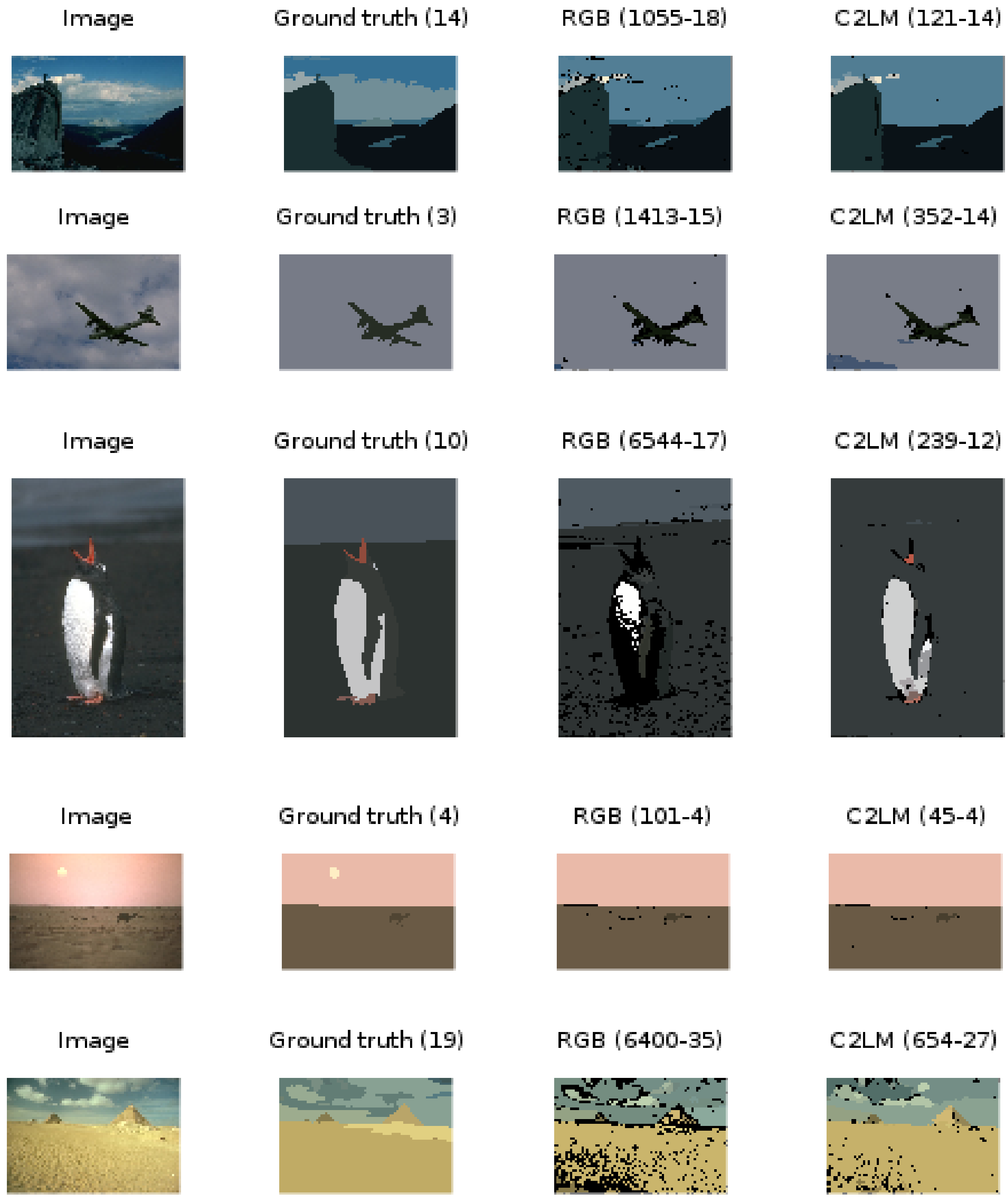

Figure 2: Comparison of the segmentations resulting from the computation of Color Mean-Shift algorithm using the Euclidean distance on the RGB components and using the perceptual distance learned with C2LM.