How to scale hyperparameters for quickshift image segmentation

Damien Garreau

Université Côte d'Azur, Inria, CNRS, LJAD, France

Abstract

Quickshift is a popular algorithm for image segmentation, used as a preprocessing step in many applications. Unfortunately, it is quite challenging to understand the hyperparameters' influence on the number and shape of superpixels produced by the method. In this paper, we study theoretically a slightly modified version of the quickshift algorithm, with a particular emphasis on homogeneous image patches with i.i.d. pixel noise and sharp boundaries between such patches. Leveraging this analysis, we derive a simple heuristic to scale quickshift hyperparameters with respect to the image size, which we check empirically.

1 INTRODUCTION

Quickshift is a clustering algorithm, which is used in image processing to obtain superpixels. It proceeds by first computing a kernel density estimate at scale k_s, then connecting each pixel to the nearest neighbor with higher density. All connections further away than d_m are removed, and the superpixels are then defined as the connected components of the resulting graph. Proposed by Vedaldi and Soatto (2008) as an efficient way to approximate the celebrated mean shift algorithm (Cheng, 1995; Comaniciu and Meer, 2002), the algorithm also appears in Rodriguez and Laio (2014). Together with SLIC (Achanta et al., 2012), compact watershed (Neubert and Protzel, 2014), and Felzenszwalb (Felzenszwalb and Huttenlocher, 2004), it is often used as a preprocessing step in more complex computer vision tasks such as image compression (Ryu et al., 2014) or semantic segmentation (Zhang et al., 2020). In the field of interpretability, quickshift is used as a default step when using LIME for images (Ribeiro et al., 2016), in order to create interpretable features.
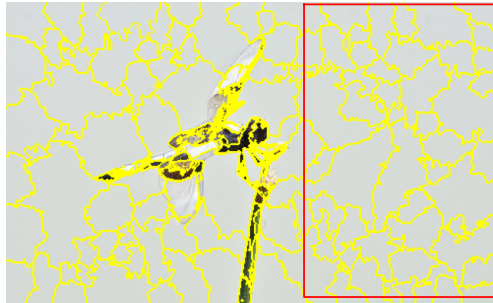

Figure 1: Segmentation of an image with quickshift with default hyperparameters ($k_s = 5$, $d_m = 10$). The superpixels are marked in yellow thick lines. The area inside the red rectangle is split up in 56 superpixels despite being quite homogeneous.

In these applications, the number of superpixels produced by quickshift can be quite critical. For instance, in the case of LIME, the number of superpixels corresponds to the dimension of the linear model trained later on by the method (Garreau and Mardaoui, 2021). Unfortunately, there is no way to know in advance the number of superpixels produced by quickshift as a function of the two main hyperparamters: k_s and d_m, in contrast for instance with SLIC. This can be particularly troublesome, in particular if the image at hand contains large, flat portions, where it is known that quickshift oversegments (see Figure 1). In contrast, boundaries between patches of different colors are seemingly always well-identified.

1.1 Summary of the paper

In this paper, we provide quantitative insights on the relationship between k_s and d_m, the two main hyperparameters of the method, and the number of superpixels produced when running the algorithm with this choice of hyperparameters. Our main theoretical contribution is the study of a modified version of quickshift on flat regions of the image. In that case, we show that this algorithm will find in average $\mathcal{A}/B(k_s, d_m)$ local maxima, where $B(\cdot, \cdot)$ is 2-homogeneous and $\mathcal{A}$ is the

area of the flat region (Theorem 3.2). Empirically, we demonstrate that these findings extend to the number of superpixels found by quickshift in real images. This yields the following heuristic: **for a given image, multiplying k_s or d_m by κ will roughly divide the number of superpixels by κ^2.** We also show how sharp boundaries between homogeneous patches are detected by quickshift (Theorem 4.1).

The paper is organized as follows: in Section 2, we recall the detailed operation of quickshift. Section 3 contains the analysis for the flat portions, Section 4 for boundaries. Only sketches of the proofs are provided, the complete version of the proofs can be found in the Appendix. Finally, we show some experimental results in Section 5. All experiments are realized with the `scikit-image` implementation of quickshift and are publicly available.[1]

1.2 Related work

Density-based clustering algorithms proceed by first computing a density estimate (Parzen, 1962), then either by looking at the level sets of the density or by hill-climbing. Quickshift belongs to the second category, and can be seen as a sample-based version of mean shift. This last method is analyzed in Arias-Castro et al. (2016), which shows that the updates converge to the correct gradient steps. Regarding quickshift, Jiang (2017) proves its consistency in the statistical sense, and gives the asymptotic size of k_s and d_m to recover the underlying density when the number of points grows to infinity. Since the result is asymptotic, it is difficult to use to pick k_s and d_m for a given image. Moreover, one of the main assumptions forbids the existence of flat regions in the theoretical analysis. A related analysis is proposed by Verdinelli and Wasserman (2018), with the same caveat. Also noting that density-based clustering algorithms tend to oversegment flat regions, Jiang et al. (2018) proposes a way to solve this problem, but this comes at the cost of computing cluster cores (Jiang and Kpotufe, 2017) of the density estimate.

2 QUICKSHIFT: A REFRESHER

We now describe quickshift in more details, introducing notation along the way. In this description, we follow the `scikit-image` implementation (Van der Walt et al., 2014), which seems to be the most popular at the moment. In all the paper, we focus on a rectangular image ξ, of size (H, W). We denote the pixel positions by $(i, j) \in I := [H] \times [W]$, where $[k] := \{1, \ldots, k\}$. The pixels values are denoted by $\xi_{i,j} \in \mathbb{R}^3$. Strictly speaking, we work directly in the CIELAB space, though quick-

shift usually takes as input RGB images and converts them to the CIELAB space.

2.1 Density estimation

Quickshift relies on a Gaussian kernel density estimate which we denote by $P_{i,j}$. The main idea is to see pixels as points in $\mathbb{R}^5$ (two space coordinates and three color coordinates). In practice, for each position pixel (i, j), not all pixels of the image are considered to build this estimate, and only pixels (u, v) close to (i, j) are taken into account. More precisely, only (u, v) inside

$$C_{i,j} := \{(u,v) \in I, \text{ s.t. } |i-u| \vee |j-v| \leqslant k_w\}, \quad (1)$$

where k_w is a positive scale hyperparameter (called *kernel width* in the `scikit-image` implementation) are considered. Now we are able to define the density estimates, computed according to

$$P_{i,j} := \sum_{(u,v)\in C_{i,j}} e^{\frac{-(i-u)^2-(j-v)^2-\|\xi_{i,j}-\xi_{u,v}\|^2}{2k_s^2}}, \quad (2)$$

where k_s is a positive scale hyperparameter (called *kernel size* and equal to 5 by default in the `scikit-image` implementation). By default, $k_w = 3k_s$, an assumption that we will always make from now on, though it is straightforward to adapt our analysis for another fixed k_w. Finally, some i.i.d. $\mathcal{N}\left(0, \sigma_0^2\right)$ noise is added on each $P_{i,j}$ to break eventual ties (with $\sigma_0 = 10^{-5}$). This procedure is described in Algorithm 1. The `scikit-learn` also has a *ratio* hyperparameter, allowing to adjust the importance of the pixel values with respect to the pixel positions. We do not consider this hyperparameter in our analysis, since it simply amounts to multiplying the pixel values by a positive constant, which does not change anything for flat regions.

Given $(i, j) \in I$, let us define for all $(u, v) \in C_{i,j}$

$$X_{u,v} := \exp\left(\frac{-\|\xi_{i,j}-\xi_{u,v}\|^2}{2k_s^2}\right) \cdot \delta_{u,v}, \quad (3)$$

where

$$\delta_{u,v} := \exp\left(\frac{-(i-u)^2-(j-v)^2}{2k_s^2}\right). \quad (4)$$

With these notation in hand, we see that

$$\forall (i,j) \in I, \quad P_{i,j} = \sum_{(u,v)\in C_{i,j}} X_{u,v},$$

and already we can spot two major difficulties in our analysis. The first is that, even if we assume the $\xi_{i,j}$s to be independent random variables, the $X_{u,v}$s are not, and as a consequence **the $P_{i,j}$s are not independent**.

As a consequence, looking at statements involving several $P_{i,j}$s is very challenging. Our approach in Section 3 will be to simplify the problem by reducing P to its main component Q, using standard tools from the theory of U-statistics. The independence of the $Q_{i,j}$s makes them much more convenient to work with. Our main challenge will be to show that nothing of value is lost when making this approximation.

Second, let us define $I_m(k_w)$ the set of points of I that are further than k_w away from the border of the image. By definition, for all $(i,j) \in I_m(k_w)$, $C_{i,j}$ is the set of points contained in a square of side $2k_w + 1$ centered in (i,j) (see Figure 4). **The picture is more complicated near the borders** of the image: for instance, right next to the border, $C_{i,j}$ has roughly half its size, which drastically lowers the density estimates near the border, even in simple cases. As we will see in Section 4, this has non-trivial consequences. Note that some implementations do normalize Eq. (2) (for instance in the GPU implementation of Fulkerson and Soatto (2010)), which removes this problem.

Algorithm 1 Density estimation.

Require: Image $\xi \in \mathbb{R}^{H \times W}$, kernel size k_s.
1: **Set:** $k_w \leftarrow 3k_s$
2: **Initialize:** $P \in \mathbb{R}^{H \times W}$
3: **for** $(i,j) \in I$ **do**
4:     $P_{i,j} \leftarrow 0$
5:     **for** $(u,v) \in C_{i,j}$ **do**
6:         $P_{i,j} \leftarrow P_{i,j} + \exp\left(\frac{-(i-u)^2 - (j-v)^2 - \|\xi_{i,j} - \xi_{u,v}\|^2}{2k_s^2}\right)$
7:     **end for**
8:     $P_{i,j} \leftarrow P_{i,j} + \mathcal{N}\left(0, \sigma_0^2\right)$
9: **end for**
10: **return** P

2.2 Graph construction

We now turn to the graph construction, the heart of the quickshift algorithm, which we describe for any array $A \in \mathbb{R}^{H \times W}$. Intuitively, quickshift moves each pixel to the nearest neighbor with higher density. When this is no longer possible, we have found a local maximum of the density. Formally, quickshift produces a directed graph $\mathcal{G}_0(A)$ with vertices I in the following way: first, all vertices are visited, and quickshift places an edge between (i,j) and $(u,v) \in C_{i,j}$ if two conditions are satisfied: (i) $A_{i,j} < A_{u,v}$, and (ii) $(i-u)^2 + (j-v)^2 + \|\xi_{i,j} - \xi_{u,v}\|^2$ is minimal among all $(u,v) \in C_{i,j}$ satisfying condition (i). Second, all edges between pixels with (5-dimensional) distance greater than d_m are removed from $\mathcal{G}_0(A)$, where d_m is a positive hyperparameter called *maximal distance* in the `scikit-image` implementation and set to 10 by default. Finally, the superpixels are defined as the *con-*

nected components of $\mathcal{G}_0(A)$. The graph construction is summarized in Algorithm 2.

Algorithm 2 Quickshift graph construction (following `skimage` implementation).

Require: Image $\xi \in \mathbb{R}^{H \times W}$, kernel size k_s, maximum distance d_m, array $A \in \mathbb{R}^{H \times W}$.
1: **set:** $k_w \leftarrow 3k_s$
2: **for** $(i,j) \in I$ **do**
3:     **initialize:** $M \leftarrow +\infty$ with same shape as $C_{i,j}$
4:     **for** $(u,v) \in C_{i,j}$ **do**
5:         $d \leftarrow (i-u)^2 + (j-v)^2 + \|\xi_{i,j} - \xi_{u,v}\|^2$
6:         **if** $A_{i,j} < A_{u,v}$ **and** $d \leqslant d_m$ **then**
7:             $M_{u,v} \leftarrow d$
8:         **end if**
9:     **end for**
10:     $(u,v) \leftarrow \arg\min M$
11:     $G_{(i,j),(u,v)} \leftarrow 1$
12: **end for**
13: **return** G

In definitive, we see quickshift as **the application of the graph construction procedure given by Algorithm 2 to the density estimate obtained by Algorithm 1**. That is, quickshift first computes P by means of Algorithm 1 and then applies Algorithm 2 to $A = P$. Other implementations exist, though they share the same basic ideas. For instance, the VLFeat library normalizes the $P_{i,j}$s by $1/(2\pi k_s)^{5+2}$.

2.3 Computational complexity

A naive implementation of the density estimation step (Algorithm 1) requires a full pass on the image, and for each pixel a pass on $C_{i,j}$. Therefore, the density estimation part of quickshift costs $\mathcal{O}\left(HW k_w^2\right)$. Regarding the graph construction (Algorithm 2), first one has to parse the image once again and look into the $C_{i,j}$ windows, which costs $\mathcal{O}\left(HW k_w^2\right)$. Subsequently, we need to find the connected components of a directed graph with $H \cdot W$ vertices and the same number of edges, which costs $\mathcal{O}\left(HW\right)$ (Hopcroft and Tarjan, 1973). In definitive, the total computational cost is linear in H and W, and quadratic in k_w. This is something to consider when increasing the k_w hyperparameter. Finally, note that the d_m hyperparameter has no influence on the computational cost, though setting small values can lead to producing many superpixels. In particular, by the graph construction, all superpixels have geometric diameter smaller than d_m.

3 HOMOGENEOUS PATCHES

In this section, we focus on flat portions of the image. That is, subsets of I such that $\xi_{i,j}$ is approximately
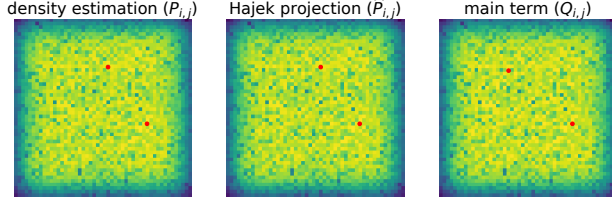
Figure 2: Density estimation under Assumption 3.1 for a 50×50 image with $\sigma = 1.0$. *Left:* using the definition, Eq. (2); *middle:* using the Hájek projection, $\widehat{P}_{i,j}$; *right:* using only the main term $Q_{i,j}$ (Eq. (5)). As predicted by Theorem 3.1, P and Q are close to each other, and so are the associated local maxima (in red). Of course, their number and exact positions can differ.

constant. As a simple model for this situation, we propose the following:

Assumption 3.1 (Flat image). For all $(i,j) \in I$, the $\xi_{i,j}$ are i.i.d. $\mathcal{N}\left(c, \sigma^2\, \mathrm{I}_3\right)$, where $c \in \mathbb{R}^3$ is an arbitrary color and $\sigma \leqslant k_s/5$.

This assumption is a simple model of the random noise coming from an electronic sensor when photographing a mono-color item (see, *e.g.*, Pratt (Chapter 1.4, 2001)). Under this assumption, we derive an approximate expression for P in Section 3.1, before introducing in Section 3.2 the modification of the quickshift algorithm that we investigate here. After precising the interplay between local maxima and superpixels in Section 3.3, we state our main result in Section 3.4.

3.1 A closer look at density estimation

Under Assumption 3.1, it is straightforward to compute the first moments of the $X_{u,v}$, and by extension those of $P_{i,j}$. Let us first define the normalization constant

$$C_p := \left(\frac{k_s^2}{k_s^2 + p\sigma^2}\right)^{3/2}.$$

Then we show in the Appendix that $\mathbb{E}\left[P_{i,j}\right] = C_2 \Delta_{i,j}$, where $\Delta_{i,j} := \sum_{(u,v)\in C_{i,j}} \delta_{u,v}$. As noted in Section 2, the $X_{u,v}$ are not independent. For instance, we show in the Appendix that $\mathrm{Cov}(X_{u,v}, X_{u',v'})$ is always positive under 3.1. Nonetheless, a careful reading of Eq. (3) reveals that $\xi_{i,j}$ draws all values of the density estimate. More precisely, $\xi_{i,j}$ is present in all the $X_{u,v}$ (there are $\mathcal{O}\left(k_s^2\right)$ of them), whereas each other individual $X_{u,v}$ is only present one time. Therefore, when k_s is large, we expect $P_{i,j}$ to behave roughly as its conditional expectation with respect to $\xi_{i,j}$ for a given (i,j), which is given by

$$Q_{i,j} := C_1 \cdot \exp\left(\frac{-\|\xi_{i,j} - c\|^2}{2(k_s^2 + \sigma^2)}\right) \cdot \Delta_{i,j}. \qquad (5)$$

We formalize this intuition by the following:

Theorem 3.1 ($P_{i,j}$ is close to $Q_{i,j}$, w.h.p.). *Assume that 3.1 holds. Suppose furthermore that $k_s \geqslant 5$. Let $(i,j) \in I_m(k_w)$. Then, for any $\varepsilon > 0$,*

$$\mathbb{P}\left(|P_{i,j} - Q_{i,j}| > \varepsilon\right) \leqslant \frac{71\sigma^2}{\varepsilon^2}.$$

The main consequence of this result is that **one can study $Q_{i,j}$ instead of $P_{i,j}$ when σ is small**, at least for a fixed $(i,j) \in I_m(k_w)$. We refer to Figure 2 for an illustration. Theorem 3.1 is the main motivation to study the graph construction step of quickshift on Q instead of P. Of course, looking at the graph construction implies looking simultaneously to several $P_{i,j}$s, a case which is not covered by Theorem 3.1. We conjecture that a uniform bound for all $(i,j) \in I_m(k_w)$ exists, though a straightforward approach *via* a union bound argument does not yield a satisfying bound.

Sketch of the proof of Theorem 3.1. Following Van der Vaart (2000), the key idea of the proof is to compute the Hájek projection of $P_{i,j}$ onto the set of random variables $\xi_{u,v}$, $(u,v) \in I$ to capture the influence of each of the $\xi_{u,v}$. We then show that the projection onto $\xi_{i,j}$, that is, $Q_{i,j}$, is the most prominent one. We start by computing the Hájek projection of $P_{i,j}$, which boils down to Gaussian integral computations under Assumption 3.1, and we find, $\forall (i,j) \in I$,

$$\widehat{P}_{i,j} = Q_{i,j} + C_1 \sum_{\substack{(u,v)\in C_{i,j} \\ (u,v)\neq(i,j)}} \left(\mathrm{e}^{\frac{-\|\xi_{u,v}-c\|^2}{2(k_s^2+\sigma^2)}} - \left(\frac{k_s^2+\sigma^2}{k_s^2+2\sigma^2}\right)^{\frac{3}{2}}\right)\delta_{u,v} \tag{6}$$

We are then able to show that $P_{i,j}$ and $\widehat{P}_{i,j}$ have similar variances when k_s is large enough. More precisely,

$$\left|\frac{\mathrm{Var}(\widehat{P}_{i,j})}{\mathrm{Var}\left(P_{i,j}\right)} - 1\right| \leqslant \frac{4}{5k_s^2}.$$

Therefore, an application of Theorem 11.2 in Van der Vaart (2000) yields:

$$\mathbb{P}\left(\left|\frac{P_{i,j} - \mathbb{E}\left[P_{i,j}\right]}{\sqrt{\mathrm{Var}\left(P_{i,j}\right)}} - \frac{\widehat{P}_{i,j} - \mathbb{E}[\widehat{P}_{i,j}]}{\sqrt{\mathrm{Var}(\widehat{P}_{i,j})}}\right| > \varepsilon\right) \leqslant \frac{1}{k_s^2 \varepsilon^2}.$$

Let us define $R_{i,j} := \widehat{P}_{i,j} - Q_{i,j}$, the rest term. Then we are able to show that $\mathrm{Var}\left(R_{i,j}\right) \leqslant \sigma^2/4$. Thus, for any given (i,j), when k_s is large enough and σ is small with respect to k_s, we can identify $P_{i,j}$ with $\widehat{P}_{i,j}$, which in turn can be identified with $Q_{i,j}$, concluding the proof of Theorem 3.1. $\qquad\square$

3.2 Modified graph construction

We now present a slight modification in the graph construction. Namely, **we remove the** $\|\xi_{i,j} - \xi_{u,v}\|^2$
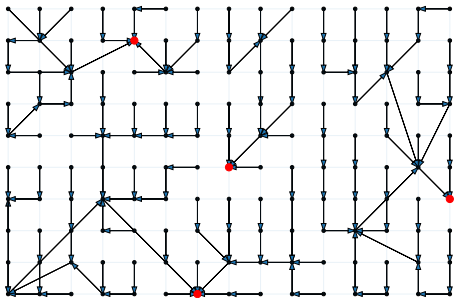
Figure 3: A sample graph $\mathcal{G}(P)$ obtained by the running Algorithm 3 on the density estimate P of an image of size 10×15. Each arrow connects a pixel to the nearest pixel having a higher density estimate. The local maxima are marked in red: there is one such local maxima per connected component of $\mathcal{G}$, as per Lemma 3.1.

term from the distance computation (line 5 of Algorithm 2). Indeed, under Assumption 3.1, $\mathbb{E}[\|\xi_{i,j} - \xi_{u,v}\|^2] = \mathcal{O}(\sigma^2)$. Since the Euclidean distance term is $\mathcal{O}(1)$, the color difference term is negligible whenever σ is small or d_m is large, and there is little difference between the output of Algorithm 2 and 3. Of course, as soon as σ is of the order of magnitude of d_m, notable differences appear. In particular, Algorithm 3 tends to produce much more superpixels in this situation.

A key observation is that, when using this modified version of quickshift, the graph construction simplifies greatly: instead of looking for values of A that are greater than $A_{i,j}$ inside $C_{i,j}$ and subsequently cut off all points with 5-dimensional distance greater than d_m, we can look directly for points in $E_{i,j} := C_{i,j} \cap D_{i,j}$, where $D_{i,j}$ is the set of points of I such that $\sqrt{(i-u)^2 + (j-v)^2} \leqslant d_m$. There are three possibilities for $E_{i,j}$, depending on the relative position of k_w and d_m: we depict these three cases in Figure 4. We call $\mathcal{G}(A)$ this procedure, described formally in Algorithm 3 and illustrated in Figure 3.

To conclude this section, we want to emphasize that the core analysis of this paper concerns $\mathcal{G}(Q)$, not $\mathcal{G}_0(P)$ which is the output of the default implementation.

Algorithm 3 Simplified quickshift graph construction.

Require: Image $\xi \in \mathbb{R}^{H \times W}$, kernel size k_s, maximum distance d_m, array $A \in \mathbb{R}^{H \times W}$.
1: **set:** $k_w \leftarrow 3k_s$
2: **for** $(i, j) \in I$ **do**
3:    $(u, v) \leftarrow$ closest element of $E_{i,j}$ s.t. $A_{i,j} < A_{u,v}$
4:    $G_{(i,j),(u,v)} \leftarrow 1$
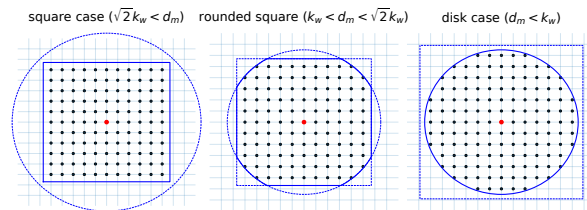5: **end for**
6: **return** G

Figure 4: Three possibilities for the lookout area of Algorithm 3. At a given $(i, j) \in I$ (in red), the algorithm looks at the nearest neighbor with higher value in $E_{i,j} = C_{i,j} \cap D_{i,j}$ (solid blue line). The shape of the intersection depends on the interplay between k_s and d_m. For instance, the default choice $k_s = 5$ and $d_m = 10$ yields $k_w = 15$, which is greater than d_m: we are in the disk case. This trichotomy remains when both k_s and d_m are multiplied by a constant factor.

3.3 Counting superpixels

We now return to our main focus: counting the number of superpixels. For a given region R of the image, we set $K_R(A)$ the number of connected components of $\mathcal{G}(A)$ intersecting R. It is challenging to investigate $K_R(A)$ directly, and we rather study the number of *local maxima*:

Definition 3.1 (Local maximum). Let $A \in \mathbb{R}^{H \times W}$ and $(i, j) \in I$. We say that (i, j) is a local maximum of A if $\forall (u, v) \in E_{i,j}$, $A_{i,j} > A_{u,v}$.

We will use the notation $N_R(A)$ to denote the number of local maxima of A inside a given region R. The main reason to study local maxima is that, on a global scale, there is an equivalence between the number of superpixels and the number of local maxima, as is demonstrated in Figure 3. This fact is formalized by our next lemma:

Lemma 3.1 (Connected components and local maxima). *Let $\mathcal{G}(A)$ be the directed graph produced by Algorithm 3 applied to an array $A \in \mathbb{R}^{H \times W}$. Then to each connected component of $\mathcal{G}(A)$ corresponds a unique local maxima of A in the sense of Definition 3.1.*

In view of Lemma 3.1, we will now focus on $N_R(A)$, which is *a priori* distinct from $K_R(A)$. Indeed, when restricting ourselves to R, it can happen that we miss the local maxima corresponding to the connected component intersecting with R. Namely, only the trivial lower bound

$$N_R(A) \leqslant K_R(A) \tag{7}$$

holds. There is no corresponding upper bound in the general case, since a rectangle can intersect many superpixels without containing *any* local maximum as depicted in Figure 5. Nevertheless, empirically, $N_R(A)$ and $K_R(A)$ are comparable for large regions (and the bound given by Eq. (7) is reasonably tight).
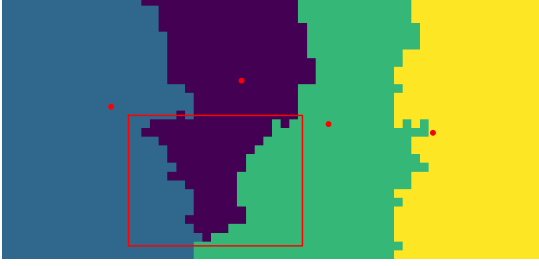
Figure 5: The number of local maxima inside a part of the image does not always coincide with the number of superpixels intersecting this area. In this example, R is a rectangle with boundaries marked in red. The number of local maxima inside R is $N_R(A) = 0$, whereas the number of superpixels intersecting R is $K_R(A) = 3$.

3.4 Average number of local maxima in flat regions

We are now ready to state and prove the main result of this section.

Theorem 3.2 (Average number of local maxima). *Assume that 3.1 holds. Let $R \subseteq I$ be a rectangle of height h and width w at distance greater than $2k_w$ from the border. Then*

$$\mathbb{E}\left[N_R(Q)\right] = \begin{cases} hw \cdot \left(\frac{1}{\pi d_m^2} + \mathcal{O}\left(\frac{1}{d_m}\right)\right) & \text{if } d_m \leqslant k_w \\ hw \cdot \left(\frac{1}{\pi(3k_w d_m - k_w^2 - d_m^2)} + \mathcal{O}\left(\frac{1}{k_w}\right)\right) \\ \qquad\qquad \text{if } k_w < d_m \leqslant \sqrt{2}k_w \\ hw \cdot \left(\frac{1}{4k_w^2} + \mathcal{O}\left(\frac{1}{k_w}\right)\right) & \text{otherwise.} \end{cases}$$

The main consequence of Theorem 3.2 is the scale of $N_R(Q)$ with respect to (i) the size of R, and (ii) the hyperparameters. In plain words, firstly, we find $N_R(Q)$ to be **proportional to the area of** R: a flat, rectangular part of the image will contain twice as many superpixels if its size is multiplied by two. Secondly, whatever the relative position of k_s and d_m is, $N_R(Q)$ **is** (-2)-**homogeneous in** (k_s, d_m): doubling both k_s and d_m will approximately divide $N_R(Q)$ by four. We illustrate the validity of Theorem 3.2 with respect to the non-modified version of quickshift (Algorithm 2 running on P) in Figure 6.

Proof of Theorem 3.2. The key of the proof is the following result, which is the main motivation for finding an i.i.d. approximation of P:

Lemma 3.2 (Key lemma). *Let $A \in \mathbb{R}^{H \times W}$ be a random array such that the $A_{i,j}$ are i.i.d. with cumulative distribution function G and associated density g. Let $R \subseteq I$. Define $N_{i,j} := \#E_{i,j}$. Then the expected number of local maxima in R produced by Algorithm 3*

applied to A is given by

$$\mathbb{E}\left[N_R(A)\right] = \sum_{(i,j) \in R} \frac{1}{N_{i,j}}.$$

Note that, as a consequence, Theorem 3.2 is not limited to a rectangular area, though it is less convenient to state for a general shape.

Proof of Lemma 3.2. We first write $N_R(A)$ as the sum of indicator that (i,j) is a local maximum. Further,

$$\mathbb{E}\left[N_R(A)\right] = \mathbb{E}\left[\sum_{(i,j) \in R} \mathbb{1}_{\forall(u,v) \in E_{i,j}^\star,\ A_{i,j} > A_{u,v}}\right]$$
$$= \sum_{(i,j) \in R} \mathbb{P}\left(\forall(u,v) \in E_{i,j}^\star,\ A_{i,j} > A_{u,v}\right),$$

where we write $E_{i,j}^\star$ short for $E_{i,j} \backslash \{(i,j)\}$. Let us condition with respect to $A_{i,j}$. Using the i.i.d. assumption on the $A_{i,j}$s, by definition of the cumulative distribution function G,

$$\mathbb{P}\left(\forall(u,v) \in E_{i,j}^\star,\ A_{i,j} > A_{u,v} \big| A_{i,j}\right) = G(A_{i,j})^{N_{i,j}-1}.$$

Integrating the last display yields the result, since

$$\int G(t)^{N_{i,j}-1} g(t)\, \mathrm{d}t = \left[\frac{G(t)^{N_{i,j}}}{N_{i,j}}\right]_{-\infty}^{+\infty} = \frac{1}{N_{i,j}}.$$

$\square$

Further, we see that the $Q_{i,j}$ are indeed i.i.d. if we are outside a band of width $2k_w$ from the border: here, $\Delta_{i,j}$ does not depend on (i,j). Thus we can apply Lemma 3.2 to the random array Q. Finally, it is just a matter of counting the number of points inside $E_{i,j}$. For instance, let us assume that $d_m \leqslant k_w$. In that case, we have to count the number of lattice points inside a disk of radius d_m. This is known as the *Gauss circle problem*, and an immediate bound (due to Gauss himself) is $N_{i,j} = \pi d_m^2 + \mathcal{O}(d_m)$. Summing over all pixels in R yields the last result in Theorem 3.2. The other cases are similar and the details can be found in the Appendix. $\square$

4 SHARP EDGES

In this short section, we focus on well-defined boundaries between homogeneous patches of the image. More precisely, we assume the following:

Assumption 4.1 (Bicolor image). There exist $j_0 \in [W]$, $c_1, c_2 \in \mathbb{R}^3$ such that the $\xi_{i,j}$ are i.i.d. $\mathcal{N}\left(c_1, \sigma^2 \mathrm{I}_3\right)$ (resp. $\mathcal{N}\left(c_2, \sigma^2 \mathrm{I}_3\right)$) on $I_\ell := [H] \times \{1, \ldots, j_0\}$ (resp. $I_r := [H] \times \{j_0, \ldots, W\}$), with $\sigma \leqslant k_s/5$.
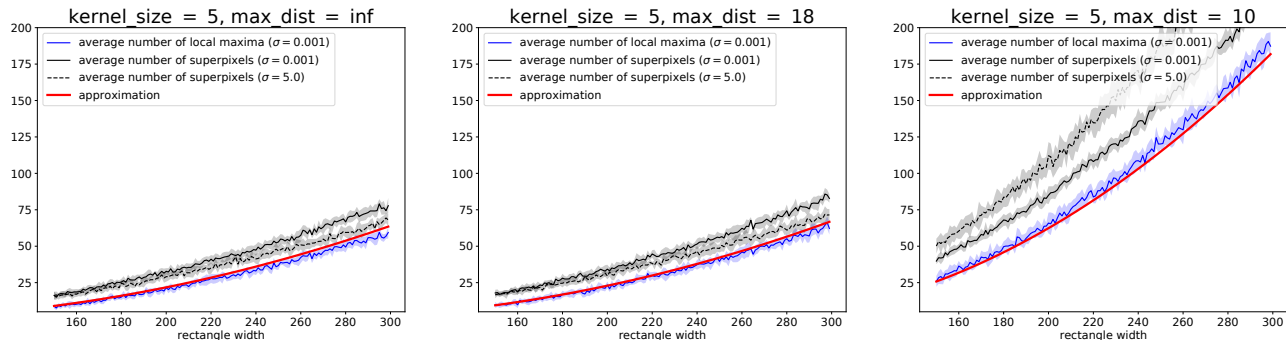
Figure 6: Evolution of the number of superpixels created by quickshift under Assumption 3.1. For each input shape, we run quickshift on ten images and count the number of local maxima and the number of superpixels in a rectangular patch far from the border of the image. We see that, when σ is small, the prediction of Theorem 3.2 (in red) is close to the empirical number of local maxima (in blue). Moreover, the number of local maxima coincides roughly with the number of superpixels (in black). When σ increases, as expected, this link weakens but we still observe a quadratic dependency in the size of the rectangle. The three choices of (k_s, d_m) correspond to the three different configurations (square, rounded square, and disk).
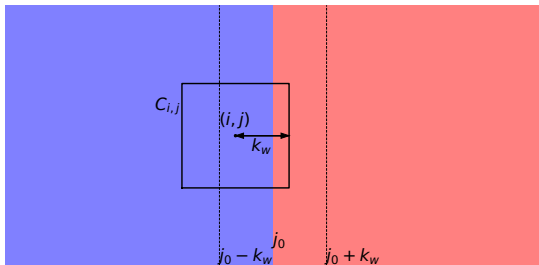

Figure 7: The bicolor setting described by Assumption 4.1. Here, c_1 is a light blue and c_2 a light red. The border between the two patches marked at j_0 is in the middle of the image.

In plain words, we consider a bicolor image with a vertical boundary at j_0 between them and i.i.d. Gaussian noise as in Section 3. We refer to Figure 7 for a visual depiction. Note that we can also work under Assumption 4.1 to understand what happens at the border of a flat image: Let us consider (i, j) a pixel near the boundary of the image. In that case, $C_{i,j}$ contains fewer points than in the center of the image since the square centered at (i, j) "overflows" the image, and therefore the sum in Eq. (2) is missing some terms. Instead of thinking of $C_{i,j}$ as reduced, we can think of $C_{i,j}$ intersecting a color patch where $\|\xi_{u,v} - \xi_{i,j}\|$ is very large for all (u, v) in that patch. In that event, the corresponding terms in Eq. (2) vanish. Without further ado, we can state the main result of this section:

Theorem 4.1 (Increasing density estimates). *As-sume that 4.1 holds. Assume further that $k_s \geqslant 5$ and that $\|c_1 - c_2\| \geqslant 3k_s$. Then, for any $(i, j) \in I_m \cap I_\ell$ such that $|j - j_0| \leqslant k_w$,*

$$\mathbb{P}\left(P_{i,j} > P_{i,j+1}\right) \geqslant 1 - 16\sigma^2 .$$

In a nutshell, Theorem 4.1 states that, in the k_w-neighborhood of the boundary, the density estimates are increasing away from the border. The main consequence of Theorem 4.1 is the **absence of local maxima near the boundary** between two homogeneous color patches when Algorithm 3 is applied to P. Indeed, the nearest neighbor of $(i, j + 1)$ with highest density will be (i, j) with high probability, and thus the edges in a band of width k_w around the boundary are all pointing away from the boundary. An additional consequence is that **the boundary is well-recognized by Algorithm 3 provided that the color difference is large enough with respect to k_s.** Indeed, for all $i \in [H]$, (i, j_0) is linked to $(i, j_0 - 1)$ and $(i, j_0 + 1)$ to (i, j_0+2) by symmetry, with high probability. Thus it is very unlikely that a point of I_ℓ belongs to the superpixel corresponding to I_r. We illustrate this phenomenon in Figure 8. We note that it is straightforward to adapt the proof of Theorem 4.1 for a horizontal boundary.

Sketch of the proof of Theorem 4.1. As in the homogeneous case, we can compute the expected estimated density under Assumption 4.1, which takes a slightly more involved expression:

$$\mathbb{E}\left[P_{i,j}\right] = C_2 \cdot \sum_{(u,v) \in C_{i,j} \cap I_\ell} \delta_{u,v} \qquad (8)$$
$$+ C_2 \cdot \exp\left(\frac{-\|c_1 - c_2\|^2}{2(k_s^2 + 2\sigma^2)}\right) \cdot \sum_{(u,v) \in C_{i,j} \cap I_r} \delta_{u,v} .$$

Using Eq. (8), we can look at the difference in expected estimated density between two neighbors. We prove that, if the color difference is large enough, the difference between the expected estimated density at two neighboring points is lower bounded by a term of

order k_s. Namely, provided that $\|c_1 - c_2\| \geqslant 3k_s$,

$$\mathbb{E}\left[P_{i,j}\right] - \mathbb{E}\left[P_{i,j+1}\right] \geqslant \frac{3k_s}{2}\,.$$

Finally, since the same upper bound on the variance of $P_{i,j}$ as in the homogeneous case holds, we can use Chebyshev's inequality to find an event of large probability Ω on which both $|P_{i,j} - \mathbb{E}\left[P_{i,j}\right]|$ and $|P_{i,j+1} - \mathbb{E}\left[P_{i,j+1}\right]|$ are smaller than $3k_s/4$. Therefore, on Ω, by the triangle inequality, $P_{i,j} > P_{i,j+1}$. $\qquad\square$
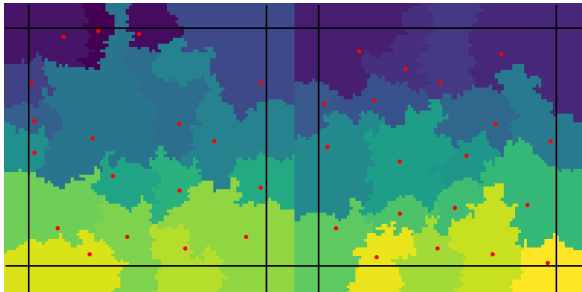

Figure 8: Illustration of Theorem 4.1: quickshift segmentation of a bicolor image with a vertical border. In this setting, the density estimate is increasing away from the border between parts of the image of different colors, with high probability. As a consequence, quickshift does not find local maxima (the red dots) in this area (marked by the thick dark lines).

5 EXPERIMENTS

In this section, we see how the claims of Section 3 extend to real images. We consider three datasets: (i) a subset of the ILSVRC2017 dataset (Russakovsky et al., 2015). From the original 5500 images, we manually picked 748 of them where a large, rectangular part of the image is visually homogeneous in color. (ii) a subset of the CityScapes dataset (Cordts et al., 2016). More precisely, the 544 images taken in Berlin from the `leftImg8bit_trainvaltest` folder. (iii) a random subset of 500 images from the `JPEGImages` folder of the Pascal VOC dataset (Everingham et al., 2015). Note that the images from the CityScapes dataset are much larger than those of the ILSVRC and PascalVOC dataset (typically 1024×2048 *vs* 300×500). We first check the scaling with respect to the size of the image in Section 5.1 and the hyperparameters in Section 5.2, before concluding with a practical use-case in Section 5.3.

5.1 Scaling with respect to the image size

According to the discussion following Theorem 3.2, for any choice of hyperparameters, the expected number of superpixels on a rectangular, homogeneous part of an image of size $h \times w$ should scale as hw. To test this,

we conducted the following experiment. For a given choice of hyperparameters, we first segmented each of the images in our datasets with quickshift, yielding n_{orig} superpixels. In a second step, we rescaled the image by a given ratio ρ, dividing both H and W by ρ, and proceeded with the same experiment, yielding n_{new} superpixels. According to Theorem 3.2, one should observe $n_{\text{orig}}/n_{\text{new}} \approx \rho^2$. We report in Table 1 the empirical average of $n_{\text{orig}}/n_{\text{new}}$ for $\rho \in \{2, 3\}$: as expected, this empirical average is close to ρ^2 in all the hyperparameter configurations that we tested. We note that this relationship is the weakest for small d_m, as showed by the great variability of the ratio.

5.2 Scaling with respect to k_s and d_m

Next, we turn to the influence of the hyperparameters. We now consider a fixed image, and multiply both k_s and d_m by a constant factor κ. According to the discussion following Theorem 3.2, one expects the average number of superpixels produced by quickshift on a flat portion of the image to be divided, roughly, by a factor κ^2. To test this hypothesis, we conducted a similar experiment to that of the previous section. First, we segmented the images of our dataset for a given choice of k_s and d_m, and counted the number of superpixels obtained (as in Section 5.1, we call this number n_{orig}). We then segmented again all images with quickshift, this time with hyperparameters $(\kappa k_s, \kappa d_m)$ and counted the number of superpixels obtained in that case (n_{new}). We report the average ratio $n_{\text{orig}}/n_{\text{new}}$ in Table 2 for $\kappa \in \{0.5, 2\}$. Again, we obtain values close to $1/\kappa^2$ as expected when d_m is large with respect to k_s. For small d_m, the ratio can be off by a constant factor.

5.3 A practical use case

Suppose that we have calibrated an image processing pipeline on downsized images with dimensions $(H/\rho, W/\rho)$ and now want to get back to the original size, in effect multiplying both H and W by a factor ρ, without the number of superpixels produced by quickshift changing too drastically. Combining the insights from Section 5.1 and 5.2, we see that running quickshift with hyperparameters $(\rho k_s, \rho d_m)$ on the original image will yield approximately the same number of superpixels. This simple heuristic provides a way to scale quickshift hyperparameters. We illustrate this process is Figure 9. In addition to recovering a similar number of superpixel, we note that their shapes are similar: increasing k_s, in certain limits, does not seem to damage the density estimation step too much.

As noted before, this heuristic works best when d_m is large, otherwise the deviations observed in Table 1 and 2 accumulate. We see two main reasons for this.

Table 1: Scaling with respect to the image size. We report the empirical average and standard deviation of $n_{\text{orig}}/n_{\text{new}}$ for all datasets and three hyperparameter configurations.

k_s	d_m	ILSVRC		Cityscapes		Pascal VOC	
		$\rho^2 = 4$	$\rho^2 = 9$	$\rho^2 = 4$	$\rho^2 = 9$	$\rho^2 = 4$	$\rho^2 = 9$
5	$+\infty$	3.8 (0.7)	7.9 (2.1)	4.3 (0.2)	10.2 (0.8)	3.9 (0.7)	8.2 (2.1)
5	18	3.3 (0.8)	6.3 (2.3)	4.0 (0.2)	9.0 (0.8)	3.3 (0.7)	6.4 (2.1)
5	10	4.3 (2.2)	9.1 (7.7)	3.9 (0.2)	8.6 (0.9)	4.3 (1.9)	8.9 (5.6)

Table 2: Scaling with respect to the hyperparameters. We report the empirical average and standard deviation of $n_{\text{orig}}/n_{\text{new}}$ for all datasets and three hyperparameter configurations.

k_s	d_m	ILSVRC		Cityscapes		Pascal VOC	
		$\kappa^{-2} = 0.25$	$\kappa^{-2} = 4$	$\kappa^{-2} = 0.25$	$\kappa^{-2} = 4$	$\kappa^{-2} = 0.25$	$\kappa^{-2} = 4$
5	$+\infty$	0.26 (0.04)	3.7 (0.4)	0.22 (0.01)	3.8 (0.1)	0.24 (0.05)	3.8 (0.4)
5	18	0.18 (0.05)	10.9 (5.3)	0.20 (0.01)	5.5 (0.8)	0.18 (0.04)	10.9 (5.4)
5	10	0.11 (0.05)	14.4 (5.5)	0.17 (0.02)	8.1 (1.3)	0.11 (0.05)	16.3 (7.6)

First, we studied and derived a heuristic from a modified version of the quickshift algorithm. As explained in Section 3.2, these two versions can be quite different when d_m is small. Second, even on flat portions of the image, the distribution of the pixel values does not quite satisfy Assumption 3.1. For instance, the variance of the pixel values is typically much higher than in the regime where the closed-form expressions of Theorem 3.2 holds. There can also be a lot of spacial dependencies between pixel values, thus breaking the independent part of our assumption. We provide additional experiments in the Appendix (Section E.2) to demonstrate this qualitatively.
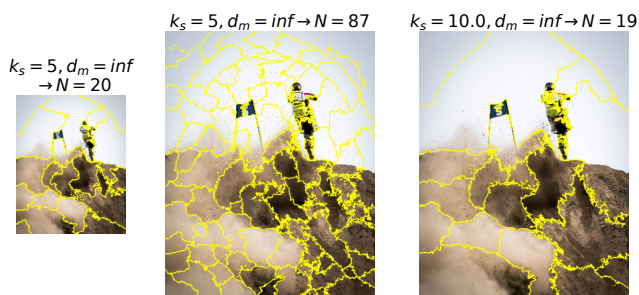

Figure 9: Scaling quickshift hyperparameters. *Left:* segmentation of a downsized image (by a factor $\rho = 2$) with $k_s = 5$ and $d_m = +\infty$, quickshift finds 20 superpixels; *middle:* segmentation of the original image with the same hyperparameters (87 superpixels); *right:* segmentation of the original image with rescaled hyperparameters ($k_s = 2 \times 5 = 10$), quickshift finds 19 superpixels, approximately the same number as in the downsized image.

6 CONCLUSION

In this paper, we investigate the relationship between the number of superpixels produced by quickshift and the choice of the two main hyperparameters of the method: k_s and d_m. Theoretically, we find that, on a flat portion of the image and for a modified version of the algorithm, this number is proportional to the size of the patch and inversely proportional to the square of k_s or d_m (depending on their relative position). Experimentally, we see that this scaling law is true to some extent for the original algorithm applied to real images. This provides a simple heuristic to keep the number of superpixels approximately constant when going from small to large images. We also show that quickshift accurately detects the borders of homogeneous patches provided that k_s is large enough with respect to the color difference.

As future work, our main focus is to tackle the original version of the algorithm, in order to capture better the behavior for small d_m. The main difficulty in doing so seems to be the extension of Lemma 3.2, which seems challenging even considering i.i.d. inputs since the lookout area is no longer deterministic. We also want to obtain a uniform version of Theorem 3.1, which in turn would give a more rigorous justification in replacing P by Q in our analysis.

References

R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11): 2274–2282, 2012.

E. Arias-Castro, D. Mason, and B. Pelletier. On the estimation of the gradient lines of a density and the consistency of the mean-shift algorithm. *The Journal of Machine Learning Research*, 17(1):1487–1514, 2016.

Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):790–799, 1995.

D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1): 98–136, 2015.

P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2004.

B. Fulkerson and S. Soatto. Really quick shift: Image segmentation on a GPU. In *European Conference on Computer Vision*, pages 350–358. Springer, 2010.

D. Garreau and D. Mardaoui. What does LIME really see in images? In *Proceedings of the 38th International Conference on Machine Learning*, pages 3620–3629, 2021.

D. Garreau and U. von Luxburg. Explaining the explainer: A first theoretical analysis of LIME. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, pages 1287–1296, 2020.

G. H. Hardy. *Ramanujan: Twelve lectures on subjects suggested by his life and work*, volume 136. American Mathematical Society, third edition, 1959.

J. Hopcroft and R. Tarjan. Algorithm 447: efficient algorithms for graph manipulation. *Communications of the ACM*, 16(6):372–378, 1973.

M. N. Huxley. The rational points close to a curve II. *Acta Arithmetica*, 93(3):201–219, 2000.

H. Jiang. On the consistency of quick shift. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

H. Jiang and S. Kpotufe. Modal-set estimation with an application to clustering. In *Artificial Intelligence and Statistics*, pages 1197–1206. PMLR, 2017.

H. Jiang, J. Jang, and S. Kpotufe. Quickshift++: Provably good initializations for sample-based mean shift. In *Proceedings of the 35th International Conference on Machine Learning*, pages 2294–2303, 2018.

P. Neubert and P. Protzel. Compact watershed and pre-emptive SLIC: On improving trade-offs of superpixel segmentation algorithms. In *2014 22nd International Conference on Pattern Recognition*, pages 996–1001. IEEE, 2014.

E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.

W. K. Pratt. *Digital image processing: PIKS inside*. John Wiley & sons, third edition, 2001.

M. T. Ribeiro, S. Singh, and C. Guestrin. "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1135–1144, 2016.

A. Rodriguez and A. Laio. Clustering by fast search and find of density peaks. *Science*, 344(6191):1492–1496, 2014.

O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.

T. Ryu, B. G. Lee, and S.-H. Lee. Image compression system using colorization and meanshift clustering methods. In Y.-S. Jeong, Y.-H. Park, C.-H. R. Hsu, and J. J. J. H. Park, editors, *Ubiquitous Information Technologies and Applications*, pages 165–172, 2014.

A. W. Van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 3rd edition, 2000.

S. Van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, and T. Yu. `scikit-image`: image processing in python. *PeerJ*, 2:e453, 2014.

A. Vedaldi and S. Soatto. Quick shift and kernel methods for mode seeking. In *European conference on computer vision*, pages 705–718. Springer, 2008.

I. Verdinelli and L. Wasserman. Analysis of a mode clustering diagram. *Electronic Journal of Statistics*, 12(2):4288–4312, 2018.

S. Zhang, Z. Ma, G. Zhang, T. Lei, R. Zhang, and
Y. Cui. Semantic image segmentation with deep
convolutional neural networks and quick shift. *Symmetry*, 12(3):427, 2020.

Supplementary material

In this appendix, we collect all missing proofs from the main paper and present some additional experimental results. It is organized as follows: Theorem 3.1, stating that the density estimates can be approximated by the $Q_{i,j}$ term, is proved in Section A. Section B is dedicated to the proof of Theorem 3.2, which gives the expected number of local maxima in a flat portion of the image. In Section C, we prove Theorem 4.1 of the paper, stating that the density estimates are increasing away from the boundary between two homogeneous patches of the image. Truly technical results are collected in Section D. Finally, we present some additional experiments, mainly in relation to the use-case, in Section E.

A   DENSITY ESTIMATES APPROXIMATION

In this section, we provide a complete proof of Theorem 3.1 of the paper. We follow the sketch of the proof provided in Section 3.1 of the paper: after providing some elementary facts about $P_{i,j}$ in Section A.1, we compute its Hájek projection onto the $\xi_{i,j}$s, $\widehat{P}_{i,j}$, in Section A.2. In Section A.3, we show that $P_{i,j}$ is close to $\widehat{P}_{i,j}$ with high probability, essentially proving that they have similar variances for large k_s. Finally, we show in Section A.4 that $\widehat{P}_{i,j}$ is close to the main term $Q_{i,j}$, and we conclude.

A.1   Elementary computations

Recall that, for any $(i,j) \in I$, we defined the observation window

$$C_{i,j} := \{(u,v) \in I, \text{ s.t. } |i - u| \vee |j - v| \leqslant k_w\}, \tag{9}$$

corresponding to all pixels of I located within a square of side $2k_w$ centered at (i,j). Our main object of interest in this section is the density estimate $P_{i,j} = \sum_{(u,v) \in C_{i,j}} X_{u,v}$, where

$$X_{u,v}^{i,j} := \exp\left(\frac{-\|\xi_{i,j} - \xi_{u,v}\|^2}{2k_s^2}\right) \delta_{u,v},$$

(Eq. (3) in the paper) and

$$\delta_{u,v}^{i,j} := \exp\left(\frac{-(i-u)^2 - (j-v)^2}{2k_s^2}\right). \tag{10}$$

As announced, we start our study by some elementary derivations, which we will use in the rest of this Appendix. Recall that we defined the normalization constant

$$\forall p \geqslant 1, \qquad C_p := \left(\frac{k_s^2}{k_s^2 + p\sigma^2}\right)^{\frac{3}{2}}. \tag{11}$$

This normalization constant is important since it appears in most of the computations involving the expected value of the $X_{u,v}$ random variables under Assumption 3.1. Indeed, we have the following:

Lemma A.1 (Moments of $X_{u,v}$). *Let $(i,j) \in I$ and $(u,v) \in C_{i,j} \backslash \{(i,j)\}$. Assume that $\xi_{i,j} \sim \mathcal{N}\left(c_1, \sigma^2 I_3\right)$ and $\xi_{u,v} \sim \mathcal{N}\left(c_2, \sigma^2 I_3\right)$ with $c_1, c_2 \in \mathbb{R}^3$. Then, for any $p \geqslant 1$,*

$$\mathbb{E}\left[X_{u,v}^p | \xi_{i,j}\right] = C_p \cdot \exp\left(\frac{-p \|\xi_{i,j} - c_2\|^2}{2(k_s^2 + p\sigma^2)}\right) \cdot \delta_{u,v}^p,$$

and

$$\mathbb{E}\left[X_{u,v}^p\right] = C_{2p} \cdot \exp\left(\frac{-p \|c_1 - c_2\|^2}{2(k_s^2 + 2p\sigma^2)}\right) \cdot \delta_{u,v}^p.$$

Under Assumption 3.1, we can easily deduce from Lemma A.1 the expected value of $X_{u,v}$, that is,

$$\mathbb{E}\left[X_{u,v}\right] = C_2 \cdot \delta_{u,v} \,. \tag{12}$$

Recall that we defined

$$\Delta_{i,j} := \sum_{(u,v)\in C_{i,j}} \delta_{u,v} \,, \tag{13}$$

by linearity we find that

$$\mathbb{E}\left[P_{i,j}\right] = C_2 \cdot \Delta_{i,j} \,. \tag{14}$$

Proof. We begin by the computation of the conditional expectation. Conditionally to $\xi_{i,j}$, we note that $X_{u,v}^p$ can be written as a product of three independent random variables. Namely,

$$\mathbb{E}\left[X_{u,v}^p|\xi_{i,j}\right] = \mathbb{E}\left[\exp\left(\frac{-p\,\|\xi_{i,j}-\xi_{u,v}\|^2}{2k_s^2}\right)\bigg|\xi_{i,j}\right] \cdot \delta_{u,v}^p$$

$$= \prod_{k=1}^{3}\mathbb{E}\left[\exp\left(\frac{-p(\xi_{i,j,k}-\xi_{u,v,k})^2}{2k_s^2}\right)\bigg|\xi_{i,j,k}\right] \cdot \delta_{u,v}^p \,. \tag{15}$$

Let us fix $k \in \{1,2,3\}$. The inner term can be written

$$\mathbb{E}\left[\exp\left(\frac{-(\xi_{i,j,k}-\xi_{u,v,k})^2}{2k_s^2}\right)\bigg|\xi_{i,j,k}\right] = \int \exp\left(\frac{-p(\xi_{i,j,k}-x)^2}{2k_s^2}\right)\cdot\exp\left(\frac{-(x-c_{2,k})^2}{2\sigma^2}\right)\frac{\mathrm{d}x}{\sigma\sqrt{2\pi}}\,,$$

since $\xi_{u,v} \sim \mathcal{N}\left(c_2,\sigma^2\,\mathrm{I}_3\right)$. We apply Lemma D.1 with $a = \xi_{i,j,k}$, $b = c_{2,k}$, $c = k_s/\sqrt{p}$, and $d = \sigma$ to obtain

$$\mathbb{E}\left[\exp\left(\frac{-p(\xi_{i,j,k}-\xi_{u,v,k})^2}{2k_s^2}\right)\bigg|\xi_{i,j,k}\right] = \left(\frac{k_s^2}{k_s^2+p\sigma^2}\right)^{\frac{1}{2}}\cdot\exp\left(\frac{-p(\xi_{i,j,k}-c_{2,k})^2}{2(k_s^2+p\sigma^2)}\right)\,.$$

Coming back to Eq. (15), we find the first statement of the lemma to be true.

To take the expectation with respect to $\xi_{i,j}$, we take the same road and first write

$$\mathbb{E}\left[X_{u,v}^p\right] = \left(\frac{k_s^2}{k_s^2+p\sigma^2}\right)^{\frac{3}{2}}\cdot\prod_{k=1}^{3}\mathbb{E}\left[\exp\left(\frac{-p(\xi_{i,j,k}-c_{2,k})^2}{2(k_s^2+p\sigma^2)}\right)\right]\cdot\delta_{u,v}^p\,. \tag{16}$$

Fix $k \in \{1,2,3\}$. The inner term can be written

$$\mathbb{E}\left[\exp\left(\frac{-p(\xi_{i,j,k}-c_{2,k})^2}{2(k_s^2+p\sigma^2)}\right)\right] = \int \exp\left(\frac{-p(x-c_{2,k})^2}{2(k_s^2+p\sigma^2)}\right)\cdot\exp\left(\frac{-(x-c_{1,k})^2}{2\sigma^2}\right)\frac{\mathrm{d}x}{\sigma\sqrt{2\pi}}\,.$$

We apply Lemma D.1 with $a = c_{2,k}$, $b = c_{1,k}$, $c^2 = (k_s^2+p\sigma^2)/p$, and $d^2 = \sigma^2$ to obtain

$$\mathbb{E}\left[\exp\left(\frac{-p(\xi_{i,j,k}-c_{2,k})^2}{2(k_s^2+p\sigma^2)}\right)\right] = \left(\frac{k_s^2+p\sigma^2}{k_s^2+2p\sigma^2}\right)^{\frac{1}{2}}\cdot\exp\left(\frac{-p(c_{1,k}-c_{2,k})^2}{2(k_s^2+2p\sigma^2)}\right)\,.$$

Coming back to Eq. (16), we have proved the second statement of the lemma. $\qquad\square$

We now introduce two important functions for our study.

Definition A.1 (ψ functions). For any $t \geqslant 0$, we let

$$\psi_1(t) := \frac{1}{(1+4t)^{\frac{3}{2}}} - \frac{1}{(1+2t)^3} \quad\text{and}\quad \psi_2(t) := \frac{1}{(1+t)^{\frac{3}{2}}(1+3t)^{\frac{3}{2}}} - \frac{1}{(1+2t)^3}\,.$$

The main reason in introducing these auxiliary functions is their appearance in the variance computations that are key to our analysis. For instance, Lemma A.1 implies that

$$\mathrm{Var}\left(X_{u,v}\right) = \left[\left(\frac{k_s^2}{k_s^2+4\sigma^2}\right)^{\frac{3}{2}} - \left(\frac{k_s^2}{k_s^2+2\sigma^2}\right)^3\right]\cdot\delta_{u,v}^2 = \psi_1\left(\frac{\sigma^2}{k_s^2}\right)\cdot\delta_{u,v}^2\,. \tag{17}$$

We also have the following:

Lemma A.2 (Covariance structure of the $X_{u,v}$). *Under Assumption 3.1, for any distinct $(u,v),(u',v') \in C_{i,j}$,*

$$\text{Cov}(X_{u,v}, X_{u',v'}) = \psi_2\left(\frac{\sigma^2}{k_s^2}\right) \cdot \delta_{u,v}\delta_{u',v'} . \tag{18}$$

Some technical facts about ψ_1 and ψ_2 are collected in Section D.3. For instance, according to Lemma D.5, $\psi_2(t)$ is positive for any $t \geq 0$. We deduce that, under Assumption 3.1, $X_{u,v}$ and $X_{u',v'}$ are positively correlated for any $(u,v),(u',v') \in C_{i,j}$.

Proof. Let $(u,v),(u',v') \in C_{i,j}$ be distinct. We write

$$\mathbb{E}\left[X_{u,v}X_{u',v'}|\xi_{i,j}\right] = \mathbb{E}\left[X_{u,v}|\xi_{i,j}\right] \cdot \mathbb{E}\left[X_{u',v'}|\xi_{i,j}\right]$$
$$= \left(\frac{k_s^2}{k_s^2 + \sigma^2}\right)^3 \cdot \exp\left(\frac{-\|\xi_{i,j} - c\|^2}{2((k_s^2 + \sigma^2)/2)}\right) \cdot \delta_{u,v}\delta_{u',v'} ,$$

where we used Lemma A.1 with $p = 1$. Marginalizing with respect to $\xi_{i,j}$, we obtain

$$\mathbb{E}\left[X_{u,v}X_{u',v'}\right] = \left(\frac{k_s^2}{k_s^2 + \sigma^2}\right)^3 \cdot \left(\frac{k_s^2 + \sigma^2}{k_s^2 + 3\sigma^2}\right)^{\frac{3}{2}} \cdot \delta_{u,v}\delta_{u',v'} ,$$

again using Lemma A.1. Eq. (12) combined with some straightforward algebra yields the promised result. $\square$

Putting together Eq. (17) and Eq. (18), we see that

$$\text{Var}\left(P_{i,j}\right) = \psi_2\left(\frac{\sigma^2}{k_s^2}\right) \cdot \sum_{\substack{(u,v),(u',v')\in C_{i,j} \\ (u,v)\neq(u',v')}} \delta_{u,v}\delta_{u',v'} + \psi_1\left(\frac{\sigma^2}{k_s^2}\right) \cdot \sum_{(u,v)\in C_{i,j}} \delta_{u,v}^2 . \tag{19}$$

It can be cumbersome to work directly with this expression. We propose the following lower bound for the variance of $P_{i,j}$:

Lemma A.3 (Lower bound on $\text{Var}\left(P_{i,j}\right)$). *Assume that 3.1 holds. Then, for any $(i,j) \in I_m(k_w)$,*

$$\text{Var}\left(P_{i,j}\right) \geq \frac{\sigma^4}{k_s^4} \cdot \Delta_{i,j}^2 .$$

Proof. On $[0, 1/25]$, both ψ_1 and ψ_2 are lower bounded by t^2 (Lemma D.4 and D.5). Therefore, according to Eq. (19),

$$\text{Var}\left(P_{i,j}\right) = \psi_2\left(\frac{\sigma^2}{k_s^2}\right) \cdot \sum_{\substack{(u,v),(u',v')\in C_{i,j} \\ (u,v)\neq(u',v')}} \delta_{u,v}\delta_{u',v'} + \psi_1\left(\frac{\sigma^2}{k_s^2}\right) \cdot \sum_{(u,v)\in C_{i,j}} \delta_{u,v}^2$$

$$\geq \frac{\sigma^4}{k_s^4} \cdot \left(\sum_{\substack{(u,v),(u',v')\in C_{i,j} \\ (u,v)\neq(u',v')}} \delta_{u,v}\delta_{u',v'} + \sum_{(u,v)\in C_{i,j}} \delta_{u,v}^2 \right)$$

$$= \frac{\sigma^4}{k_s^4} \cdot \Delta_{i,j}^2 .$$

$\square$

We also have an upper bound on the variance of $P_{i,j}$.

Lemma A.4 (Upper bound on $\text{Var}\left(P_{i,j}\right)$). *Assume that 3.1 holds. In addition, suppose that $k_s \geq 5$. Then, for any $(i,j) \in I_m(k_w)$,*

$$\text{Var}\left(P_{i,j}\right) \leq 107\sigma^4 .$$

Proof. Since $\sigma^2 \leqslant k_s^2/25$, according to Lemma D.4, we have

$$\mathrm{Var}\,(X_{u,v}) \leqslant \frac{6\sigma^4}{k_s^4} \cdot \delta_{u,v}^2 \,,$$

and according to Lemma D.5,

$$\mathrm{Cov}(X_{u,v}, X_{u',v'}) \leqslant \frac{2\sigma^4}{k_s^4} \cdot \delta_{u,v}\delta_{u',v'} \,.$$

Therefore,

$$\mathrm{Var}\,(P_{i,j}) \leqslant \frac{6\sigma^4}{k_s^4} \sum \delta_{u,v}^2 + \frac{2\sigma^4}{k_s^4} \sum \delta_{u,v}\delta_{u',v'} = \frac{2\sigma^4}{k_s^4} \Delta_{i,j}^2 + \frac{4\sigma^4}{k_s^4} \sum \delta_{u,v}^2 \,.$$

Now, we use Lemma D.2 to bound $\Delta_{i,j}$ and Eq. (31) to bound the sum of squares. We obtain

$$\mathrm{Var}\,(P_{i,j}) \leqslant \frac{2\sigma^4}{k_s^4} \cdot \frac{(5k_s + 2)^4}{2^4} + \frac{4\sigma^4}{k_s^4} \cdot \frac{(9k_s + 5)^2}{25} \,.$$

We conclude by studying the last display as a function of k_s on $[5, +\infty)$. $\qquad\square$

Remark A.1. The main message of this section is that one can compute the moments of $X_{u,v}$ (and thus $P_{i,j}$) under parametric assumptions on the pixel values. That the distribution of the noise is Gaussian is not crucial, the same computations could be made with another p.d.f., leading to different expressions for the moments and thus the variance.

A.2   Hájek projection of the density estimates

In this section, we study the Hájek projection of $P_{i,j}$ under Assumption 3.1. We refer to Chapter 11 in Van der Vaart (2000) for an introduction to Hájek projections. We start by the derivation of the projection itself, which is given without proof in the paper as Eq. (6).

Proposition A.1 (Hajek projection of the density estimates). *Under Assumption 3.1, the Hajek projection of $P_{i,j}$ onto the set of random variables $\{\xi_{i,j}, (i,j) \in I\}$ is given by*

$$\forall (i,j) \in I, \quad \widehat{P}_{i,j} = C_1 \cdot \exp\left(\frac{-\|\xi_{i,j} - c\|^2}{2(k_s^2 + \sigma^2)}\right) \cdot \Delta_{i,j} + C_1 \cdot \sum_{\substack{(u,v)\in C_{i,j} \\ (u,v)\neq(i,j)}} \left[\exp\left(\frac{-\|\xi_{u,v} - c\|^2}{2(k_s^2 + \sigma^2)}\right) - \left(\frac{k_s^2 + \sigma^2}{k_s^2 + 2\sigma^2}\right)^{\frac{3}{2}}\right] \cdot \delta_{u,v} \,.$$

Proof. According to Lemma 11.10 in Van der Vaart (2000),

$$\widehat{P}_{i,j} = \mathbb{E}\,[P_{i,j}] + \sum_{(u,v)\in C_{i,j}} \left(\mathbb{E}\,[P_{i,j}|\xi_{u,v}] - \mathbb{E}\,[P_{i,j}]\right), \tag{20}$$

since the $\xi_{u,v}$ with $(u,v) \in C_{i,j}$ are the only random variables from which $P_{i,j}$ depends. By linearity, computing $\widehat{P}_{i,j}$ is thus a matter of computing $\mathbb{E}\,[P_{i,j}|\xi_{u,v}]$, for all $(u,v) \in C_{i,j}$.

If $(u,v) = (i,j)$, then Lemma A.1 with $p = 1$ gives us

$$\mathbb{E}\,[P_{i,j}|\xi_{i,j}] = C_1 \cdot \sum_{(u,v)\in C_{i,j}} \exp\left(\frac{-\|\xi_{i,j} - c\|^2}{2(k_s^2 + \sigma^2)}\right) \cdot \delta_{u,v} = C_1 \cdot \exp\left(\frac{-\|\xi_{i,j} - c\|^2}{2(k_s^2 + \sigma^2)}\right) \cdot \Delta_{i,j} \,.$$

Let us now assume that $(u,v) \neq (i,j)$. By linearity of the conditional expectation, the main computation is thus $\mathbb{E}\,[X_{u',v'}|\xi_{u,v}]$. There are two cases. First, $(u,v) = (u',v')$. Then similarly to the first part of the proof, we obtain

$$\mathbb{E}\,[X_{u,v}|\xi_{u,v}] = C_1 \cdot \exp\left(\frac{-\|\xi_{u,v} - c\|^2}{2(k_s^2 + \sigma^2)}\right) \cdot \delta_{u,v} \,.$$

Second, if $(u, v) \neq (u', v')$, then, by independence, $\mathbb{E}\left[X_{u',v'} | \xi_{u,v}\right] = \mathbb{E}\left[X_{u',v'}\right]$. Keeping in mind that $X_{i,j} = 1$ a.s., we deduce that

$$\mathbb{E}\left[P_{i,j} | X_{u,v}\right] = \sum_{(u',v') \in C_{i,j}} \mathbb{E}\left[X_{u',v'} | X_{u,v}\right]$$

$$= 1 + \mathbb{E}\left[X_{u,v} | \xi_{u,v}\right] + \sum_{\substack{(u',v') \in C_{i,j} \\ (u',v') \neq (i,j) \\ (u',v') \neq (u,v)}} \mathbb{E}\left[X_{u',v'}\right].$$

Therefore

$$\mathbb{E}\left[P_{i,j} | X_{u,v}\right] - \mathbb{E}\left[P_{i,j}\right] = C_1 \cdot \left[\exp\left(\frac{-\|\xi_{u,v} - c\|^2}{2(k_s^2 + \sigma^2)}\right) - \left(\frac{k_s^2 + \sigma^2}{k_s^2 + 2\sigma^2}\right)^{\frac{3}{2}}\right] \cdot \delta_{u,v}.$$

We conclude the proof by simplifying the $\mathbb{E}\left[P_{i,j}\right]$ in Eq. (20) with the one in the $(u, v) = (i, j)$ term. $\square$

Next, we will show that the variance ratio $\mathrm{Var}(\widehat{P}_{i,j}) / \mathrm{Var}(P_{i,j})$ is close to 1. We first derive the variance of $\widehat{P}_{i,j}$.

Lemma A.5 (Variance of the Hajek projection). *Under Assumption 3.1,*

$$\mathrm{Var}(\widehat{P}_{i,j}) = \psi_2\left(\frac{\sigma^2}{k_s^2}\right) \cdot \left(\Delta_{i,j}^2 + \sum_{\substack{(u,v) \in C_{i,j} \\ (u,v) \neq (i,j)}} \delta_{u,v}^2\right)$$

$$= 2\psi_2\left(\frac{\sigma^2}{k_s^2}\right) \cdot \sum_{(u,v) \in C_{i,j}} \delta_{u,v}^2 + \psi_2\left(\frac{\sigma^2}{k_s^2}\right) \cdot \sum_{\substack{(u,v),(u',v') \in C_{i,j} \\ (u,v) \neq (u',v')}} \delta_{u,v}\delta_{u',v'} + \psi_2\left(\frac{\sigma^2}{k_s^2}\right).$$

Proof. By construction, the Hájek projection is a sum of independent variables, thus the main computation is that of the variance of the exponential term. Using Lemma A.1, we see that

$$\mathrm{Var}\left(\exp\left(\frac{-\|\xi_{i,j} - c\|^2}{2(k_s^2 + \sigma^2)}\right)\right) = \left(\frac{k_s^2 + \sigma^2}{k_s^2 + 3\sigma^2}\right)^{\frac{3}{2}} - \left(\frac{k_s^2 + \sigma^2}{k_s^2 + 2\sigma^2}\right)^3.$$

Multiplying the previous display by C_1^2, we recognize $\psi_2(\sigma^2/k_s^2)$. Simple algebra yields the result, keeping in mind the definition of $\Delta_{i,j}$ and that $\delta_{i,j} = 1$. $\square$

A.3 $P_{i,j}$ is close to $\widehat{P}_{i,j}$

In this section, we show that $P_{i,j}$ is close to $\widehat{P}_{i,j}$ with high probability. The main difficulty here is to prove that the variance ratio is close to 1, which is achieved by the next proposition.

Proposition A.2 (Controlling the variance ratio). *Assume that 3.1 holds. Assume further that $k_s \geq 5$. Then, for any $(i, j) \in I_m(k_w)$,*

$$\left|\frac{\mathrm{Var}(\widehat{P}_{i,j})}{\mathrm{Var}(P_{i,j})} - 1\right| \leq \frac{4}{5k_s^2}.$$

Proof. Using Lemma A.5 and Eq. (19), we first write

$$\left|\mathrm{Var}(\widehat{P}_{i,j}) - \mathrm{Var}(P_{i,j})\right| \leq \left|2\psi_2\left(\frac{\sigma^2}{k_s^2}\right) - \psi_1\left(\frac{\sigma^2}{k_s^2}\right)\right| \cdot \sum_{(u,v) \in C_{i,j}} \delta_{u,v}^2 + \psi_2\left(\frac{\sigma^2}{k_s^2}\right)$$

$$\leq \frac{3\sigma^4}{k_s^4} \cdot \sum_{(u,v) \in C_{i,j}} \delta_{u,v}^2 + \frac{3\sigma^4}{2k_s^4},$$

where we used the fact that $|2\psi_2(t) - \psi_1(t)| \leq 3t^2$ and $\psi_2(t) \leq 3t^2/2$ for all $t \in [0, 1/25]$ (a consequence of Lemma D.6 and D.5).

Now recall that, according to Lemma A.3, $\mathrm{Var}\,(P_{i,j}) \geqslant \frac{\sigma^4}{k_s^4}\Delta_{i,j}^2$. We deduce that

$$
\begin{aligned}
\left| \frac{\mathrm{Var}(\widehat{P}_{i,j})}{\mathrm{Var}\,(P_{i,j})} - 1 \right| &\leqslant \frac{3\sigma^4}{k_s^4} \frac{k_s^4 \sum_{(u,v)\in C_{i,j}} \delta_{u,v}^2}{\sigma^4 \Delta_{i,j}^2} + \frac{3\sigma^4}{2k_s^4} \cdot \frac{k_s^4}{\sigma^4 \Delta_{i,j}^2} \\
&\leqslant \frac{3}{4k_s^2} + \frac{3}{2} \cdot \frac{1}{(2k_s+1)^4} \,. \qquad \text{(Lemma D.3 and D.2)}
\end{aligned}
$$

We deduce the result by studying the last display as a function of k_s on $[5, +\infty)$. $\qquad \square$

Transferring this control on the variance ratio to the random variable and its projection is a classical idea when dealing with Hájek projections:

Corollary A.1 ($P_{i,j}$ and $\widehat{P}_{i,j}$ are close, in probability). *Assume that 3.1 holds. Assume further that $k_s \geqslant 5$. Let $\varepsilon > 0$. Then, for any fixed $(i,j) \in I_m(k_w)$,*

$$
\mathbb{P}\left(\left| \frac{P_{i,j} - \mathbb{E}\,[P_{i,j}]}{\sqrt{\mathrm{Var}\,(P_{i,j})}} - \frac{\widehat{P}_{i,j} - \mathbb{E}[\widehat{P}_{i,j}]}{\sqrt{\mathrm{Var}(\widehat{P}_{i,j})}} \right| > \varepsilon \right) \leqslant \frac{1}{k_s^2 \varepsilon^2} \,.
$$

Proof. The assumptions of Proposition A.2 are satisfied, therefore

$$
\left| \frac{\mathrm{Var}(\widehat{P}_{i,j})}{\mathrm{Var}\,(P_{i,j})} - 1 \right| \leqslant \frac{4}{5k_s^2} \,,
$$

which is smaller than $1/30$ by our choice of k_s. Since $\sqrt{\cdot}$ is $3/5$-Lipschitz on $[1 - 1/30, 1 + 1/30]$, it holds that

$$
\left| \sqrt{\frac{\mathrm{Var}(\widehat{P}_{i,j})}{\mathrm{Var}\,(P_{i,j})}} - 1 \right| \leqslant \frac{3}{5} \cdot \frac{4}{5k_s^2} \leqslant \frac{1}{2k_s^2} \,. \tag{21}
$$

A careful reading of the proof of Theorem 11.2 in Van der Vaart (2000) reveals that a factor 2 appears, which concludes the proof. $\qquad \square$

A.4   Main term

We now show that $R_{i,j} = \widehat{P}_{i,j} - Q_{i,j}$ is negligible in probability when σ^2 is small. More precisely, we derive a variance bound for $R_{i,j}$.

Lemma A.6 ($R_{i,j}$ is negligible). *Assume that 4.1 holds. Assume further that $k_s \geqslant 5$. Then*

$$
\mathrm{Var}\,(R_{i,j}) \leqslant \frac{\sigma^2}{4} \,.
$$

Proof. It is clear that $\mathbb{E}\,[R_{i,j}] = 0$ and from the proof of Lemma A.5, we see that

$$
\begin{aligned}
\mathrm{Var}\,(R_{i,j}) &= \psi_2\left(\frac{\sigma^2}{k_s^2} \right) \cdot \sum_{\substack{(u,v)\in C_{i,j} \\ (u,v)\neq(i,j)}} \delta_{u,v}^2 \\
&\leqslant \frac{3\sigma^4}{2k_s^4} \cdot \frac{(9k_s+5)^2}{25} \,. \qquad \text{(Eq. (31))}
\end{aligned}
$$

Since $\sigma^2 \leqslant k_s^2/25$, we see that

$$
\mathrm{Var}\,(R_{i,j}) \leqslant \frac{3(9k_s+5)^2}{2 \cdot 25 \cdot 25 k_s^2} \cdot \sigma^2 \,.
$$

We conclude by studying the previous display as a function of k_s on $[5, +\infty)$. $\qquad \square$

To conclude this section, we now prove Theorem 3.1 of the paper, which is re-stated here for completeness' sake.

Theorem A.1 ($P_{i,j}$ **is close to** $Q_{i,j}$, **with high probability**). *Assume that 3.1 holds. Suppose furthermore that $k_s \geqslant 5$. Let $(i,j) \in I_m(k_w)$. Then, for any $\varepsilon > 0$,*

$$\mathbb{P}\left(|P_{i,j} - Q_{i,j}| > \varepsilon\right) \leqslant \frac{71\sigma^2}{\varepsilon^2} \,.$$

Proof. We first write

$$\mathbb{P}\left(|P_{i,j} - Q_{i,j}| > \varepsilon\right) \leqslant \mathbb{P}(|P_{i,j} - \widehat{P}_{i,j}| > \varepsilon/2) + \mathbb{P}\left(|R_{i,j}| > \varepsilon/2\right) \,. \tag{22}$$

Let us focus on the second term, and notice that $\mathbb{E}\left[R_{i,j}\right] = 0$. Since the assumptions of Lemma A.6 are satisfied, we know that $\mathrm{Var}\left(R_{i,j}\right) \leqslant \sigma^2/4$. Therefore, by Chebyshev's inequality,

$$\mathbb{P}\left(|R_{i,j}| > \varepsilon/2\right) \leqslant \frac{2^2\sigma^2}{4\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2} \,.$$

Now we turn back to the first term in Eq. (22). Noting that $\mathbb{E}\left[P_{i,j}\right] = \mathbb{E}[\widehat{P}_{i,j}]$, we have

$$\mathbb{P}(|P_{i,j} - \widehat{P}_{i,j}| > \varepsilon/2) = \mathbb{P}\left(\left|\frac{P_{i,j} - \mathbb{E}\left[P_{i,j}\right]}{\sqrt{\mathrm{Var}(\widehat{P}_{i,j})}} - \frac{\widehat{P}_{i,j} - \mathbb{E}\left[\widehat{P}_{i,j}\right]}{\sqrt{\mathrm{Var}(\widehat{P}_{i,j})}}\right| > \frac{\varepsilon}{2\sqrt{\mathrm{Var}(\widehat{P}_{i,j})}}\right)$$

$$\leqslant \mathbb{P}\left(\left|\frac{P_{i,j} - \mathbb{E}\left[P_{i,j}\right]}{\sqrt{\mathrm{Var}\left(P_{i,j}\right)}} - \frac{\widehat{P}_{i,j} - \mathbb{E}\left[\widehat{P}_{i,j}\right]}{\sqrt{\mathrm{Var}(\widehat{P}_{i,j})}}\right| > \frac{\varepsilon}{4\sqrt{\mathrm{Var}(\widehat{P}_{i,j})}}\right) \tag{23}$$

$$+ \mathbb{P}\left(|P_{i,j} - \mathbb{E}\left[P_{i,j}\right]| \cdot \left|\frac{1}{\sqrt{\mathrm{Var}\left(P_{i,j}\right)}} - \frac{1}{\sqrt{\mathrm{Var}(\widehat{P}_{i,j})}}\right| > \frac{\varepsilon}{4\sqrt{\mathrm{Var}(\widehat{P}_{i,j})}}\right) \,. \tag{24}$$

According to Corollary A.1, Eq. (23) is upper bounded by

$$\frac{16\,\mathrm{Var}(\widehat{P}_{i,j})}{k_s^2\varepsilon^2} \leqslant \frac{16\,\mathrm{Var}\left(P_{i,j}\right)}{k_s^2\varepsilon^2} \leqslant \frac{16 \cdot 107 \cdot \sigma^4}{k_s^2\varepsilon^2} \leqslant \frac{69\sigma^2}{\varepsilon^2} \,,$$

where we used, successively, the fact that projection reduces variance, Lemma A.4, and $\sigma^2 \leqslant k_s^2/25$. Finally, we rewrite Eq. (24) as

$$\mathbb{P}\left(|P_{i,j} - \mathbb{E}\left[P_{i,j}\right]| > \frac{\varepsilon}{4\left|\sqrt{\frac{\mathrm{Var}(\widehat{P}_{i,j})}{\mathrm{Var}(P_{i,j})}} - 1\right|}\right) \leqslant \frac{16\,\mathrm{Var}\left(P_{i,j}\right)}{\varepsilon^2} \cdot \left|\sqrt{\frac{\mathrm{Var}(\widehat{P}_{i,j})}{\mathrm{Var}\left(P_{i,j}\right)}} - 1\right|^2 \qquad \text{(Chebyshev)}$$

$$\leqslant \frac{16 \cdot 107 \cdot \sigma^4}{\varepsilon^2} \cdot \frac{1}{4k_s^4} \qquad \text{(Lemma A.4 and Eq. (21))}$$

$$\leqslant \frac{16 \cdot 107 \cdot \sigma^2}{4\varepsilon^2 \cdot 25 \cdot 25} \qquad (\sigma^2 \leqslant k_s^2/25 \text{ and } k_s \geqslant 5)$$

$$\leqslant \frac{\sigma^2}{\varepsilon^2} \,.$$

We conclude by summing all bounds. $\qquad\qquad\square$

Remark A.2. Here, trying to obtain a uniform bound, that is, a meaningful upper bound for $\mathbb{P}\left(\|P - Q\|_\infty > \varepsilon\right)$ seems challenging. Indeed, since the typical deviations of P are of order σ, we need ε to be of order σ for Theorem 3.1 to be useful. If we were to extend the bound given by Theorem 3.1 by a union bound argument, a factor $H \cdot W$ would appear.

B HOMOGENEOUS PATCHES

In this section, we provide a detailed proof of Theorem 3.2 of the paper. We start by providing a proof of Lemma 3.1 of the paper in Section B.1. We follow up by some area computations in Section B.2. In Section B.3, we show that the number of lattice points inside a square, a rounded square, or a disk, is close to the area of the geometric form. We conclude the proof in Section B.4.

B.1 Proof of Lemma 3.1

The main goal of this section is to prove the following (Lemma 3.1 in the paper).

Lemma B.1 (Connected components and local maxima). *Let $\mathcal{G}(A)$ be the directed graph produced by Algorithm 3 applied to an array $A \in \mathbb{R}^{H \times W}$. Then to each connected component of $\mathcal{G}(A)$ corresponds a unique local maxima of A in the sense of Definition 3.1.*

Proof. Let $\mathcal{C}$ be a connected component of $\mathcal{G}(A)$. We split the proof in existence and uniqueness of the local maximum.

Existence. Let us pick any point in $\mathcal{C}$. If this point has no outgoing edge, then it is a local maximum by definition. Otherwise, we follow outgoing edges until we meet a vertex that has no outgoing edge. This always happens since there is a finite number of points in $\mathcal{C}$ and loops are prohibited by construction of the outgoing edges.

Uniqueness. Let us now suppose that there are two distinct local maxima in $\mathcal{C}$, say (i_1, j_1) and (i_2, j_2). Since $\mathcal{C}$ is a connected component of $\mathcal{G}(A)$, there is a path between (i_1, j_1) and (i_2, j_2). This path cannot have directed edges flowing out of (i_1, j_1) (resp. (i_2, j_2)): by construction of $\mathcal{G}(A)$, it would mean that there exists a vertex (i', j') inside E_{i_1, j_1} (resp. E_{i_2, j_2}) with higher value. Thus the path connecting (i_1, j_1) and (i_2, j_2) has directed edges going towards them. Let us call (i'_1, j'_1) and (i'_2, j'_2) the vertices at the origin of these edges. Of course, one has $(i'_1, j'_1) \neq (i'_2, j'_2)$, since a given vertex has only one outgoing edge by construction of the graph. Therefore, we can repeat the reasoning above and find two distinct vertices with edges flowing towards (i'_1, j'_1) and (i'_2, j'_2). This is absurd, since there is only a finite number of vertices in $\mathcal{C}$. Therefore the local maximum is unique. $\qquad\square$

B.2 Area computations

Computing the area of a square of half side k_w or of a disk of radius d_m is straightforward: we obtain $4k_w^2$ and πd_m^2, respectively. The only difficulty is computing the area of a rounded square, the case of interest when $k_w \leqslant d_m \leqslant \sqrt{2}k_w$, which we will assume in this section. We begin by the computation of a *circle segment* (see Figure 10).

Lemma B.2 (Area of a circle segment). *Let d_m and k_w be two real numbers such that $k_w < d_m < \sqrt{2}k_w$. Then the area of the circle segment is given by $\gamma(k_w, d_m)$, with*

$$\forall s, d, \qquad \gamma(s, d) := d^2 \arctan \frac{\sqrt{d^2 - s^2}}{s} - s\sqrt{d^2 - s^2}\,.$$

Before proving Lemma B.2, let us note that a direct consequence is that we can deduce that the area of the rounded square. Let us define

$$B(s, d) := \pi d^2 - 4\gamma(s, d) = \pi d^2 - 4d^2 \arctan \frac{\sqrt{d^2 - s^2}}{s} + 4s\sqrt{d^2 - s^2}\,, \tag{25}$$

that is, the area of a disk of radius d to which we subtract four times the area of the circle segment of parameters (s, d). Then the area of the rounded square of parameters (k_w, d_m) is given by $B(k_w, d_m)$. Notice that we recover the limit cases: when $d_m = k_s$, $B(k_w, d_m) = \pi d_m^2$, and when $d_m = \sqrt{2}k_s$, $B(k_w, d_m) = 4k_s^2$, as expected.

Proof. The area of the circle segment is given by the difference between the area of the angular sector OAB and the area of the triangle OAB. Let us recall that $OC = s$ and $OA = d$. According to Pythagoras theorem, $AC = \sqrt{d^2 - s^2}$, and therefore $\tan \theta = \frac{\sqrt{d^2 - s^2}}{s}$. We deduce that the area of the angular sector is given by $d^2 \arctan \frac{\sqrt{d^2 - s^2}}{s}$. Finally, the area of the triangle is $AB \times OC/2$, that is, $s\sqrt{d^2 - s^2}$. $\qquad\square$
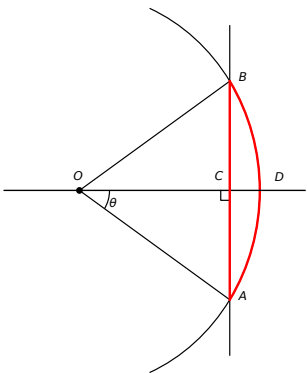
Figure 10: A circle segment with parameters (s, d) is outlined in red, obtained by intersecting of the disk of radius $OD = d$ centered in O and the square of half-side $OC = s$. The area of the portion of the plane enclosed by the red curve is given by Lemma B.2.
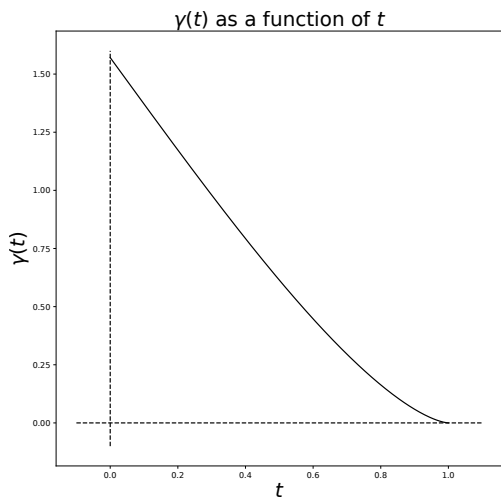

Figure 11: Plot of the γ function, the normalized version of the γ function derived in Lemma B.2.

B.3   Counting lattice points

The main goal of this section is to prove a binding lemma: in order to count the number of lattice points inside on of the three geometrical shapes, it is sufficient to compute the area of these figures, if one is ready to loose $\mathcal{O}\left(k_w + d_m\right)$ terms. The idea of the proof is rather simple: show that the area expressions are Lipschitz with respect to k_w and d_m, and then use the idea of Gauss historical bound for the Gauss circle problem. As in Section B.2, the only challenging case is that of the rounded square. We thus focus our attention on this case for now on, and start by studying γ more precisely.

Lemma B.3 (γ is Lipschitz in each coordinate). *Let s and d be fixed numbers such that $s < d < \sqrt{2}s$. Then the function $\gamma(\cdot, d)$ is d-Lipschitz on $[0, d]$ and the function $\gamma(s, \cdot)$ is $3s$-Lipschitz on $[s, \sqrt{2}s]$.*

Proof. Let us first consider the case where d is fixed. We can write

$$\gamma(ds, d) = d^2 \arctan \frac{\sqrt{d^2 - d^2 s^2}}{ds} - ds \cdot \sqrt{d^2 - d^2 s^2}$$

$$= d^2 \left[\arctan \frac{\sqrt{1 - s^2}}{s} - s\sqrt{1 - s^2} \right]$$

$$\gamma(ds, d) = d^2 \gamma(s) \,,$$

where we let $\gamma(\cdot)$ denote $\gamma(\cdot, 1)$. As seen on Figure 11, $\gamma(\cdot)$ is well-behaved. More precisely, one can check that $\gamma'(t) = -2t\sqrt{1 - t^2}$. The maximum of $|\gamma'(t)|$ on $[0, 1]$ is 1, attained at $t = \sqrt{2}/2$. As a consequence, γ is 1-Lipschitz. For any h small enough, we write

$$|\gamma(s, d) - \gamma(s + h, d)| = d^2 \left| \gamma\left(\frac{s + h}{d}\right) - \gamma\left(\frac{s}{d}\right) \right|$$

$$\leqslant d^2 \left| \frac{s + h}{d} - \frac{s}{d} \right| = dh \,,$$

where we used the 1-Lipschitzness of γ in the inequality. We deduce that $\gamma(\cdot, d)$ is d-Lipschitz.

Next let us consider that s is fixed. We directly compute the partial derivative of γ with respect to d and we obtain $2d \arctan \frac{\sqrt{d^2 - s^2}}{s}$. It is an increasing function of d, thus taking its maximum at the rightmost possible value for d, which is $d = \sqrt{2}s$. The extreme value is $2\sqrt{2}s \arctan 1 = \frac{\pi \sqrt{2}}{2} s (\approx 2.22s)$, and we can deduce the result. $\qquad\square$

We can now state and prove the main result of this section:

Proposition B.1 (Counting lattice points). *Let s and d be positive real numbers. Then*

- *if $d < s$, the number of lattice points inside a disk of radius d is given by $\pi d^2 + \mathcal{O}(d)$;*

- *if $s < d < \sqrt{2}s$, the number of lattice points inside a rounded square of parameters (s, d) is given by $B(s, d) + \mathcal{O}(s)$;*

- *if $\sqrt{2}s < d$, the number of lattice points inside a square of half side s is given by $4s^2 + \mathcal{O}(s)$.*

Proof. We begin by the first case, which is known as the *Gauss circle problem* in the literature. Let $n(d)$ denote the number of lattice points inside the disk of radius d. For each lattice point inside the disk, we can draw a square of side 1 centered at the point. These squares are non-overlapping, and there total area is less than that of a disk of radius $d + \sqrt{2}/2$: the limit case is that of a lattice point lying exactly on the boundary of the disk. In the same fashion, the total area cannot be less than that of a disk of radius $d - \sqrt{2}/2$ (if $d \leqslant \sqrt{2}/2$, then there is no need to consider this case). Since the total area coincide with the number of lattice points inside the disk, we have obtained the following bound:

$$\pi(d - \sqrt{2}/2)^2 \leqslant n(d) \leqslant \pi(d + \sqrt{2}/2)^2 \,.$$

From this last display, we immediately deduce that $\left| n(d) - \pi d^2 \right| = \mathcal{O}(d)$. This line of proof is actually the historical one, proposed by Gauss himself (Hardy, 1959). We directly extend this reasoning to the square by considering an inner square of side $2s - 1$ and an outer square of side $2s + 1$.

The rounded square case is slightly more involved: essentially, one has to look at an inner rounded square of parameters $(s - 1/2, d - \sqrt{2}/2)$ and an outer rounded square of parameters $(s + 1/2, d + \sqrt{2}/2)$. Controlling the error amounts to bounding $B(s, d) - B(s - 1/2, d - \sqrt{2}/2)$ (the outer case is similar). This is where Lemma B.3 comes into play: we write

$$\left| B(s, d) - B(s - 1/2, d - \sqrt{2}/2) \right| = \left| B(s, d) - B(s - 1/2, d) + B(s - 1/2, d) - B(s - 1/2, d - \sqrt{2}/2) \right|$$

$$\leqslant |B(s, d) - B(s - 1/2, d)| + \left| B(s - 1/2, d) - B(s - 1/2, d - \sqrt{2}/2) \right|$$

$$\leqslant \frac{d}{2} + 4s^2 - (2s - 1)^2 + \frac{3s\sqrt{2}}{2} = \mathcal{O}(s) \,,$$

where we used the (bi-)Lipschitzness of γ in the second inequality. $\qquad\square$

Remark B.1. Let us call $E(d)$ the difference between $n(d)$, the number of lattice points inside a disk of radius d, and πd^2, the area of that disk. Following the classical argument, we showed that $|E(d)| \leqslant 2\sqrt{2}\pi d$. This is not the best bound, which is currently $|E(d)| = \mathcal{O}\left(d^{\frac{131}{208}}\right)$ (Huxley, 2000). Using this bound would improve the error made in Theorem 3.2 by considering πd_m^2 instead of $N_{i,j}$, but additional work is required to extend the argument to the rounded square.

B.4 Average number of local maxima

We are now able to prove Theorem 3.2 of the paper. Let us recall the statement of Theorem 3.2:

Theorem B.1 (Average number of local maxima). *Assume that 3.1 holds. Let $R \subseteq I$ be a rectangle of height h and width w at distance greater than $2k_w$ from the border. Then*

$$
\mathbb{E}\left[N_R(Q)\right] = \begin{cases} hw \cdot \left(\frac{1}{\pi d_m^2} + \mathcal{O}\left(\frac{1}{d_m}\right)\right) & \text{if } d_m \leqslant k_w \\ hw \cdot \left(\frac{1}{\pi(3k_w d_m - k_w^2 - d_m^2)} + \mathcal{O}\left(\frac{1}{k_w}\right)\right) & \text{if } k_w < d_m \leqslant \sqrt{2}k_w \\ hw \cdot \left(\frac{1}{4k_w^2} + \mathcal{O}\left(\frac{1}{k_w}\right)\right) & \text{otherwise.} \end{cases}
$$

Proof. Let us first focus on the disk case, that is, $d_m \leqslant k_w$. According to Lemma 3.2 of the paper,

$$
\mathbb{E}\left[N_R(Q)\right] = \sum_{(i,j)\in R} \frac{1}{N_{i,j}} \,. \tag{26}
$$

Since R is at distance greater than $2k_w > d_m$ from the border of the image, $E_{i,j}$ is not intersecting the boundaries of the image, and in particular $N_{i,j}$ is constant. Moreover, according to Proposition B.1, we know that $N_{i,j} = \pi d_m^2 + \mathcal{O}(d_m)$. We deduce the result since there are hw terms in the sum in Eq. (26), and since

$$
\frac{1}{N_{i,j}} = \frac{1}{\pi d_m^2 + \mathcal{O}(d_m)} = \frac{1}{\pi d_m^2(1 + \mathcal{O}(1/d))} = \frac{1}{\pi d_m^2}\left(1 + \mathcal{O}\left(\frac{1}{d_m}\right)\right) \,.
$$

The square case is similar. Finally, in the rounded square case, the only difference is that we replaced $B(k_w, d_m)$ by the more readable $\pi(3k_w d_m - k_w^2 - d_m^2)$. This is justified by Lemma D.9. Indeed,

$$
\left| \frac{1}{B(k_w, d_m)} - \frac{1}{\pi(3k_w d_m - k_w^2 - d_m^2)} \right| = \frac{\left| B(k_w, d_m) - \pi(3k_w d_m - k_w^2 - d_m^2) \right|}{B(k_w, d_m) \cdot \pi(3k_w d_m - k_w^2 - d_m^2)} \leqslant \frac{0.04 d_m^2}{\pi\left(k_w^2\right)^2} \,,
$$

since a disk of radius k_w is always included in the intersection in this configuration. Since $k_w \geqslant d_m/\sqrt{2}$, we deduce that

$$
\left| \frac{1}{B(k_w, d_m)} - \frac{1}{\pi(3k_w d_m - k_w^2 - d_m^2)} \right| = \mathcal{O}\left(\frac{1}{d_m^3}\right) \,,
$$

and the leading term in the approximation comes from replacing $N_{i,j}$ by B. $\qquad\square$

Remark B.2. The main reason for substituting $\pi(3k_w d_m - k_w^2 - d_m^2)$ to $B(k_w, d_m)$ is the clarity of the exposition. Both expressions are 2-homogeneous in (k_w, d_m) and we could have stated Theorem 3.2 of the paper with B at the denominator in the second case.

C SHARP BOUNDARIES

In this section, we prove Theorem 4.1 of the paper, which is true in the bicolor setting (Assumption 4.1). The organization of this section follows the sketch of the proof: in Section C.1 , we compute the expected value of $P_{i,j}$ under Assumption 4.1, and the difference $P_{i,j+1} - P_{i,j}$. In Section C.2, we study the variance of $P_{i,j}$ and show that the same bound holds. We conclude in Section C.3.

C.1 Expectation computation

The expected value of $P_{i,j}$ under Assumption 4.1 differs notably from the one computed under Assumption 3.1 (in Section A.1 of this Appendix).

Lemma C.1 (Expected density, bi-color setting). *Assume that 4.1 holds. Then, for all $(i,j) \in I_\ell$, we have*

$$\mathbb{E}\left[P_{i,j}\right] = C_2 \cdot \left[\sum_{(u,v) \in C_{i,j} \cap I_\ell} \delta_{u,v} + \exp\left(\frac{- \|c_1 - c_2\|^2}{2(k_s^2 + 2\sigma^2)} \right) \cdot \sum_{(u,v) \in C_{i,j} \cap I_r} \delta_{u,v} \right] .$$

Proof. Straightforward from Lemma A.1. $\qquad\square$

We now show that the expected density decreases near the boundary, provided that the color change is large enough.

Lemma C.2 (Expected density decreases near the boundary). *Assume that 4.1 holds. Assume further that $\|c_1 - c_2\| \geqslant 3k_s$ and $k_s \geqslant 5$. Then, for any $(i,j) \in I_\ell$,*

$$\mathbb{E}\left[P_{i,j}\right] - \mathbb{E}\left[P_{i,j+1}\right] \geqslant \frac{3k_s}{2} .$$

Proof. We start by noticing that another way to read Lemma C.1 is

$$\mathbb{E}\left[P_{i,j}\right] = C_2 \cdot \left[\Delta_{i,j} - \left(1 - \exp\left(\frac{- \|c_1 - c_2\|^2}{2(k_s^2 + 2\sigma^2)} \right) \right) \cdot \sum_{(u,v) \in C_{i,j} \cap I_r} \delta_{u,v} \right] .$$

Thus the difference that interests us here can be written

$$\mathbb{E}\left[P_{i,j}\right] - \mathbb{E}\left[P_{i,j+1}\right] = C_2 \cdot \left(1 - e^{\frac{-\|c_1 - c_2\|^2}{2(k_s^2 + 2\sigma^2)}} \right) \cdot \left[\sum_{(u,v) \in C_{i,j+1} \cap I_r} \delta_{u,v}^{i,j+1} - \sum_{(u,v) \in C_{i,j} \cap I_r} \delta_{u,v}^{i,j} \right] .$$

Since $\delta_{u,v}^{i,j+1} = \delta_{u,v-1}^{i,j}$, most of the terms cancel out in the right-hand side of the last display, and we are left with

$$\mathbb{E}\left[P_{i,j}\right] - \mathbb{E}\left[P_{i,j+1}\right] = C_2 \cdot \left(1 - e^{\frac{-\|c_1 - c_2\|^2}{2(k_s^2 + 2\sigma^2)}} \right) \cdot \sum_{u=i-k_w}^{i+k_w} \delta_{u,j_0}^{i,j} . \tag{27}$$

We now proceed to lower bound each term in Eq. (27). First, since we assumed $\sigma^2 \leqslant k_s^2/25$, it is easy to check that $C_2 \geqslant 4/5$. Next, using Eq. (30) (we assumed that $k_s \geqslant 5$), we see that

$$\sum_{u=i-k_w}^{i+k_w} \delta_{u,v_0}^{i,j} \geqslant 2 \cdot k_s + 1 \geqslant 2k_s ,$$

since the worst case is at the border, where $j = j_0$. The last term requires a bit more attention. We start by writing

$$\frac{\|c_1 - c_2\|^2}{2(k_s^2 + 2\sigma^2)} \geqslant \frac{9k_s^2}{2(k_s^2 + 2\sigma^2)} \qquad\qquad (\text{since } \|c_1 - c_2\| \geqslant 3k_s)$$

$$\geqslant \frac{9 \cdot 25}{2 \cdot 27} . \qquad\qquad (\text{since } \sigma^2 \leqslant k_s^2/25)$$

We deduce that

$$1 - e^{\frac{-\|c\|^2}{2(k_s^2 + 2\sigma^2)}} \geqslant 1 - e^{\frac{-9 \cdot 25}{2 \cdot 27}} \geqslant \frac{49}{50} .$$

Multiplying together the individual numerical bounds, we find the promised result. $\qquad\square$

C.2   Variance computation

We now prove that the variance of $P_{i,j}$ in the bicolor case is upper bounded by the variance in the homogeneous case.

Lemma C.3 (Variance of $P_{i,j}$, bicolor setting). *Assume that 4.1 holds. Assume further that $k_s \geqslant 5$. Then*

$$\mathrm{Var}\left(P_{i,j}\right) \leqslant 107\sigma^4 .$$

Remark C.1. There is no point in trying to derive a specific bound since the worst case is far from the border when $P_{i,j}$ coincides with its homogeneous version. In that event, both variance coincides. Lemma C.3 simply states that this is the worst case scenario.

Proof. When computing the variance of the density estimate, all terms are identical to the homogeneous case when looking at the individual variances of $X_{u,v}$ for $(u, v) \in C_{i,j} \cap I_\ell$. The covariance terms are also the same if both points lie in the same part of $C_{i,j}$.

Let us compute $\mathrm{Var}\left(X_{u,v}\right)$ when $(u, v) \in C_{i,j} \cap I_r$. Using Lemma A.1, we find that

$$\mathbb{E}\left[X_{u,v}\right] = \left(\frac{k_s^2}{k_s^2 + 2\sigma^2}\right)^{\frac{3}{2}} \cdot \exp\left(\frac{-\|c_1 - c_2\|^2}{2(k_s^2 + 2\sigma^2)}\right) \cdot \delta_{u,v} . \tag{28}$$

and

$$\mathbb{E}\left[X_{u,v}^2\right] = \left(\frac{k_s^2}{k_s^2 + 4\sigma^2}\right)^{\frac{3}{2}} \cdot \exp\left(\frac{-\|c_1 - c_2\|^2}{k_s^2 + 4\sigma^2}\right) \cdot \delta_{u,v}^2 .$$

Putting the last two displays together, we see that

$$\mathrm{Var}\left(X_{u,v}\right) = \left[\left(\frac{k_s^2}{k_s^2 + 4\sigma^2}\right)^{\frac{3}{2}} \cdot \mathrm{e}^{\frac{-\|c_1 - c\|^2}{k_s^2 + 4\sigma^2}} - \left(\frac{k_s^2}{k_s^2 + 2\sigma^2}\right)^3 \cdot \mathrm{e}^{\frac{-\|c_1 - c_2\|^2}{k_s^2 + 2\sigma^2}}\right] \cdot \delta_{u,v}^2 .$$

According to Lemma D.7, the left term is upper bounded by $\psi_1(\sigma^2/k_s^2)$, provided that $\|c\| \geqslant 3k_s$ and $\sigma^2 \leqslant k_s^2/25$, which we assumed.

Finally, let us look at covariance terms with point $(u, v) \in C_{i,j} \cap I_\ell$ and $(u', v') \in C_{i,j} \cap I_r$. In that case, we first write

$$\mathbb{E}\left[X_{u,v}X_{u',v'}|\xi_{i,j}\right] = \left(\frac{k_s^2}{k_s^2 + \sigma^2}\right)^3 \cdot \exp\left(\frac{-\|\xi_{i,j} - c_1\|^2}{2(k_s^2 + \sigma^2)}\right) \cdot \exp\left(\frac{-\|\xi_{i,j} - c_2\|^2}{2(k_s^2 + \sigma^2)}\right) ,$$

by independence. The key computation here is

$$\int \exp\left(\frac{-(x - c_{1,k})^2}{2(k_s^2 + \sigma^2)} + \frac{-(x - c_{2,k})^2}{2(k_s^2 + \sigma^2)} + \frac{-(x - c_{1,k})^2}{2\sigma^2}\right) \frac{\mathrm{d}x}{\sigma\sqrt{2\pi}} = \left(\frac{k_s^2 + \sigma^2}{k_s^2 + 3\sigma^2}\right)^{\frac{1}{2}} \cdot \exp\left(\frac{-(c_{1,k} - c_{2,k})^2}{k_s^2 + 3\sigma^2}\right) ,$$

for $k \in \{1, 2, 3\}$, since $(i, j) \in I_\ell$. In definitive, we obtain that

$$\mathbb{E}\left[X_{u,v}X_{u',v'}\right] = \frac{k_s^6}{(k_s^2 + \sigma^2)^{\frac{3}{2}}(k_s^2 + 3\sigma^2)^{\frac{3}{2}}} \cdot \exp\left(\frac{-\|c_1 - c_2\|^2}{k_s^2 + 3\sigma^2}\right) \cdot \delta_{u,v}\delta_{u',v'} .$$

Now, the expectation of $X_{u,v}$ is unchanged:

$$\mathbb{E}\left[X_{u,v}\right] = \left(\frac{k_s^2}{k_s^2 + 2\sigma^2}\right)^{\frac{3}{2}} \cdot \delta_{u,v} .$$

The expectation of $X_{u',v'}$ is as in Eq. (28). Putting all of this together, we have

$$\mathrm{Cov}(X_{u,v}, X_{u',v'}) = \left[\frac{k_s^6}{(k_s^2 + \sigma^2)^{\frac{3}{2}}(k_s^2 + 3\sigma^2)^{\frac{3}{2}}} \cdot \mathrm{e}^{\frac{-\|c_1 - c_2\|^2}{k_s^2 + 3\sigma^2}} - \left(\frac{k_s^2}{k_s^2 + 2\sigma^2}\right)^3 \cdot \mathrm{e}^{\frac{-\|c_1 - c_2\|^2}{2(k_s^2 + 2\sigma^2)}}\right] \cdot \delta_{u,v}\delta_{u',v'} .$$

In Lemma D.8, we show that the bracketed term is smaller than $\psi_2\left(\frac{\sigma^2}{k_s^2}\right)$. Thus $\mathrm{Var}\left(P_{i,j}\right)$ is smaller than in the unicolor case, and the bound is thus identical to that of Lemma A.4. $\square$

C.3 Putting everything together

We are now able to prove Theorem 4.1 of the paper, which we re-state here:

Theorem C.1 (Decreasing density estimates). *Assume that 4.1 holds. Assume further that $k_s \geqslant 5$ and that $\|c_1 - c_2\| \geqslant 3k_s$. Then, for any $(i,j) \in I_m \cap I_\ell$ such that $|j - j_0| \leqslant k_w$,*

$$\mathbb{P}\left(P_{i,j} > P_{i,j+1}\right) \geqslant 1 - 16\sigma^2 .$$

Proof. We first use the Chebyshev's inequality and Lemma C.3 to show that

$$\mathbb{P}\left(|P_{i,j} - \mathbb{E}\left[P_{i,j}\right]| \geqslant \frac{3k_s}{4}\right) \leqslant \frac{\operatorname{Var}\left(P_{i,j}\right)}{3^2 k_s^2 / 4^2} \leqslant \frac{107\sigma^4 \cdot 4^2}{3^2 k_s^2} .$$

Since $\sigma^2 \leqslant k_s^2/25$,

$$\mathbb{P}\left(|P_{i,j} - \mathbb{E}\left[P_{i,j}\right]| \geqslant \frac{107 \cdot 16 k_s}{9 \cdot 25}\right) \leqslant 8\sigma^2 .$$

We have the same bound for $P_{i,j+1}$. Noting that $\frac{3k_s}{4}$ is exactly half the minimal gap between both expectations given by Lemma C.2, we conclude by a union bound argument. $\qquad\square$

D   TECHNICAL RESULTS

We collect in this section technical lemmatas used throughout the proofs.

D.1   Expectation computations

In this section we collect technical facts related to Gaussian computations.

Lemma D.1 (Key Gaussian computation). *Let a, b be real numbers, and c, d be positive numbers. Then it holds that*

$$\int \exp\left(\frac{-(x-a)^2}{2c^2} + \frac{-(x-b)^2}{2d^2}\right) \frac{\mathrm{d}x}{d\sqrt{2\pi}} = \left(\frac{c^2}{c^2 + d^2}\right)^{\frac{1}{2}} \cdot \exp\left(\frac{-(a-b)^2}{2(c^2 + d^2)}\right) .$$

Proof. See for instance Lemma 11.1 in Garreau and von Luxburg (2020). $\qquad\square$

D.2   Deterministic part

In this section, we collect some technical facts about $\delta_{u,v}$ and $\Delta_{i,j}$.

Lemma D.2 (Bounding $\Delta_{i,j}$). *Assume that $k_s \geqslant 5$. Then, for all $(i,j) \in I_m(k_w)$,*

$$(2k_s + 1)^2 \leqslant \Delta_{i,j} \leqslant \frac{(5k_s + 2)^2}{4} .$$

Proof. By definition of $\Delta_{i,j}$, since $(i,j) \in I_m(k_w)$, we have

$$\Delta_{i,j} = \left(2 \sum_{u=1}^{k_w} \exp\left(\frac{-u^2}{2k_s^2}\right) + 1\right)^2 . \tag{29}$$

The key idea of the proof is a series-integral comparison of the sum appearing in the previous display. Since the mapping $u \mapsto \mathrm{e}^{-u^2/(2k_s^2)}$ is decreasing on $\mathbb{R}_+$, we write

$$\forall u \in \{1, \ldots, k_w\}, \qquad \int_u^{u+1} \mathrm{e}^{\frac{-t^2}{2k_s^2}} \, \mathrm{d}t \leqslant \exp\left(\frac{-u^2}{2k_s^2}\right) \leqslant \int_{u-1}^u \mathrm{e}^{\frac{-t^2}{2k_s^2}} \, \mathrm{d}t .$$

We then sum these inequalities over u. On one side,

$$\sum_{u=1}^{k_w} \exp\left(\frac{-u^2}{2k_s^2}\right) \geqslant \int_1^{k_w+1} e^{\frac{-t^2}{2k_s^2}} \, dt$$

$$= k_s \cdot \int_{1/k_s}^{\frac{k_w+1}{k_s}} e^{\frac{-s^2}{2}} \, ds \qquad\qquad (s = t/k_s)$$

$$\geqslant k_s \cdot \int_{1/5}^{3} e^{\frac{-s^2}{2}} \, ds \, . \qquad\qquad (k_w = 3k_s \text{ and } k_s \geqslant 5)$$

Evaluating numerically the integral in last display, we find

$$\sum_{u=1}^{k_w} \exp\left(\frac{-u^2}{2k_s^2}\right) \geqslant 1.0512 \cdot k_s \geqslant k_s \, . \qquad\qquad (30)$$

Coming back to Eq. (29), we get the promised lower bound.

In the other direction, we find that

$$\sum_{u=1}^{k_w} e^{\frac{-u^2}{2k_s^2}} \leqslant \int_0^{k_w} e^{\frac{-u^2}{2k_s^2}} \, du$$

$$= k_s \cdot \int_0^3 e^{\frac{-u^2}{2}} \, du \approx 1.24 \cdot k_s$$

$$\leqslant \frac{5k_s}{2 \cdot 2} \, .$$

Coming back to Eq. (29), we get the promised upper bound. $\qquad\square$

Lemma D.3 (Sum of squares is negligible). *Assume that $k_s \geqslant 5$. Then, for any $(i,j) \in I_m(k_w)$, it holds that*

$$\sum_{(u,v) \in C_{i,j}} \delta_{u,v}^2 \leqslant \frac{\Delta_{i,j}^2}{4k_s^2} \, .$$

Proof. Similarly to the proof of LemmaD.2, we first write that

$$\sum_{(u,v) \in C_{i,j}} \delta_{u,v}^2 = \left(2 \cdot \sum_{u=1}^{k_w} e^{\frac{-u^2}{k_s^2}} + 1\right)^2 \, ,$$

since we consider that $(i,j) \in I_m(k_w)$. Since the mapping $u \mapsto e^{-u^2/k_s^2}$ is decreasing on $\mathbb{R}_+$, we have

$$\sum_{u=1}^{k_w} e^{\frac{-u^2}{k_s^2}} \leqslant \int_0^{k_w} e^{\frac{-u^2}{k_s^2}} \, du$$

$$= k_s \cdot \int_0^3 e^{-t^2} \, dt \, ,$$

where we used $k_w = 3k_s$. Numerically, we find this integral to be smaller than $9/10$, and we deduce that

$$\sum_{(u,v) \in C_{i,j}} \delta_{u,v}^2 \leqslant \frac{(9k_s + 5)^2}{25} \, . \qquad\qquad (31)$$

Recall that, since we assumed $(i,j) \in I_m(k_w)$ and $k_s \geqslant 5$, according to Lemma D.2,

$$\Delta_{i,j}^2 \geqslant (2k_s + 1)^4 \, .$$

We notice that, for any $k_s \geqslant 5$,

$$\frac{k_s^2(18k_s + 10)^2}{100(2k_s + 1)^4} \leqslant \frac{81}{400} \, ,$$

and we deduce the result. $\qquad\square$

D.3 Facts about the ψ functions

In this section, we collect facts about ψ_1 and ψ_2 that are used throughout the proofs.

Lemma D.4 (Bounds on ψ_1). *For any $t \geqslant 0$, $\psi_1(t) \geqslant 0$. Moreover,*

$$\forall t \in \left[0, \frac{1}{25}\right], \qquad 4t^2 \leqslant \psi_1(t) \leqslant 6t^2 \,.$$

Proof. For any $t \geqslant 0$, we have

$$(1 + 2t)^6 - (1 + 4t)^3 = 12t^2 + 96t^3 + 240t^4 + 192t^5 + 64t^6 \geqslant 0 \,,$$

which proves the first statement. The second statement follows trough a function study. □

Lemma D.5 (Bounds on ψ_2). *For any $t \geqslant 0$, $\psi_2(t) \geqslant 0$. Moreover,*

$$\forall t \in \left[0, \frac{1}{25}\right], \qquad t^2 \leqslant \psi_2(t) \leqslant \frac{3t^2}{2} \,.$$

Proof. For any $t \geqslant 0$, we have

$$(1 + 2t)^6 - (1 + t)^3(1 + 3t)^3 = 3t^2 + 24t^3 + 69t^4 + 84t^5 + 37t^6 \geqslant 0 \,,$$

which proves the first statement. The second statement follows trough a function study. □

Lemma D.6 (A technical bound). *For any $t \in [0, 1/25]$,*

$$|2\psi_1(t) - \psi_2(t)| \leqslant 3t^2 \,.$$

Proof. Straightforward function study. □

D.4 Variance bounds

We detail here the variance bounds used in Section C.2.

Lemma D.7 (Upper bounding the variance in the bicolor case). *Assume that $\sigma^2 \leqslant k_s^2/25$ and that $\|c\| \geqslant 3k_s$. Then*

$$\left(\frac{k_s^2}{k_s^2 + 4\sigma^2}\right)^{\frac{3}{2}} e^{\frac{-\|c\|^2}{k_s^2 + 4\sigma^2}} - \left(\frac{k_s^2}{k_s^2 + 2\sigma^2}\right)^3 e^{\frac{-\|c\|^2}{k_s^2 + 2\sigma^2}} \leqslant \psi_1\left(\frac{\sigma^2}{k_s^2}\right) \,. \tag{32}$$

Proof. Let us first study the limit case: $\|c\| = 3k_s$. In that case, we see Eq. (32) as a difference of functions of $s = \sigma^2/k_s^2$. Studying this difference on $[0, 1/25]$, we see that it is negative on this interval, and thus Eq. (32) is true when $\|c\| = 3k_s$. Next, we see the left term of Eq. (32) as a function of $t = \|c\|^2$. We are going to prove that this function is decreasing. Indeed, the derivative at t is given by

$$\Phi(t) := \frac{-k_s^3}{(k_s^2 + 4\sigma^2)^{\frac{5}{2}}} e^{\frac{-t}{k_s^2 + 4\sigma^2}} + \frac{k_s^6}{(k_s^2 + 2\sigma^2)^4} e^{\frac{-t}{k_s^2 + 2\sigma^2}} \,.$$

Asking for $\Phi(t)$ to be negative is equivalent to asking for

$$\exp\left(t\left(\frac{-1}{k_s^2 + 2\sigma^2} + \frac{1}{k_s^2 + 4\sigma^2}\right)\right) < \frac{(k_s^2 + 2\sigma^2)^4}{k_s^3(k_s^2 + 4\sigma^2)^{\frac{5}{2}}} \,.$$

Since $k_s^2 + 2\sigma^2 < k_s^2 + 4\sigma^2$, the left-hand side of the last display is a decreasing function of t, and we just have to consider the limit case $t = 9k_s^2$. Setting again $s = \sigma^2/k_s^2$, one can check that

$$\exp\left(9\left(\frac{-1}{1 + 2s} + \frac{1}{1 + 4s}\right)\right) < \frac{(1 + 2s)^4}{(1 + 4s)^{5/2}}$$

for any $s \in [0, 1/25]$. This concludes the proof. □

Lemma D.8 (Upper bounding the covariance in the bicolor case). *Assume that $\sigma^2 \leqslant k_s^2/25$ and that $\|c\| \geqslant 3k_s$. Then*

$$\frac{k_s^6}{(k_s^2 + \sigma^2)^{\frac{3}{2}}(k_s^2 + 3\sigma^2)^{\frac{3}{2}}} \cdot e^{\frac{-\|c\|^2}{k_s^2 + 3\sigma^2}} - \left(\frac{k_s^2}{k_s^2 + 2\sigma^2}\right)^3 \cdot e^{\frac{-\|c\|^2}{2(k_s^2 + 2\sigma^2)}} \leqslant \psi_2\left(\frac{\sigma^2}{k_s^2}\right) .$$

Proof. This time the proof is much simpler. Indeed, since $k_s^2 + 3\sigma^2 < 2(k_s^2 + 2\sigma^2)$, one finds that

$$e^{\frac{-\|c\|^2}{k_s^2 + 3\sigma^2}} < e^{\frac{-\|c\|^2}{2(k_s^2 + 2\sigma^2)}} ,$$

and we deduce the result since this last display is less than 1. $\qquad\square$

D.5 Approximate expression for the rounded square area

The exact expression of B can be cumbersome to use. We sometimes prefer to use the following approximation, which gives close enough values for all practical purposes:

Lemma D.9 (Approximate expression for B). *Assume that $s < d < \sqrt{2}s$. Then*

$$\left|B(s, d) - \pi(3sd - s^2 - d^2)\right| \leqslant 0.04d^2 .$$

Proof. Recall that, for any s, d such that $s < d < \sqrt{2}s$,

$$\gamma(s, d) = d^2\gamma(t), \quad \text{with} \quad \gamma(t) = \arctan\frac{\sqrt{1 - t^2}}{t} - t\sqrt{1 - t^2} .$$

Numerically, we find that

$$\left|\gamma(t) - \frac{\pi}{2}\left(1 - \frac{3t}{2} + \frac{t^2}{2}\right)\right| \leqslant 0.04 .$$

We deduce that

$$\left|\gamma(s, d) - \frac{\pi}{2}\left(d^2 - \frac{3sd}{2} + \frac{s^2}{2}\right)\right| \leqslant 0.04d^2 .$$

The result follows by definition of B. $\qquad\square$

E ADDITIONAL EXPERIMENTS

In this section, we present additional experimental results. We start with further verification of the validity of Theorem 3.2 of the paper in Section E.1. In Section E.2, we provide additional insights on the distribution of pixel values for real images. Finally, we showcase more qualitative results for the scaling use-case in Section E.3.

E.1 Checking Theorem 3.2

In this section, we provide some additional experimental checks of Theorem 3.2 on synthetic data, namely with lower and higher k_s. They are summarized in Figure 12. As in the paper, we generated ten random images with increasing shape and counted the number of superpixels in the central square area. The paper presents results for $k_s = 5$ (which is the default choice), which we reproduce here for easier comparison (middle row of Figure 12).

E.2 Flat portions of real images

In this section, we showcase the limits of Assumption 3.1 when dealing with real images, expanding the discussion at the beginning of Section 5 in the paper. In Figure 13, we display the histogram of pixel values for an image of our dataset. More specifically, we took the same image as in Figure 1 of the paper, which we converted to CIELAB. Subsequently, we selected the pixels contained in the red rectangle and reported for each channel the histogram of the values. The main differences with our model are:
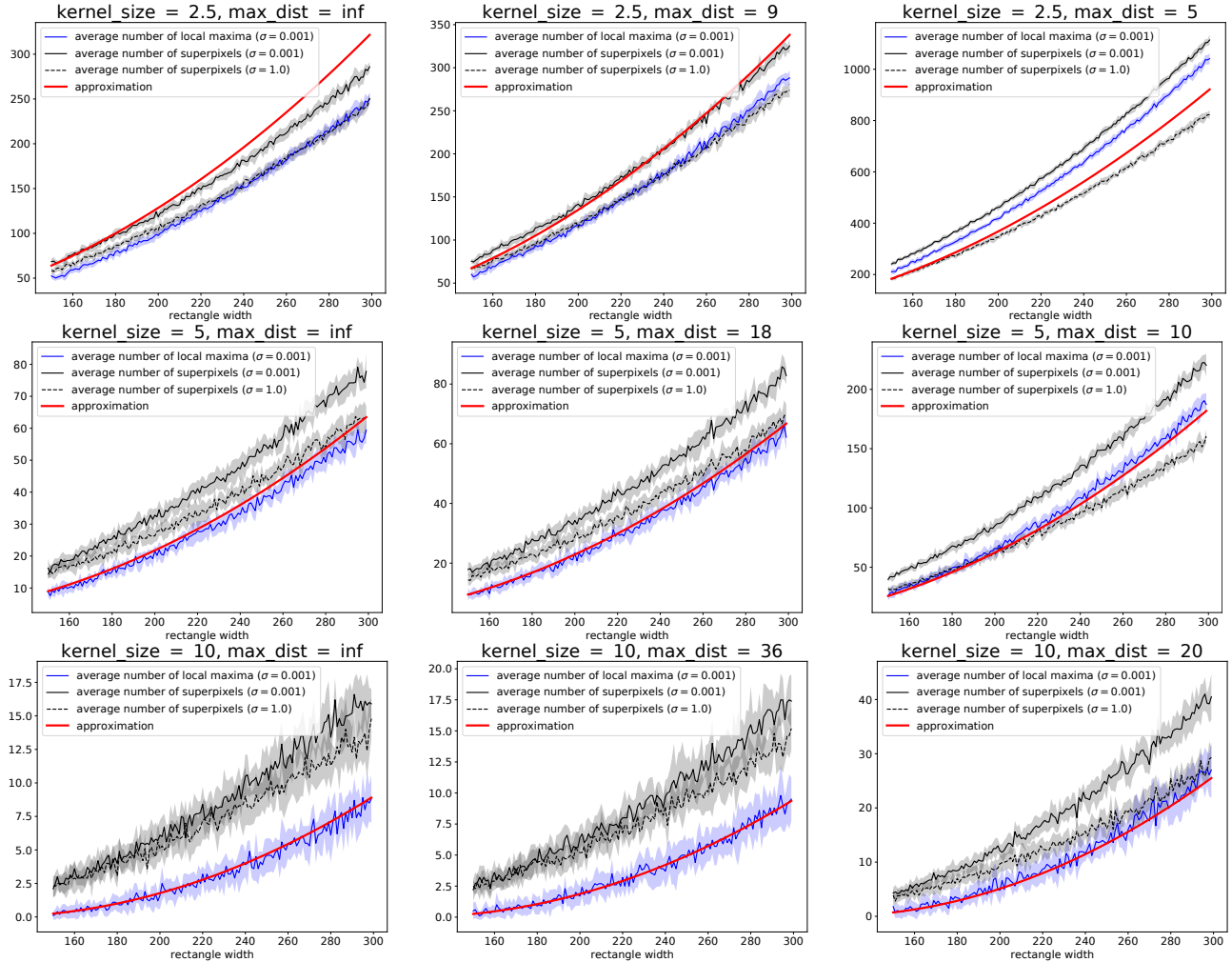
Figure 12: For small values of k_s, the averaging effect is not strong enough and the fit is not perfect between Theorem 3.2 and empirical runs of quickshift. However, we still recover the right scale for the number of superpixels. The fit improves as k_s increases. Notice the difference in the scale of the y-axes.

- **higher variance:** though rather small, the variance on each channel is higher than the values with which Theorem 3.2 of the paper is concerned. Moreover, it is not constant throughout all channels.

- **distribution is not Gaussian:** the pixel values distribution is not Gaussian. A Kolmogorov-Smirnov test rejects H_0 : "L is normally distributed" at any level. Moreover, the b channel is clearly multimodal (hinting that there are actually several colors in the rectangle).

- **spacial dependencies:** often, there is a source of light somewhere in the image drawing the L values. In our example, the Pearson correlation between the first line of L values and their indices is -0.68, far from 0, the theoretical value under Assumption 3.1.

This is a consistent behavior on all the flat regions of images that we tested.

E.3   Scaling hyperparemeters

In this section, we present more qualitative results for the rescaling use-case described in Section 5.3 of the paper. All experiments presented here are done with the ILSVRC dataset. We took $k_s = 5$ and a downsizing ratio equals to 2, that is, the downsized image has size $H/2 \times W/2$, rounded to the closest integer. In Figure 14 we take $d_m = +\infty$, in Figure 15 $d_m = 18$, and $d_m = 10$ in Figure 16. For each image, we produce superpixels
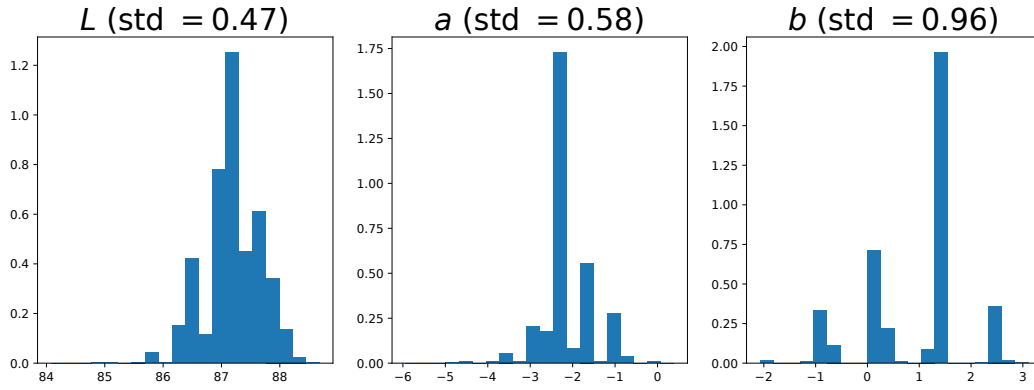
Figure 13: Histogram of CIELAB pixel values in the rectangle outlined in Figure 1 of the paper. Under Assumption 3.1, we should observe three bell curves with same width.

with the parameters indicated in the title. The **left image is the downsized version**, while the **middle and right image have the original size**. We denote by N the number of superpixels in the image. The heuristic proposed in the paper amounts to multiply both k_s and d_m by the downsizing ratio (here, 2), in order to get *approximately* the same number of superpixels in downsized image as in the original image.
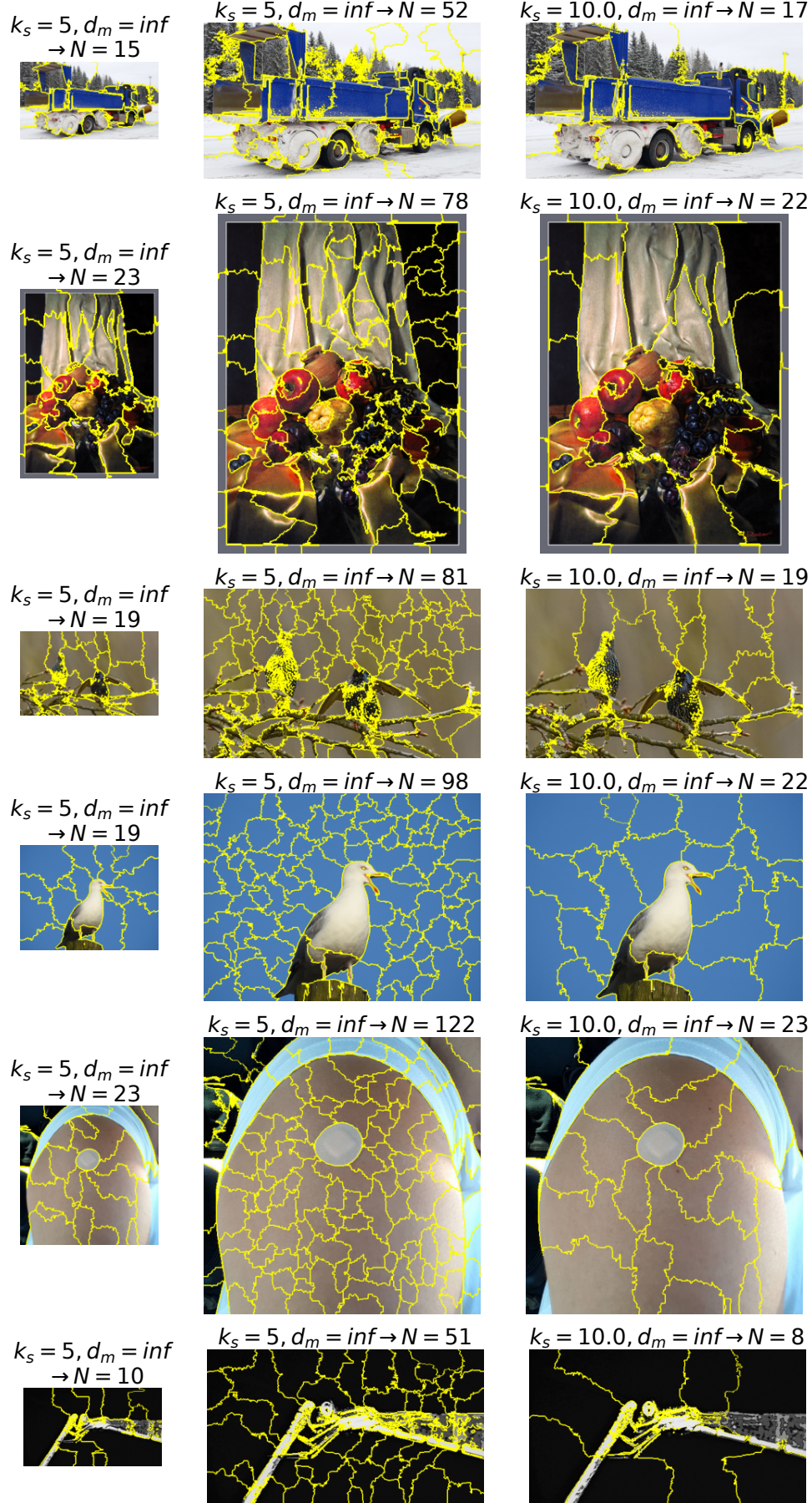
Figure 14: Additional results for the rescaling experiments. Original hyperparameters: $k_s = 5$ and $d_m = +\infty$.
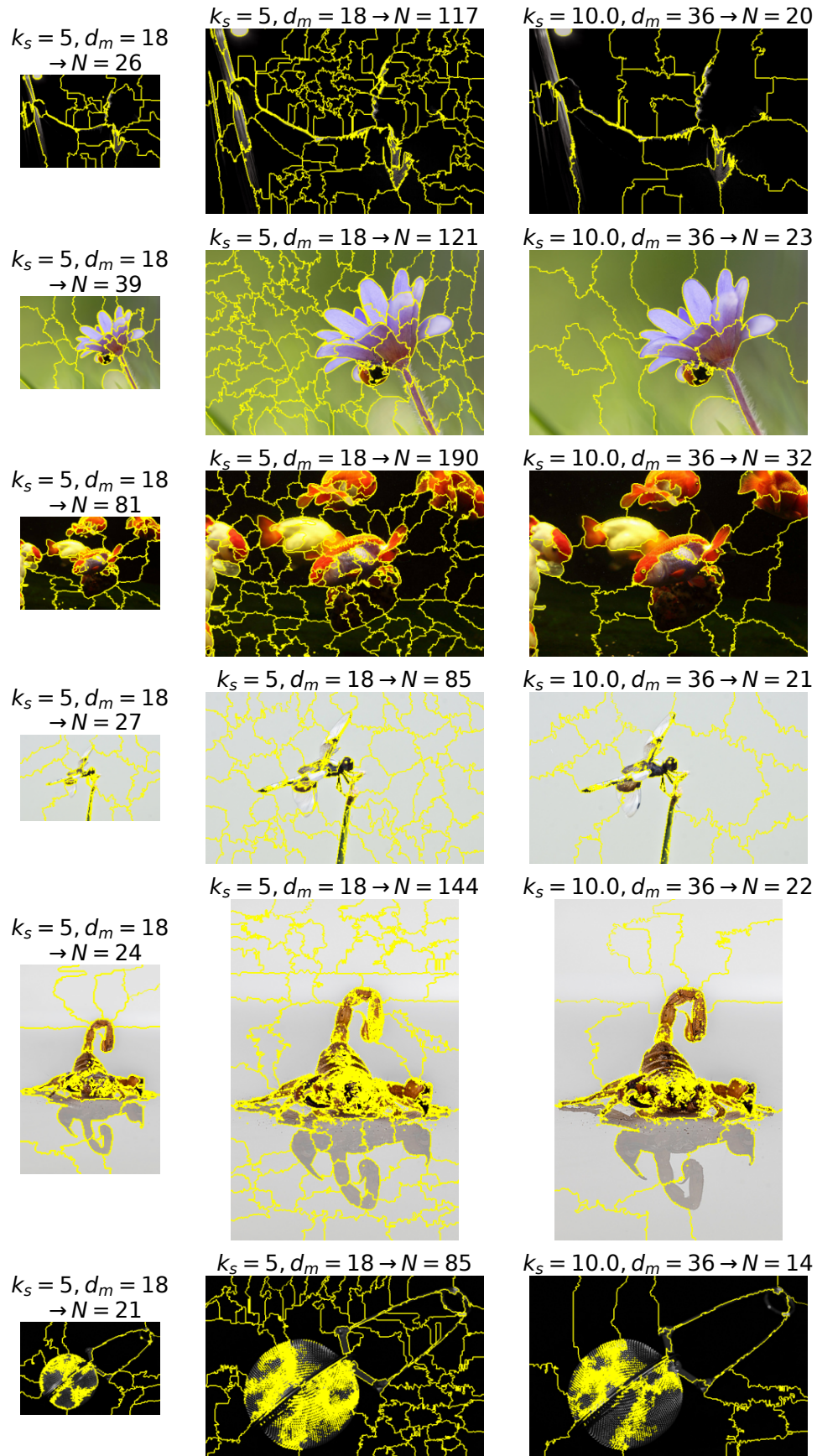
Figure 15: Additional results for the rescaling experiments. Original hyperparameters: $k_s = 5$ and $d_m = 18$.
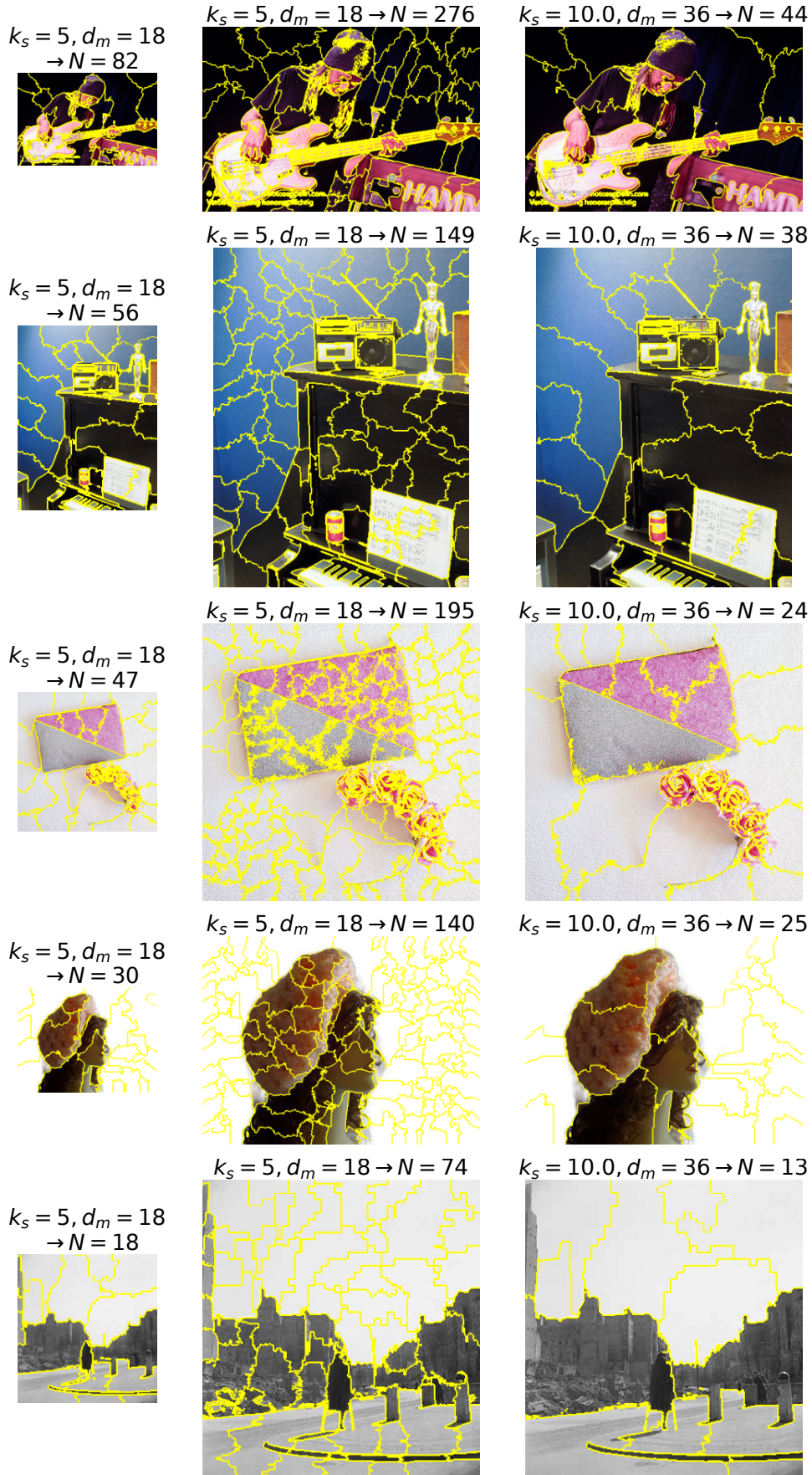
Figure 16: Additional results for the rescaling experiments. Original hyperparameters: $k_s = 5$ and $d_m = 10$.