Modeling Perceptual Color Differences by Local Metric Learning

Supplementary Material - Paper ID 922

1 Overview of the supplementary material

This supplementary material is organised into two parts. In Section 2 we provide the proofs of the lemmas and the theorem presented in Section 3.3 of the paper, while Section 3 presents some examples of image segmentation.

2 Theoretical analysis

This section presents the proofs of Lemma 1 and Theorem 1 from Section 3.3 of the paper. Lemma 1 is proved in Section 2.1 and Theorem 1 is proved in Section 2.2.

2.1 Generalization bound per region C_j

First, we recall our optimization problem considered in each region C_j:

$$\underset{\mathbf{M}_j \succeq 0}{\arg\min} F_{T_j}(\mathbf{M}_j) \quad (1)$$

where

$$F_{T_j}(\mathbf{M}_j) = \hat{\varepsilon}_T(\mathbf{M}_j) + \lambda_j \|\mathbf{M}_j\|^2_{\mathcal{F}},$$
$$\hat{\varepsilon}_T(\mathbf{M}_j) = \frac{1}{n_j} \sum_{(\mathbf{x},\mathbf{x}',\Delta E_{00}) \in T} l(\mathbf{M}_j, (\mathbf{x}, \mathbf{x}', \Delta E_{00})),$$
$$\text{and } \; l(\mathbf{M}_j, (\mathbf{x}, \mathbf{x}', \Delta E_{00})) = \left| (\mathbf{x} - \mathbf{x}')^T \mathbf{M}_j (\mathbf{x} - \mathbf{x}') - \Delta E_{00}\left(\mathbf{x}, \mathbf{x}'\right)^2 \right|.$$

Here $\hat{\varepsilon}_T(\mathbf{M}_j)$ stands for the empirical risk of a matrix $\mathbf{M}_j$ over a training set T_j, of size n_j, drawn from an unknown distribution $P(C_j)$. The true risk $\varepsilon_{P(C_j)}(\mathbf{M}_j)$ is defined as follows:

$$\varepsilon_{P(C_j)}(\mathbf{M}_j) = \mathbb{E}_{(\mathbf{x},\mathbf{x}',\Delta E_{00}) \sim P(C_j)} \left[l(\mathbf{M}_j, (\mathbf{x}, \mathbf{x}', \Delta E_{00})) \right].$$

In this section, $T_j[i]$ denotes the training set obtained from T_j by replacing the i^{th} example of T_j by a new independent one. Moreover, we have $\Delta_{\max} = \max_{(\mathbf{x},\mathbf{x}',\Delta E_{00}) \sim P} \left\{ \Delta E_{00}\left(\mathbf{x}, \mathbf{x}'\right) \right\}$ and $D_j = \max_{(\mathbf{x},\mathbf{x}',\Delta E_{00}) \sim P(C_j)} \left(\|\mathbf{x} - \mathbf{x}'\| \right) \leq 1$[1].

To derive such a generalization bound, we need to consider loss functions that fulfill two properties: k-lipschitz continuity (Definition 1) and (σ, m)-admissibility (Definition 2).

Definition 1 (k-lipschitz continuity). *A loss function $l(\mathbf{M}, (\mathbf{x}, \mathbf{x}', \Delta E_{00}))$ is k-lipschitz w.r.t. its first argument if, for any matrices $\mathbf{M}$, $\mathbf{M}'$ and any example $(\mathbf{x}, \mathbf{x}', \Delta E_{00})$, there exists $k \geq 0$ such that:*

$$|l(\mathbf{M},(\mathbf{x}, \mathbf{x}', \Delta E_{00})) - l(\mathbf{M}',(\mathbf{x}, \mathbf{x}', \Delta E_{00}))| \leq k \|\mathbf{M} - \mathbf{M}'\|_{\mathcal{F}}.$$

This k-lipschitz property ensures that the loss deviation does not exceed the deviation between matrices $\mathbf{M}$ and $\mathbf{M}'$ with respect to a positive constant k.

[1] We assume the examples to be normalized such that $\|\mathbf{x}\| \leq 1$.

Definition 2 **((σ, m)-admissibility).** *A loss function $l(\mathbf{M_j}, (\mathbf{x}, \mathbf{x}', \Delta E_{00}))$ is (σ, m)-admissible, w.r.t. $\mathbf{M_j}$, if it is convex w.r.t. its first argument and for two examples $(\mathbf{x}, \mathbf{x}', \Delta E_{00}(\mathbf{x}, \mathbf{x}'))$ and $(\mathbf{t}, \mathbf{t}', \Delta E_{00}(\mathbf{t}, \mathbf{t}'))$, we have:*

$$\left| l(\mathbf{M_j}, (\mathbf{x}, \mathbf{x}', \Delta E_{00}(\mathbf{x}, \mathbf{x}')) - l(\mathbf{M_j}, (\mathbf{t}, \mathbf{t}', \Delta E_{00}(\mathbf{t}, \mathbf{t}')))\right| \leq \sigma \left| \Delta E_{00}(\mathbf{x}, \mathbf{x}') - \Delta E_{00}(\mathbf{t}, \mathbf{t}')\right| + m.$$

Definition 2 bounds the difference between the losses of two examples by a value only related to the ΔE_{00} values plus a constant independent from $\mathbf{M_j}$. Let us introduce a last concept which is required to derive a generalization bound.

Definition 3 **(Uniform stability).** *In a region C_j, a learning algorithm has a uniform stability in $\frac{\mathcal{K}}{n_j}$, with $\mathcal{K} \geq 0$ a constant, if $\forall i$,*

$$\sup_{(\mathbf{x}, \mathbf{x}', \Delta E_{00}) \sim P(C_j)} \left| l(\mathbf{M_j}, (\mathbf{x}, \mathbf{x}', \Delta E_{00})) - l(\mathbf{M^i_j}, (\mathbf{x}, \mathbf{x}', \Delta E_{00}))\right| \leq \frac{\mathcal{K}}{n_j},$$

where $\mathbf{M_j}$ is the matrix learned on the training set T_j and $\mathbf{M^i_j}$ is the matrix learned on the training set obtained by replacing the i^{th} example of T_j by a new independent one.

The uniform stability guarantees that the solutions learned with two close training sets are not significantly different and that the variation converges in $O(1/n)$.

To prove our main theorem, we need four additional lemmas and one additional theorem which are not presented in the paper. For the sake of readability, we number these with capital letters.

Lemma A (k-lipschitz continuity) *Let $\mathbf{M}_j$ and $\mathbf{M'}_j$ be two matrices for a region C_j and $(\mathbf{x}, \mathbf{x}', \Delta E_{00})$ be an example. Our loss $l(\mathbf{M}_j, (\mathbf{x}, \mathbf{x}', \Delta E_{00}))$ is k-lipschitz with $k = D_j^2$.*

Proof.

$$\begin{aligned}
&|l(\mathbf{M}_j, (\mathbf{x}, \mathbf{x}', \Delta E_{00})) - l(\mathbf{M'}_j, (\mathbf{x}, \mathbf{x}', \Delta E_{00}))| \\
&\quad = \left| \left| (\mathbf{x} - \mathbf{x}')^T \mathbf{M}_j (\mathbf{x} - \mathbf{x}') - \Delta E_{00}\,(\mathbf{x}, \mathbf{x}')^2 \right| - \left| (\mathbf{x} - \mathbf{x}')^T \mathbf{M'}_j (\mathbf{x} - \mathbf{x}') - \Delta E_{00}\,(\mathbf{x}, \mathbf{x}')^2 \right| \right| \\
&\quad \leq \left| (\mathbf{x} - \mathbf{x}')^T \mathbf{M}_j (\mathbf{x} - \mathbf{x}') - (\mathbf{x} - \mathbf{x}')^T \mathbf{M'}_j (\mathbf{x} - \mathbf{x}') \right| && (2.1) \\
&\quad = \left| (\mathbf{x} - \mathbf{x}')^T (\mathbf{M}_j - \mathbf{M'}_j) (\mathbf{x} - \mathbf{x}') \right| \\
&\quad \leq \|\mathbf{x} - \mathbf{x}'\| \|\mathbf{M}_j - \mathbf{M'}_j\|_{\mathcal{F}} \|\mathbf{x} - \mathbf{x}'\| && (2.2) \\
&\quad \leq D_j^2 \|\mathbf{M}_j - \mathbf{M'}_j\|_{\mathcal{F}} && (2.3)
\end{aligned}$$

Inequality 2.1 is due to the triangle inequality, 2.2 is obtained by application of the Cauchy-Schwarz inequality and some classical norm properties. 2.3 comes from the definition of D_j. Setting $k = D_j^2$ gives the Lemma.

We now provide an additional lemma useful for proving Lemma C on the (σ, m)-admissibility of our loss function.

Lemma B *Let $\mathbf{M}_j$ be an optimal solution of Problem 1, we have*

$$\|\mathbf{M}_j\| \leq \frac{\Delta_{\max}}{\sqrt{\lambda_j}}.$$

Proof. Since $\mathbf{M}_j$ is an optimal solution of Problem 1, we have then:

$$\begin{aligned}
&F_{T_j}(\mathbf{M}_j) \leq F_{T_j}(\mathbf{0}) \\
\Leftrightarrow\ & \frac{1}{n_j} \sum_{(\mathbf{x}, \mathbf{x}', \Delta E_{00}) \in T_j} l(\mathbf{M}_j, (\mathbf{x}, \mathbf{x}', \Delta E_{00})) + \lambda_j \|\mathbf{M}_j\|_{\mathcal{F}}^2 \leq \frac{1}{n_j} \sum_{(\mathbf{x}, \mathbf{x}', \Delta E_{00}) \in T_j} l(\mathbf{0}, (\mathbf{x}, \mathbf{x}', \Delta E_{00})) + \lambda_j \|\mathbf{0}\|_{\mathcal{F}}^2 \\
\Rightarrow\ & \lambda_j \|\mathbf{M}_j\|_{\mathcal{F}}^2 \leq \frac{1}{n_j} \sum_{(\mathbf{x}, \mathbf{x}', \Delta E_{00}) \in T_j} l(\mathbf{0}, (\mathbf{x}, \mathbf{x}', \Delta E_{00})) && (3.1) \\
\Rightarrow\ & \lambda_j \|\mathbf{M}_j\|_{\mathcal{F}}^2 \leq \Delta_{\max}^2 && (3.2) \\
\Rightarrow\ & \|\mathbf{M}_j\|_{\mathcal{F}} \leq \frac{\Delta_{\max}}{\sqrt{\lambda_j}}.
\end{aligned}$$

Inequality 3.1 comes from the fact that our loss is always positive and that $\|\mathbf{0}\|_{\mathcal{F}} = 0$. 3.2 is obtained by noting that $l(\mathbf{0}, (\mathbf{x}, \mathbf{x}', \Delta E_{00})) \leq \Delta_{\max}$.

Lemma C ((σ, m)-admissibility) *Let $(\mathbf{x}, \mathbf{x}', \Delta E_{00}(\mathbf{x}, \mathbf{x}'))$ and $(\mathbf{t}, \mathbf{t}', \Delta E_{00}(\mathbf{t}, \mathbf{t}'))$ be two examples and $\mathbf{M}_j$ be the optimal solution of Problem* (1). *The loss $l(\mathbf{M}_j, (\mathbf{x}, \mathbf{x}', \Delta E_{00}))$ is (σ, m)-admissible with $\sigma = 2\Delta_{\max}$ and $m = \frac{2D_j^2 \Delta_{\max}}{\sqrt{\lambda_j}}$.*

Proof.

$$
\begin{aligned}
&|l(\mathbf{M}_j, (\mathbf{x}, \mathbf{x}', \Delta E_{00}(\mathbf{x}, \mathbf{x}'))) - l(\mathbf{M}_j, (\mathbf{t}, \mathbf{t}', \Delta E_{00}(\mathbf{t}, \mathbf{t}')))| \\
&\quad = \left| \left| (\mathbf{x} - \mathbf{x}')^T \mathbf{M}_j (\mathbf{x} - \mathbf{x}') - \Delta E_{00}(\mathbf{x}, \mathbf{x}')^2 \right| - \left| (\mathbf{t} - \mathbf{t}')^T \mathbf{M}_j (\mathbf{t} - \mathbf{t}') - \Delta E_{00}(\mathbf{t}, \mathbf{t}')^2 \right| \right| \\
&\quad \leq \left| (\mathbf{x} - \mathbf{x}')^T \mathbf{M}_j (\mathbf{x} - \mathbf{x}') - (\mathbf{t} - \mathbf{t}')^T \mathbf{M}_j (\mathbf{t} - \mathbf{t}') \right| + \left| \Delta E_{00}(\mathbf{t}, \mathbf{t}')^2 - \Delta E_{00}(\mathbf{x}, \mathbf{x}')^2 \right| \qquad (4.1) \\
&\quad \leq \left| (\mathbf{x} - \mathbf{x}')^T \mathbf{M}_j (\mathbf{x} - \mathbf{x}') \right| + \left| (\mathbf{t} - \mathbf{t}')^T \mathbf{M}_j (\mathbf{t} - \mathbf{t}') \right| + \left| \Delta E_{00}(\mathbf{t}, \mathbf{t}')^2 - \Delta E_{00}(\mathbf{x}, \mathbf{x}')^2 \right| \qquad (4.2) \\
&\quad \leq 2 \max_{(\mathbf{x}, \mathbf{x}')} \left\{ \left| (\mathbf{x} - \mathbf{x}')^T \mathbf{M}_j (\mathbf{x} - \mathbf{x}') \right| \right\} + \left| \Delta E_{00}(\mathbf{t}, \mathbf{t}')^2 - \Delta E_{00}(\mathbf{x}, \mathbf{x}')^2 \right| \\
&\quad \leq \frac{2D^2 \Delta_{\max}}{\sqrt{\lambda_j}} + \left| \Delta E_{00}(\mathbf{t}, \mathbf{t}')^2 - \Delta E_{00}(\mathbf{x}, \mathbf{x}')^2 \right| \qquad (4.3) \\
&\quad \leq \frac{2D^2 \Delta_{\max}}{\sqrt{\lambda_j}} + |\Delta E_{00}(\mathbf{t}, \mathbf{t}') + \Delta E_{00}(\mathbf{x}, \mathbf{x}')| \, |\Delta E_{00}(\mathbf{t}, \mathbf{t}') - \Delta E_{00}(\mathbf{x}, \mathbf{x}')| \\
&\quad \leq \frac{2D^2 \Delta_{\max}}{\sqrt{\lambda_j}} + 2\Delta_{\max} |\Delta E_{00}(\mathbf{t}, \mathbf{t}') - \Delta E_{00}(\mathbf{x}, \mathbf{x}')| .
\end{aligned}
$$

Inequalities 4.1 and 4.2 are obtained by applying the triangle inequality respectively twice and once, 4.3 comes from the fact that $\|\mathbf{M}_j\|_{\mathcal{F}} \leq \frac{\Delta_{\max}}{\sqrt{\lambda_j}}$ (Lemma B) and that $\|\mathbf{x} - \mathbf{x}'\| \leq D_j^2$. Setting $\sigma = 2\Delta_{\max}$ and $m = \frac{2D_j^2 \Delta_{\max}}{\sqrt{\lambda_j}}$ gives the Lemma.

We will now prove the uniform stability of our algorithm but before to present this proof, we need the following Lemma.

Lemma D *Let $F_{T_j}()$ and $F_{T_j[i]}()$ be the functions to optimize, $\mathbf{M}_j$ and $\mathbf{M^i}_j$ their corresponding minimizers, and λ_j the regularization parameter used in our algorithm. Let $\Delta\mathbf{M}_j = \mathbf{M}_j - \mathbf{M^i}_j$, then, we have, for any $t \in [0, 1]$,*

$$\|\mathbf{M}_j\|_{\mathcal{F}}^2 - \|\mathbf{M}_j - t\Delta\mathbf{M}_j\|_{\mathcal{F}}^2 + \|\mathbf{M^i}_j\|_{\mathcal{F}}^2 - \|\mathbf{M^i}_j + t\Delta\mathbf{M}_j\|_{\mathcal{F}}^2 \leq \frac{2kt}{\lambda_j n_j} \|\Delta\mathbf{M}_j\|_{\mathcal{F}}. \qquad (5)$$

Proof. This proof is similar to the proof of Lemma 20 in [1] which we recall for the sake of completeness. $\hat{\varepsilon}_{T_j[i]}()$ is a convex function, thus, for any $t \in [0, 1]$, we can write:

$$\hat{\varepsilon}_{T_j[i]}(\mathbf{M}_j - t\Delta\mathbf{M}_j) - \hat{\varepsilon}_{T_j[i]}(\mathbf{M}_j) \leq t \left(\hat{\varepsilon}_{T_j[i]}(\mathbf{M^i}_j) - \hat{\varepsilon}_{T_j[i]}(\mathbf{M}_j) \right), \qquad (6)$$

$$\hat{\varepsilon}_{T_j[i]}(\mathbf{M^i}_j + t\Delta\mathbf{M}_j) - \hat{\varepsilon}_{T_j[i]}(\mathbf{M^i}_j) \leq t \left(\hat{\varepsilon}_{T_j[i]}(\mathbf{M}_j) - \hat{\varepsilon}_{T_j[i]}(\mathbf{M^i}_j) \right). \qquad (7)$$

By summing inequalities 6 and 7 we obtain

$$\hat{\varepsilon}_{T_j[i]}(\mathbf{M}_j - t\Delta\mathbf{M}_j) - \hat{\varepsilon}_{T_j[i]}(\mathbf{M}_j) + \hat{\varepsilon}_{T_j[i]}(\mathbf{M^i}_j + t\Delta\mathbf{M}_j) - \hat{\varepsilon}_{T_j[i]}(\mathbf{M^i}_j) \leq 0. \qquad (8)$$

Since $\mathbf{M}_j$ and $\mathbf{M^i}_j$ are minimizers of $F_{T_j}()$ and $F_{T_j[i]}()$, we can write:

$$F_{T_j}(\mathbf{M}_j) - F_{T_j}(\mathbf{M}_j - t\Delta\mathbf{M}_j) \leq 0, \qquad (9)$$

$$F_{T_j[i]}(\mathbf{M^i}_j) - F_{T_j[i]}(\mathbf{M^i}_j + t\Delta\mathbf{M}_j) \leq 0. \qquad (10)$$

By summing inequalities 9 and 10, we obtain

$$\hat{\varepsilon}_{T_j}(\mathbf{M}_j) - \hat{\varepsilon}_{T_j}(\mathbf{M}_j - t\Delta\mathbf{M}_j) + \lambda_j\|\mathbf{M}_j\|^2_{\mathcal{F}} - \lambda_j\|\mathbf{M}_j - t\Delta\mathbf{M}_j\|^2_{\mathcal{F}} + \\ \hat{\varepsilon}_{T_j[i]}(\mathbf{M^i}_j) - \hat{\varepsilon}_{T_j[i]}(\mathbf{M^i}_j + t\Delta\mathbf{M}_j) + \lambda_j\|\mathbf{M^i}_j\|^2_{\mathcal{F}} - \lambda_j\|\mathbf{M^i}_j + t\Delta\mathbf{M}_j\|^2_{\mathcal{F}} \leq 0. \quad (11)$$

We can now sum inequalities 8 and 11 to obtain

$$\hat{\varepsilon}_{T_j}(\mathbf{M}_j) - \hat{\varepsilon}_{T_j[i]}(\mathbf{M}_j) - \hat{\varepsilon}_{T_j}(\mathbf{M}_j - t\Delta\mathbf{M}_j) + \hat{\varepsilon}_{T_j[i]}(\mathbf{M}_j - t\Delta\mathbf{M}_j) + \\ \lambda_j\|\mathbf{M}_j\|^2_{\mathcal{F}} - \lambda_j\|\mathbf{M}_j - t\Delta\mathbf{M}_j\|^2_{\mathcal{F}} + \lambda_j\|\mathbf{M^i}_j\|^2_{\mathcal{F}} - \lambda_j\|\mathbf{M^i}_j + t\Delta\mathbf{M}_j\|^2_{\mathcal{F}} \leq 0. \quad (12)$$

From 12, we can write:

$$\lambda_j\|\mathbf{M}_j\|^2_{\mathcal{F}} - \lambda_j\|\mathbf{M}_j - t\Delta\mathbf{M}_j\|^2_{\mathcal{F}} + \lambda_j\|\mathbf{M^i}_j\|^2_{\mathcal{F}} - \lambda_j\|\mathbf{M^i}_j + t\Delta\mathbf{M}_j\|^2_{\mathcal{F}} \leq B \quad (13)$$

with

$$B = \hat{\varepsilon}_{T_j[i]}(\mathbf{M}_j) - \hat{\varepsilon}_{T_j}(\mathbf{M}_j) + \hat{\varepsilon}_{T_j}(\mathbf{M}_j - t\Delta\mathbf{M}_j) - \hat{\varepsilon}_{T_j[i]}(\mathbf{M}_j - t\Delta\mathbf{M}_j).$$

We are now looking for a bound on B:

$$\begin{aligned}
B &\leq \left|\hat{\varepsilon}_{T_j}(\mathbf{M}_j - t\Delta\mathbf{M}_j) - \hat{\varepsilon}_{T_j[i]}(\mathbf{M}_j - t\Delta\mathbf{M}_j) + \hat{\varepsilon}_{T_j[i]}(\mathbf{M}_j) - \hat{\varepsilon}_{T_j}(\mathbf{M}_j)\right| \\
&\leq \frac{1}{n_j}\left|\sum_{(\mathbf{x},\mathbf{x}',\Delta E_{00})\in T_j} l(\mathbf{M}_j - t\Delta\mathbf{M}_j,(\mathbf{x},\mathbf{x}',\Delta E_{00})) - \sum_{(\mathbf{t},\mathbf{t}',\Delta E_{00})\in T_j[i]} l(\mathbf{M}_j - t\Delta\mathbf{M}_j,(\mathbf{t},\mathbf{t}',\Delta E_{00})) + \right. \\
&\qquad \left.\sum_{(\mathbf{t},\mathbf{t}',\Delta E_{00})\in T_j[i]} l(\mathbf{M}_j,(\mathbf{t},\mathbf{t}',\Delta E_{00})) - \sum_{(\mathbf{x},\mathbf{x}',\Delta E_{00})\in T_j} l(\mathbf{M}_j,(\mathbf{x},\mathbf{x}',\Delta E_{00}))\right| \\
&= \frac{1}{n_j}\left|l(\mathbf{M}_j - t\Delta\mathbf{M}_j,(\mathbf{x_i},\mathbf{x_i}',\Delta E_{00})) - l(\mathbf{M}_j - t\Delta\mathbf{M}_j,(\mathbf{t_i},\mathbf{t_i}',\Delta E_{00})) + \right. \\
&\qquad \left. l(\mathbf{M}_j,(\mathbf{t_i},\mathbf{t_i}',\Delta E_{00})) - l(\mathbf{M}_j,(\mathbf{x_i},\mathbf{x_i}',\Delta E_{00}))\right| && (14.1) \\
&\leq \frac{1}{n_j}\left(\left|l(\mathbf{M}_j - t\Delta\mathbf{M}_j,(\mathbf{x_i},\mathbf{x_i}',\Delta E_{00})) - l(\mathbf{M}_j,(\mathbf{x_i},\mathbf{x_i}',\Delta E_{00}))\right| + \right. \\
&\qquad \left.\left|l(\mathbf{M}_j,(\mathbf{t_i},\mathbf{t_i}',\Delta E_{00})) - l(\mathbf{M}_j - t\Delta\mathbf{M}_j,(\mathbf{t_i},\mathbf{t_i}',\Delta E_{00}))\right|\right) && (14.2) \\
&\leq \frac{1}{n_j}\left(k\|\mathbf{M}_j - t\Delta\mathbf{M}_j - \mathbf{M}_j\|_{\mathcal{F}} + k\|\mathbf{M}_j - \mathbf{M}_j + t\Delta\mathbf{M}_j\|_{\mathcal{F}}\right) && (14.3) \\
&\leq \frac{2kt}{n_j}\|\Delta\mathbf{M}_j\|_{\mathcal{F}}.
\end{aligned}$$

Equality 14.1 comes from the fact that T_j and $T_j[i]$ only differ by their i^{th} example, inequality 14.2 is due to the triangle inequality and 14.3 is obtained thanks to the k-lipschitz property of our loss (Lemma A).

Then combining the bound on B with equation 13 and dividing both sides by λ_j gives the Lemma.

We can now show the uniform stability property of the approach.

Lemma E (Uniform stability) *Given a training sample T_j of n_j examples drawn i.i.d. from $P(C_j)$, our algorithm has a uniform stability in $\frac{\mathcal{K}}{n_j}$ with $\mathcal{K} = \frac{2D_j^4}{\lambda_j}$.*

Proof. By setting $t = \frac{1}{2}$ in Lemma D, one can obtain for the left hand side:

$$\|\mathbf{M}_j\|^2_{\mathcal{F}} - \|\mathbf{M}_j - \frac{1}{2}\Delta\mathbf{M}_j\|^2_{\mathcal{F}} + \|\mathbf{M^i}_j\|^2_{\mathcal{F}} - \|\mathbf{M^i}_j + \frac{1}{2}\Delta\mathbf{M}_j\|^2_{\mathcal{F}} = \frac{1}{2}\|\Delta\mathbf{M}_j\|^2_{\mathcal{F}}$$

and thus:

$$\frac{1}{2}\|\Delta\mathbf{M}_j\|_{\mathcal{F}}^2 \leq \frac{2k\frac{1}{2}}{\lambda_j n_j}\|\Delta\mathbf{M}_j\|_{\mathcal{F}},$$

which implies

$$\|\Delta\mathbf{M}_j\|_{\mathcal{F}} \leq \frac{2k}{\lambda_j n_j}.$$

Since our loss is k-lipschitz (Lemma A) we have:

$$\begin{aligned}|l(\mathbf{M}_j, (\mathbf{x}, \mathbf{x}', \Delta E_{00})) - l(\mathbf{M'}_j, (\mathbf{x}, \mathbf{x}', \Delta E_{00}))| &\leq k\|\Delta\mathbf{M}_j\|_{\mathcal{F}} \\ &\leq \frac{2k^2}{\lambda_j n_j}.\end{aligned}$$

In particular,

$$\sup_{(\mathbf{x}, \mathbf{x}', \Delta E_{00})} |l(\mathbf{M}_j, (\mathbf{x}, \mathbf{x}', \Delta E_{00})) - l(\mathbf{M'}_j, (\mathbf{x}, \mathbf{x}', \Delta E_{00}))| \leq \frac{2k^2}{\lambda_j n_j}.$$

By recalling that $k = D_j^2$ (Lemma A) and setting $\mathcal{K} = \frac{2k^2}{\lambda_j}$, we get the lemma.

We now recall the McDiarmid inequality [2], used to prove our main theorem.

Theorem A (McDiarmid inequality) *Let $X_1, ..., X_n$ be n independent random variables taking values in X and let $Z = f(X_1, ..., X_n)$. If for each $1 \leq i \leq n$, there exists a constant c_i such that*

$$\sup_{x_1,...,x_n,x_i' \in \mathcal{X}} |f(x_1, ..., x_n) - f(x_1, ..., x_i', ..., x_n)| \leq c_i, \forall 1 \leq i \leq n,$$

$$\textit{then for any } \epsilon > 0, Pr\left[|Z - \mathbb{E}\left[Z\right]| \geq \epsilon\right] \leq 2\exp\left(\frac{-2\epsilon^2}{\sum_{i=1}^n c_i^2}\right).$$

Using Lemma E about the stability of our algorithm and the McDiarmid inequality we can derive our generalization bound. For this purpose, we replace Z by $D_{T_j} = \varepsilon_{P(C_j)}(\mathbf{M}_j) - \hat{\varepsilon}_{T_j}(\mathbf{M}_j)$ in Theorem A and we need to bound $\mathbb{E}_{T_j}\left[D_{T_j}\right]$ and $\left|D_{T_j} - D_{T_j[i]}\right|$, which is done in the following two lemmas.

Lemma F *For any learning method of estimation error D_{T_j} and satisfying a uniform stability in $\frac{\mathcal{K}}{n_j}$, we have*

$$\mathbb{E}_{T_j}\left[D_{T_j}\right] \leq \frac{\mathcal{K}}{n_j}.$$

Proof.

$$\begin{aligned}
\mathbb{E}_{T_j}\left[D_{T_j}\right] &\leq \mathbb{E}_{T_j}\left[\mathbb{E}_{(\mathbf{x},\mathbf{x}',\Delta E_{00})}\left[l(\mathbf{M}_j, (\mathbf{x}, \mathbf{x}', \Delta E_{00}))\right] - \hat{\varepsilon}_{T_j}(\mathbf{M}_j)\right] \\
&\leq \mathbb{E}_{T_j,(\mathbf{x},\mathbf{x}',\Delta E_{00})}\left[\left|l(\mathbf{M}_j, (\mathbf{x}, \mathbf{x}', \Delta E_{00})) - \frac{1}{n_j}\sum_{(\mathbf{x_k},\mathbf{x_k'},\Delta E_{00})\in T_j} l(\mathbf{M}_j, (\mathbf{x_k}, \mathbf{x_k'}, \Delta E_{00}))\right|\right] \\
&\leq \mathbb{E}_{T_j,(\mathbf{x},\mathbf{x}',\Delta E_{00})}\left[\left|\frac{1}{n_j}\sum_{(\mathbf{x_k},\mathbf{x_k'},\Delta E_{00})\in T_j} (l(\mathbf{M}_j, (\mathbf{x}, \mathbf{x}', \Delta E_{00})) - l(\mathbf{M}_j, (\mathbf{x_k}, \mathbf{x_k'}, \Delta E_{00})))\right|\right] \\
&\leq \mathbb{E}_{T_j,(\mathbf{x},\mathbf{x}',\Delta E_{00})}\left[\left|\frac{1}{n_j}\sum_{(\mathbf{x_k},\mathbf{x_k'},\Delta E_{00})\in T_j} (l(\mathbf{M}_j[k], (\mathbf{x_k}, \mathbf{x_k'}, \Delta E_{00})) - l(\mathbf{M}_j, (\mathbf{x_k}, \mathbf{x_k'}, \Delta E_{00})))\right|\right] \quad (15.1) \\
&\leq \frac{\mathcal{K}}{n_j}. \quad (15.2)
\end{aligned}$$

Inequality 15.1 comes from the fact that T_j and $(\mathbf{x}, \mathbf{x}', \Delta E_{00})$ are drawn i.i.d. from the distribution $P(C_j)$ and thus we do not change the expected value by replacing one example with another, 15.2 is obtained by applying triangle inequality followed by the property of uniform stability (Lemma E).

Lemma G *For any matrix $\mathbf{M}_j$ learned by our algorithm using n_j training examples, and any loss function l satisfying the (σ, m)-admissibility, we have*

$$\left| D_{T_j} - D_{T_j[k]} \right| \leq \frac{2\mathcal{K} + (\Delta_{\max}\sigma + m)}{n_j}.$$

Proof.

$$\begin{aligned}
\left| D_{T_j} - D_{T_j[i]} \right| &= \left| \varepsilon_{P(C_j)}(\mathbf{M}_j) - \hat{\varepsilon}_{T_j}(\mathbf{M}_j) - \left(\varepsilon_{P(C_j)}(\mathbf{M^i}_j) - \hat{\varepsilon}_{T_j[i]}(\mathbf{M^i}_j) \right) \right| \\
&= \left| \varepsilon_{P(C_j)}(\mathbf{M}_j) - \hat{\varepsilon}_{T_j}(\mathbf{M}_j) - \varepsilon_{P(C_j)}(\mathbf{M^i}_j) + \hat{\varepsilon}_{T_j[i]}(\mathbf{M^i}_j) - \hat{\varepsilon}_{T_j}(\mathbf{M^i}_j) + \hat{\varepsilon}_{T_j}(\mathbf{M^i}_j) \right| \\
&\leq \left| \varepsilon_{P(C_j)}(\mathbf{M}_j) - \varepsilon_{P(C_j)}(\mathbf{M^i}_j) \right| + \left| \hat{\varepsilon}_{T_j}(\mathbf{M^i}_j) - \hat{\varepsilon}_{T_j}(\mathbf{M}_j) \right| + \left| \hat{\varepsilon}_{T_j[i]}(\mathbf{M^i}_j) - \hat{\varepsilon}_{T_j}(\mathbf{M^i}_j) \right| \quad (16.1) \\
&\leq \mathbb{E}_{(\mathbf{x},\mathbf{x}',\Delta E_{00})} \left[\left| l(\mathbf{M}_j, (\mathbf{x}, \mathbf{x}', \Delta E_{00})) - l(\mathbf{M^i}_j, (\mathbf{x}, \mathbf{x}', \Delta E_{00})) \right| \right] + \\
&\qquad \left| \hat{\varepsilon}_{T_j}(\mathbf{M^i}_j) - \hat{\varepsilon}_{T_j}(\mathbf{M}_j) \right| + \left| \hat{\varepsilon}_{T_j[i]}(\mathbf{M^i}_j) - \hat{\varepsilon}_{T_j}(\mathbf{M^i}_j) \right| \quad (16.2) \\
&\leq \frac{\mathcal{K}}{n_j} + \left| \hat{\varepsilon}_{T_j}(\mathbf{M^i}_j) - \hat{\varepsilon}_{T_j}(\mathbf{M}_j) \right| + \left| \hat{\varepsilon}_{T_j[i]}(\mathbf{M^i}_j) - \hat{\varepsilon}_{T_j}(\mathbf{M^i}_j) \right| \quad (16.3) \\
&\leq \frac{\mathcal{K}}{n_j} + \frac{1}{n_j} \sum_{(\mathbf{x},\mathbf{x}',\Delta E_{00}) \in T_j} \left| l(\mathbf{M^i}_j, (\mathbf{x}, \mathbf{x}', \Delta E_{00})) - l(\mathbf{M}_j, (\mathbf{x}, \mathbf{x}', \Delta E_{00})) \right| + \\
&\qquad \left| \hat{\varepsilon}_{T_j[i]}(\mathbf{M^i}_j) - \hat{\varepsilon}_{T_j}(\mathbf{M^i}_j) \right| \\
&\leq \frac{\mathcal{K}}{n_j} + \frac{\mathcal{K}}{n_j} + \left| \hat{\varepsilon}_{T_j[i]}(\mathbf{M^i}_j) - \hat{\varepsilon}_{T_j}(\mathbf{M^i}_j) \right| \quad (16.4) \\
&= \frac{2\mathcal{K}}{n_j} + \frac{1}{n_j} \left| l(\mathbf{M^i}_j, (\mathbf{t}, \mathbf{t}', \Delta E_{00})) - l(\mathbf{M^i}_j, (\mathbf{x}, \mathbf{x}', \Delta E_{00})) \right| \quad (16.5) \\
&\leq \frac{2\mathcal{K}}{n_j} + \frac{1}{n_j} \left(\sigma \left| \Delta E_{00}(\mathbf{t}, \mathbf{t}') - \Delta E_{00}(\mathbf{x}, \mathbf{x}') \right| + m \right) \quad (16.6) \\
&\leq \frac{2\mathcal{K} + (\Delta_{\max}\sigma + m)}{n_j}. \quad (16.7)
\end{aligned}$$

Inequalities 16.1 and 16.2 are due to the triangle inequality. 16.3 and 16.4 come from the uniform stability (Lemma E). 16.5 comes from the fact that T_j and $T_j[i]$ only differ by their i^{th} example. 16.6 comes from the (σ, m)-admissibility of our loss (Lemma C). Eventually, 16.7 is obtained by noting that $|\Delta E_{00}(\mathbf{t}, \mathbf{t}') - \Delta E_{00}(\mathbf{x}, \mathbf{x}')| \leq \Delta_{\max}$.

Theorem 1 (Generalization bound) *With probability $1 - \delta$, for any matrix $\mathbf{M}_j$ related to a region C_j, $0 \leq j \leq K$, learned with Algorithm 1, we have:*

$$\varepsilon_{P(C_j)}(\mathbf{M}_j) \leq \hat{\varepsilon}_{T_j}(\mathbf{M}_j) + \frac{2D_j^4}{\lambda_j n_j} + \left(\frac{2D_j^4}{\lambda_j} + \Delta_{\max}\left(\frac{2D_j^2}{\sqrt{\lambda_j}} + 2\Delta_{\max} \right) \right) \sqrt{\frac{\ln(\frac{2}{\delta})}{2n_j}}.$$

Proof. Using the McDiarmid inequality (Theorem A) and Lemma G we can write:

$$\begin{aligned}
\Pr\left[\left| D_{T_j} - \mathbb{E}_{T_j}\left[D_{T_j} \right] \right| \geq \epsilon \right] &\leq 2\exp\left(-\frac{2\epsilon^2}{\sum_{j=1}^{n} \left(\frac{2\mathcal{K} + (5\sigma + m)}{n_j} \right)^2} \right) \\
&\leq 2\exp\left(-\frac{2\epsilon^2}{\frac{1}{n_j}\left(2\mathcal{K} + (5\sigma + m) \right)^2} \right).
\end{aligned}$$

Then, by setting:

$$\delta = 2\exp\left(-\frac{2\epsilon^2}{\frac{1}{n_j}\left(2\mathcal{K} + (5\sigma + m)\right)^2}\right)$$

we obtain:

$$\epsilon = (2\mathcal{K} + (\Delta_{\max}\sigma + m))\sqrt{\frac{\ln\left(\frac{2}{\delta}\right)}{2n_j}}$$

and thus:

$$\Pr\left[\left|D_{T_j} - \mathbb{E}_{T_j}\left[D_{T_j}\right]\right| < \epsilon\right] > 1 - \delta.$$

Then, with probability $1 - \delta$:

$$\begin{aligned} & D_{T_j} < \mathbb{E}_{T_j}\left[D_{T_j}\right] + \epsilon \\ \Leftrightarrow \quad & \varepsilon_{P(C_j)}(\mathbf{M}_j) - \hat{\varepsilon}_{T_j}(\mathbf{M}_j) < \mathbb{E}_{T_j}\left[D_{T_j}\right] + \epsilon \\ \Leftrightarrow \quad & \varepsilon_{P(C_j)}(\mathbf{M}_j) < \hat{\varepsilon}_{T_j}(\mathbf{M}_j) + \frac{\mathcal{K}}{n_j} + (2\mathcal{K} + (\Delta_{\max}\sigma + m))\sqrt{\frac{\ln\left(\frac{2}{\delta}\right)}{2n_j}} \quad \text{by Lemma F.} \end{aligned}$$

Replacing $\mathcal{K}$, σ and m by their respective values in the last equation gives the lemma.

2.2 Generalization bound for Algorithm 1

We consider the partition $C_0, C_1, \ldots, C_K$ over pairs of examples considered by Algorithm 1. We first recall the concentration inequality that will help us to derive the bound.

Proposition 1 ([3]). *Let $(n_0, n_1, \ldots, n_K)$ an IID multinomial random variable with parameters $n = \sum_{j=0}^{K}$ and $(p(C_0), p(C_1), \ldots, p(C_K))$. By the Breteganolle-Huber-Carol inequality we have: $Pr\left\{\sum_{j=0}^{K}\left|\frac{n_j}{n} - p(C_j)\right| \geq \eta\right\} \leq 2^K \exp\left(\frac{-n\eta^2}{2}\right)$, hence with probability at least $1 - \delta$,*

$$\sum_{j=0}^{K}\left|\frac{n_j}{n} - p(C_j)\right| \leq \sqrt{\frac{2K\ln 2 + 2\ln(1/\delta)}{n}}. \tag{17}$$

We are now ready to prove the main theorem of the paper.

Theorem 2 *Let $C_0, C_1, \ldots, C_K$ be the regions considered, then for any set of metrics $\mathbf{M} = \{\mathbf{M}_0, \ldots, \mathbf{M}_K\}$ learned by Algorithm 1 from a data sample T of n pairs with our approach, we have with probability at least $1 - \delta$ that*

$$[l(\mathbf{M}, (\mathbf{x}, \mathbf{x}', \Delta E_{00}))] \leq \hat{\varepsilon}_{T_j}(\mathbf{M}) + L_B\sqrt{\frac{2(K+1)\ln 2 + 2\ln 2/\delta}{n}} + \Bigg(\frac{2(KD^4+1)}{\lambda n} + \frac{2(KD^4+1)}{\lambda} + \Delta_{\max}(\frac{2(KD^2+1)}{\sqrt{\lambda}} + 2(K$$

, where $D = \max 1 \leq j \leq KD_j$, $L_B = \max\{\frac{\Delta_{max}}{\sqrt{\lambda}}, \Delta_{\max}^2\}$ is the bound on the loss function and $\lambda = \min_{0\leq j\leq K}\lambda_j$ the minimum regularization parameter among the $K+1$ learning problems used in Algorithm 1.

Proof. Let n_j be the number points of T that fall into the partition C_j. $(n_0, n_1, \ldots, n_K)$ is a IID multinomial random variable with parameters n and $(p(C_0), p(C_1), \ldots, p(C_K))$.

$$
\begin{aligned}
&\left|\mathbb{E}_{(\mathbf{x},\mathbf{x}',\Delta E_{00})\sim P}\left[l(\mathbf{M},(\mathbf{x},\mathbf{x}',\Delta E_{00}))\right] - \hat{\varepsilon}_T(\mathbf{M})\right| \\
&= \left|\sum_{j=0}^{K} \mathbb{E}_{(\mathbf{x},\mathbf{x}',\Delta E_{00})\sim P|(\mathbf{x},\mathbf{x}',\Delta E_{00})\in C_j}\left[l(\mathbf{M},(\mathbf{x},\mathbf{x}',\Delta E_{00}))\right] p(C_j) - \hat{\varepsilon}_T(\mathbf{M})\right| \\
&= \Bigg|\sum_{j=0}^{K} \mathbb{E}_{(\mathbf{x},\mathbf{x}',\Delta E_{00})\sim P|(\mathbf{x},\mathbf{x}',\Delta E_{00})\in C_j}\left[l(\mathbf{M},(\mathbf{x},\mathbf{x}',\Delta E_{00}))\right] p(C_j) \\
&\quad - \sum_{j=0}^{K} \mathbb{E}_{(\mathbf{x},\mathbf{x}',\Delta E_{00})\sim P|(\mathbf{x},\mathbf{x}',\Delta E_{00})\in C_j}\left[l(\mathbf{M},(\mathbf{x},\mathbf{x}',\Delta E_{00}))\right] \frac{n_j}{n} \\
&\quad + \sum_{j=0}^{K} \mathbb{E}_{(\mathbf{x},\mathbf{x}',\Delta E_{00})\sim P|(\mathbf{x},\mathbf{x}',\Delta E_{00})\in C_j}\left[l(\mathbf{M},(\mathbf{x},\mathbf{x}',\Delta E_{00}))\right] \frac{n_j}{n} - \hat{\varepsilon}_T(\mathbf{M})\Bigg| \\
&\leq \Bigg|\sum_{j=0}^{K} \mathbb{E}_{(\mathbf{x},\mathbf{x}',\Delta E_{00})\sim P|(\mathbf{x},\mathbf{x}',\Delta E_{00})\in C_j}\left[l(\mathbf{M},(\mathbf{x},\mathbf{x}',\Delta E_{00}))\right] p(C_j) \\
&\quad - \sum_{j=0}^{K} \mathbb{E}_{(\mathbf{x},\mathbf{x}',\Delta E_{00})\sim P|(\mathbf{x},\mathbf{x}',\Delta E_{00})\in C_j}\left[l(\mathbf{M},(\mathbf{x},\mathbf{x}',\Delta E_{00}))\right] \frac{n_j}{n}\Bigg| \\
&\quad + \left|\sum_{j=0}^{K} \mathbb{E}_{(\mathbf{x},\mathbf{x}',\Delta E_{00})\sim P|(\mathbf{x},\mathbf{x}',\Delta E_{00})\in C_j}\left[l(\mathbf{M},(\mathbf{x},\mathbf{x}',\Delta E_{00}))\right] \frac{n_j}{n} - \hat{\varepsilon}_T(\mathbf{M})\right| \\
&\leq \sum_{j=0}^{K} \mathbb{E}_{(\mathbf{x},\mathbf{x}',\Delta E_{00})\sim P|(\mathbf{x},\mathbf{x}',\Delta E_{00})\in C_j}\left|\left[l(\mathbf{M},(\mathbf{x},\mathbf{x}',\Delta E_{00}))\right]\right| \left|p(C_j) - \frac{n_j}{n}\right| && (18.1) \\
&\quad + \left|\sum_{j=0}^{K} \mathbb{E}_{(\mathbf{x},\mathbf{x}',\Delta E_{00})\sim P|(\mathbf{x},\mathbf{x}',\Delta E_{00})\in C_j}\left[l(\mathbf{M},(\mathbf{x},\mathbf{x}',\Delta E_{00}))\right] \frac{n_j}{n} - \sum_{j=0}^{K}\frac{n_j}{n}\hat{\varepsilon}_{T_j}(\mathbf{M})\right. \\
&\leq \sum_{j=0}^{K} L_B \left|p(C_j) - \frac{n_j}{n}\right| && (18.2) \\
&\quad + \left|\sum_{j=0}^{K} \frac{n_j}{n}\left(\mathbb{E}_{(\mathbf{x},\mathbf{x}',\Delta E_{00})\sim P|(\mathbf{x},\mathbf{x}',\Delta E_{00})\in C_j}\left[l(\mathbf{M},(\mathbf{x},\mathbf{x}',\Delta E_{00}))\right] - \hat{\varepsilon}_{T_j}(\mathbf{M})\right)\right| \\
&\leq L_B\sqrt{\frac{2(K+1)\ln 2 + 2\ln 2/\delta}{n}} + \sum_{j=0}^{K}\frac{n_j}{n}\left|\mathbb{E}_{(\mathbf{x},\mathbf{x}',\Delta E_{00})\sim P|(\mathbf{x},\mathbf{x}',\Delta E_{00})\in C_j}\left[l(\mathbf{M},(\mathbf{x},\mathbf{x}',\Delta E_{00}))\right] - \hat{\varepsilon}_{T_j}(\mathbf{M})\right| \\
&\leq L_B\sqrt{\frac{2(K+1)\ln 2 + 2\ln 2/\delta}{n}} + \\
&\quad \sum_{j=0}^{K}\frac{n_j}{n}\left(\frac{2D_j^4}{\lambda_j n_j} + \left(\frac{2D_j^4}{\lambda_j} + \Delta_{\max}\left(\frac{2D_j^2}{\sqrt{\lambda_j}} + 2\Delta_{\max}\right)\right)\sqrt{\frac{\ln(\frac{4(K+1)}{\delta})}{2n_j}}\right) && (18.3) \\
&\leq L_B\sqrt{\frac{2(K+1)\ln 2 + 2\ln 2/\delta}{n}} + \left(\frac{2(KD^4+1)}{\lambda n} + \frac{2(KD^4+1)}{\lambda}\right. \\
&\quad \left. + \Delta_{\max}\left(\frac{2(KD^2+1)}{\sqrt{\lambda}} + 2\Delta_{\max}\right)\right)\sqrt{\frac{\ln(\frac{4(K+1)}{\delta})}{2n}} && (18.4)
\end{aligned}
$$

Line (18.1) uses the fact L_B bounds the loss function. 18.2 is obtained by applying proposition 1 with probability $1-\delta/2$. **??** by applying Lemma (1) for each of the $(K+1)$ learning problems T_j, $0 \leq j \leq n$ with probability $1-\delta/(2(K+1))$. The last line (18.4) is obtained by cancelling out the n_j and noting that $\sqrt{n_j} \leq \sqrt{n}$ and taking $D = max_{1\leq i\leq n} D_j$, note that $D_0 = 1$ and corresponding to the partition used by the global metric. The final result is obtained by the union bound.

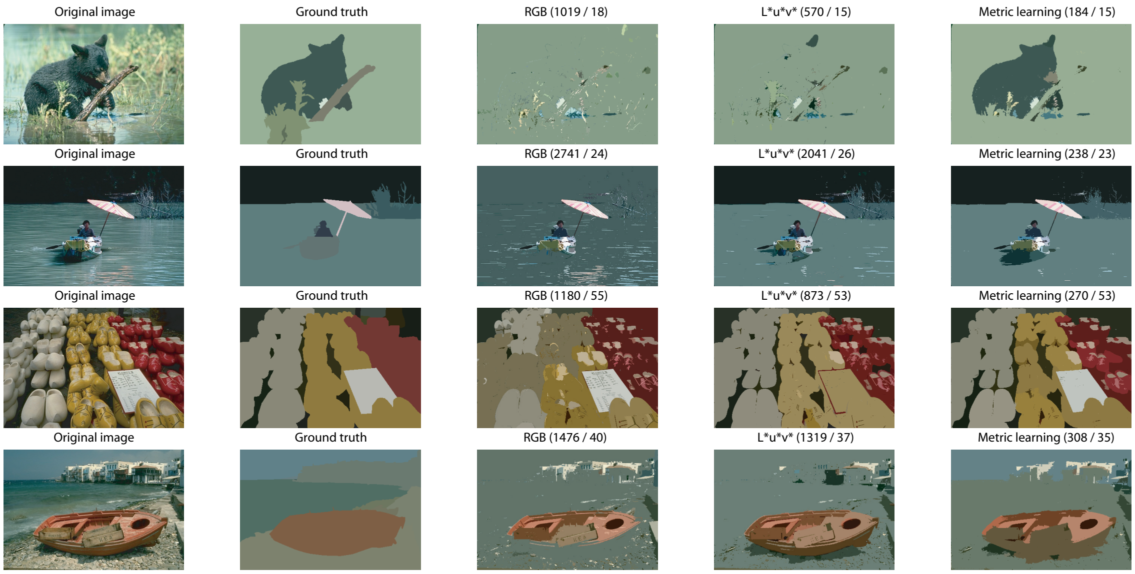
3 Image Segmentation

In this section, we illustrate the application of the color mean-shift algorithm presented in our paper. We apply color mean-shift on RGB components, on $L^*u^*v^*$ components and by using our learned distance directly in the RGB components. The overall quantitative results for the Berkeley dataset are provided in the paper and we propose to show some qualitative results on this dataset in Figure 1. As explained in the paper, the number of segments in the resulting images is not a parameter of the algorithm, as a consequence it is not easy to obtain images with the same number of segments for the three algorithms (RGB, $L^*u^*v^*$ and Metric learning). Thus, given an image, by playing with the color distance threshold, we have tried to obtain the same segment numbers as the corresponding ground truth for the three algorithms. However, the color mean-shift algorithm provides some very small segments, specially for the RGB and $L^*u^*v^*$ color spaces. Consequently, for each test, in Figure 1, we have mentioned between brackets, first, the number of segments, and second, the number of segments whose size is more than 100 pixels. For fair comparison, we use this last number as reference for each image, i.e. this number is almost constant and close to the ground truth for each row.

It is worth mentioning that the ground truth segmentation has always very few segments. Thus, starting from a large number of small segments, the used algorithm is grouping them by considering their color differences. Consequently, the used color distance is crucial when we want to obtain small number of segments as provided by the ground truth. We can see in Figure 1, that when working in the RGB or $L^*u^*v^*$ color spaces, some segments that are perceptually different are merged while some other similar segments are not. Most of the time, the color mean-shift is working well when using our distance. This point was already checked quantitatively on the whole Berkeley dataset in the paper.

References

1. Olivier Bousquet and Andr Elisseeff. Stability and generalization. *JMLR*, 2:499–526, 2002.
2. C. McDiarmid. *Surveys in Combinatorics*, chapter On the method of bounded differences, pages 148–188. Cambridge University Press, 1989.
3. Aad W. van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes*. Springer, 2000.


Original image
Ground truth
RGB (1019 / 18)
L*u*v* (570 / 15)
Metric learning (184 / 15)
Original image
Ground truth
RGB (2741 / 24)
L*u*v* (2041 / 26)
Metric learning (238 / 23)
Original image
Ground truth
RGB (1180 / 55)
L*u*v* (873 / 53)
Metric learning (270 / 53)
Original image
Ground truth
RGB (1476 / 40)
L*u*v* (1319 / 37)
Metric learning (308 / 35)