Dynamic spectral residual superpixels

Jianchao Zhang [a,1], Angelica I. Aviles-Rivero [b,1,*], Daniel Heydecker [b,1], Xiaosheng Zhuang [a], Raymond Chan [a], Carola-Bibiane Schönlieb [b]

[a] *City University of Hong Kong (CityU), Hong Kong*
[b] *DPMMS and DAMTP, University of Cambridge, UK*

ARTICLE INFO

ABSTRACT

We consider the problem of segmenting an image into superpixels in the context of k-means clustering, in which we wish to decompose an image into local, homogeneous regions corresponding to the underlying objects. Our novel approach builds upon the widely used Simple Linear Iterative Clustering (SLIC), and incorporate a measure of objects' structure based on the spectral residual of an image. Based on this combination, we propose a modified initialisation scheme and search metric, which keeps fine-details. This combination leads to better adherence to object boundaries, while preventing unnecessary segmentation of large, uniform areas, and remaining computationally tractable in comparison to other methods. We demonstrate through numerical and visual experiments that our approach outperforms the state-of-the-art techniques.

1. Introduction

Image segmentation has been a widely explored task in computer vision yet a still open problem. In particular, superpixels segmentation has become a pre-processing tool for several applications including classification [1,2], optical flow [3,4], colour transfer [5], depth estimation [6] and tracking [7]. The central idea of superpixels is to split a given image in multiple clusters, which reflect semantically meaningful regions.

There are several advantages of using superpixel representation instead of working at pixel-wise level. Firstly, an application becomes computationally and representationally efficient as the number of primitives are significantly reduced. Secondly, the natural redundancy in an image is exploited, and therefore, features can be extracted on representative regions whilst reducing noise and increasing discriminative information [8–10].

Since the pioneering work of Ren and Malik [8], the community has devoted to develop different algorithmic approaches to improve over [8]. These approaches can be roughly divided into graph-based methods e.g. [8,11], path-based approaches e.g. [12],

density-based models e.g. [13], contour models e.g. [14] and clustering methods e.g. [9,15].

Out of all of the approaches reported in the literature, the Simple Linear Iterative Clustering (SLIC) [9] is perhaps the most popular method that offers a good performance whilst demanding low computational cost, by building on Lloyd's algorithm [16] for k-means. The central idea of SLIC is to perform the superpixels partition based on an iterative scheme that searches for similarities between points, ensuring that at each step we assign points to the nearest cluster from the previous step.

The ability of SLIC to obtain a good segmentation with low computational cost comes from the observation that, by using a similarity metric, one can greatly reduce the number of distance calculations required. However, SLIC is also limited by its own construction, and in particular, by its search range, and one can thus observe two major limitations. Firstly, SLIC tends to segment large uniform regions in an image with more superpixels than are intuitively necessary, which limits resolution in other parts of the image. Secondly, in structure-rich parts of the image, the final superpixel size is much smaller than the search radius of SLIC, which leads to many unnecessary distance computation. Finally, since we expect structure-rich parts of the image to have a higher density of superpixels, it may be efficient to perform the initial seeding of cluster centres in anticipation of this inhomogeneity.

In this work, motivated by the drawbacks discussed above, we propose a new algorithmic approach, exhibited in Fig. 1, that improves upon the SLIC approach. We show that our approach out-

---

* Corresponding author.
E-mail addresses: jzhang537-c@my.cityu.edu.hk (J. Zhang), ai323@cam.ac.uk (A.I. Aviles-Rivero), dh489@cam.ac.uk (D. Heydecker), xzhuang7@cityu.edu.hk (X. Zhuang), rchan.sci@cityu.edu.hk (R. Chan), cbs31@cam.ac.uk (C.-B. Schönlieb).
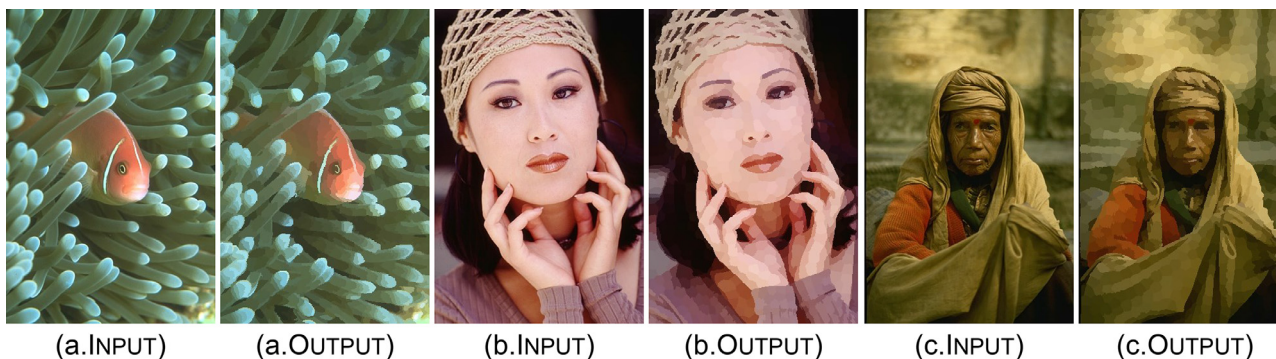[1] Equal contribution.

Fig. 1. Input-Output examples of our proposed approach for three images with different characteristics such as diverse structure and colours. Our approach seeks to adhere better to the boundaries through a structure measurement whilst connecting meaningful regions, for example see the eyes and mouth at the three outputs.
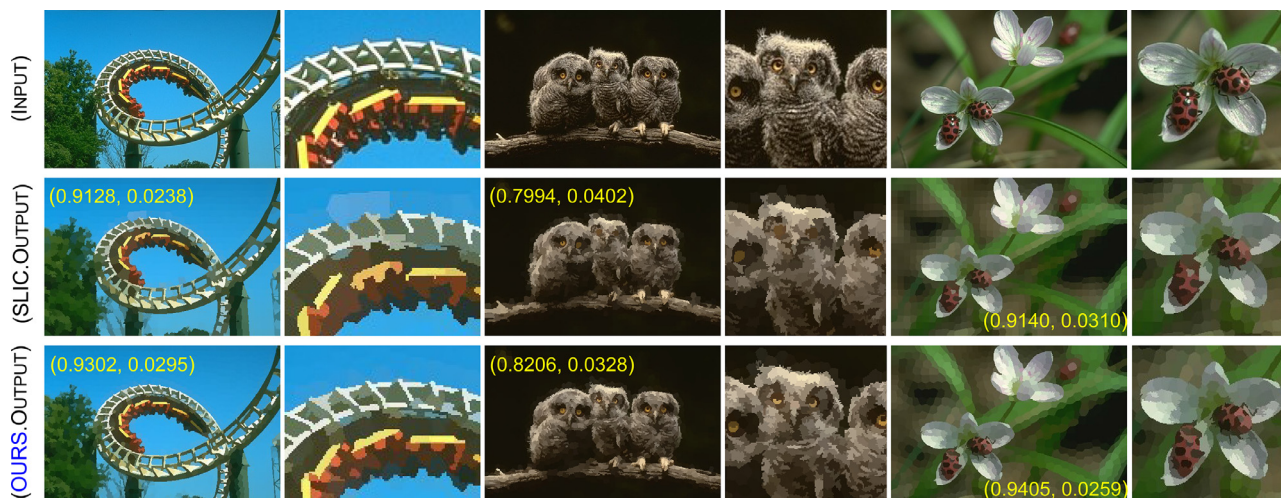

Fig. 2. Illustration output of our approach against SLIC. SLIC tends to over-segment uniform areas with more superpixels than necessary, such as the sky in the first image, and fails to preserve fine-structures, such as the owl's eyes and the roller coaster at the zoom-in views. The numerical results in parenthesis denote (Undersegmentation Error, Boundary Recall). The lower the undersegmentation error the better whilst the higher the boundary recall the better.

performs SLIC and several works on the body of literature. Our main contributions are as follows.

- We propose a new superpixel approach, which incorporates the *saliency function* $\mathcal{S}(x)$ of Hou et al. [17] as a proxy for object density. This leads to the following advantages.
  - By incorporating the saliency $\mathcal{S}(x)$ into the distance computation, we can prevent unnecessary over-segmentation of large, uniform regions, such as the sky in the first example of Fig. 2, and allowing greater focus on structure-rich parts of the image.
  - We propose a new seeding strategy, based on the inhomogenity described by $\mathcal{S}$. This allows for greater resolution changes at fewer iterations by focusing on relevant structures, and hence keeping fine-details of the structures in the final segmentation.
- We extensively evaluate our approach with a large range of numerical and visual experiments.
- We demonstrate that our two major contributions mitigate the major drawbacks of the state-of-the-art techniques, by reporting the lowest undersegmentation error and highest boundary recall.

2. Related work

In this section, we review the body of literature in turn. We then highlight the advantages of clustering based methods, and their current drawbacks that motivate our new algorithmic approach.

There have been different attempts in the literature to improve superpixels segmentation. A set of approaches tackle the problem using graph representation of the image. Within this umbrella, let us mention the *normalised cuts* technique [8,18], and some works relating to graph random walks [19,20]. Other techniques reported in the literature include a greedy-type algorithm [21] which forms clusters by seeking 'evidence of boundaries', and the technique of [22] which forces a segmentation with the topology of a square lattice. However, although promising results are reported, the computational time is often very high. Another perspective has been followed by local mode-seeking algorithms including the well-known Quick Shift, which partition is based on an approximation of kernelised mean-shift [13]. However, there is no control on the number of superpixels or compactness.

Another set of approaches addressed the superpixel partition problem as the task of finding the shortest path between seeds, for instance using the well-known Dijkstra algorithm, as reported in Tang et al. [12], Fu et al. [23], Condori et al. [24]. However, this type of approach is usually unable to control the compactness. We briefly mention other methods for superpixels segmentation. A body of work has proposed algorithms for image segmentation based on geometric methods such as mean shift [25], watershed methods [26,27], and gradient flows [14,28]. For an extensive review of the literature, we refer the reader to [10].

In particular, in this work we focus on what is probably the most popular superpixel category, which is clustering based approaches. The basis of this perspective builds on Lloyd's algorithm [16] for k-means clustering. The main idea of this algorithm is to partition a set of observations into k clusters, in which each observation is assigned to the cluster with the nearest mean, and produces excellent results at the cost of high computational intensity. Within this category, one can find a top reference approach called Simple Linear Iterative Clustering (SLIC). SLIC was proposed by Achanta et al. [9], in which authors propose a local version of the Lloyd's algorithm, which is computationally much simpler while keeping excellent segmentation quality.

Following this philosophy, different algorithmic approaches have been proposed including [29–32]. Most recently, in Achanta and Susstrunk [33] authors proposed an improved version of SLIC, in which they proposed to compute a polygonal partition to adapt better to the geometry of the objects in the image. Maierhofer et al. [15] proposed a *dynamic refinement* of this method, called dSLIC, which seeds the initial cluster points inhomogenously and allows the search radius to vary across clusters, both according to a measure of local object density. This allows better capturing of fine details in structure-rich regions, and further reduces computational complexity by eliminating unnecessary searches.

As with many other problems in image analysis, machine learning techniques have enjoyed considerable success in image segmentation since the emergence of Deep Neural Networks. We highlight the particular success of *Deeplab*, a deep-convolutional network [34]; a later work by Marmanis [35] seeks to improve boundary adherence by incorporating explicit edge-detection. A review of such methods can be found in [36].

Let us also mention the closely related problem of *salient object detection*. In this problem, one has the simpler goal of identifying which regions in an image contain *salient* or novel information, and which contain only patterns and structures repeated throughout the image. This problem shares some similarities with the problem of image segmentation; for instance, one might hope that the salient objects are identified as superpixel regions. A hugely successful method in this problem, based on Fourier analysis, was proposed by Hou et al. [17], which inspires our current approach. More recent works include techniques based on graphs [37] or machine-learning [38,39].

3. Proposed approach

In this section, we describe in detail our superpixel approach. Firstly, we formalise the definition of superpixels in terms of a clustering task. We then introduce the details of both our new measure of structure function and our initialisation strategy.

We view an input image, of width $\mathcal{A}$ and height $\mathcal{B}$, as a map $I : \mathcal{X} \to \Omega$, where $\mathcal{X} = [\mathcal{A}] \times [\mathcal{B}]$ is a rectangular domain, and $\Omega \subset \mathbb{R}^3$ is the image domain with an appropriate colour domain. We also define a metric d on $\mathcal{X} \times \Omega$, representing the similarity of points in space and with different colour values, and a *feature map* $\mathcal{F}$, which takes a subset $S \subset \mathcal{X}$ and returns a pair in $\mathcal{X} \times \Omega$. *k-means clustering* now seeks a partition of $\mathcal{X}$ into path-connected sets $\{S_i\}_{i=1}^n$ such that, for each i, S_i is exactly the set of points where the infimum $\inf_j d((x, I(x)), \mathcal{F}(S_j))$ is attained at $j = i$.

We first give a very brief explanation on SLIC as our approach builds upon it. SLIC uses Lab colour space as this is a representation of the visible colours which simulates human vision.

Definition 1 (Lab colour space). The Lab colour space L_{ab} describes mathematically all perceivable colours in the three dimensions l for lightness and a and b for the colour opponents green-red and blue-yellow. The range of coordinates for l are 0 to 100 and bounded intervals for a and b respectively, the bounds on which depend on the convention used.

Given this particular choice of coordinates for our colour space, SLIC chooses the following distance measure: For $\mathbf{p}_1, \mathbf{p}_2 \in \mathcal{X} \times \Omega$, $\mathbf{p}_i = [\mathbf{x}_i, \mathbf{l}_i]^T$, $\mathbf{x}_i \in \mathcal{X}$, $\mathbf{l}_i \in \Omega$ define:

$$d(\mathbf{p}_1, \mathbf{p}_2) = \sqrt{d_s^2 + \left(\frac{d_c}{S}\right)^2 m^2} \quad \text{where}$$

$$d_s(\mathbf{p}_1, \mathbf{p}_2) = \|\mathbf{x}_1 - \mathbf{x}_2\|_2 \quad \text{and} \quad d_c(\mathbf{p}_1, \mathbf{p}_2) = \|\mathbf{l}_1 - \mathbf{l}_2\|_2,$$

and m is a parameter which tunes the importance of spatial as compared to Lab-distance. At the practical level, the value of m strongly impacts the shape of the superpixels found.

3.1. Object density measure via spectral residual

The key strength of dSLIC [15] over SLIC is the recognition that objects in an image are not distributed uniformly, and that image segmentation can exploit this to improve segmentation results and computational efficiency. Our approach is to exploit this same principle further, and use the strength of the *Spectral Residual* approach proposed by [17] as a better measure of object detection.

We briefly review the Fourier analysis leading to the definition of the spectral residual in Hou and Zhang [17]. For an image I, we write $\mathfrak{F}I$ for the Fourier transform, which is a matrix of the same dimensions as I, and whose arguments we will write as two-dimensional frequencies, f. The log-spectrum of an image I is then given by

$$\mathcal{L}(f) = \log(\Re(\mathfrak{F}I)(f)) \tag{1}$$

where $\Re$ denotes the real part; we also write $\mathcal{P}(f) = \Im(\mathfrak{F}I)(f)$ for the imaginary part, or phase spectrum. The key insight of [17] is that much of the information contained within $\mathcal{L}$ is redundant, because $\mathcal{L}$ is, to a good approximation, locally linear. These features are then encapsulated in the local average $\mathcal{A}(f) = (h_n \star \mathcal{L})(f)$, where h_n is the matrix consisting entirely of $\frac{1}{|\mathcal{X}|}$, and the *residual* log-spectrum, corresponding to the salient features, is given by the following expression:

$$\mathcal{R}(f) = \mathcal{L}(f) - \mathcal{A}(f) = \mathcal{L}(f) - (h_n \star \mathcal{L})(f). \tag{2}$$

The final saliency map, which we take as our measure of object density, is then given by recombining with the phase spectrum, inverting the Fourier transform and adjusting the resulting function. Therefore, our proposed function reads:

$$\mathcal{SR}(x) = g_\sigma \star \left(\mathfrak{F}^{-1}\left\{e^{\mathcal{R}(f) + \mathcal{P}(f)}\right\}\right)^2(x), \tag{3}$$

where the squaring ensures that the quantity considered is non-negative, and the convolution with a Gaussian kernel g_σ ensures that the final result is smooth. We found that, for our algorithm, $\sigma = 20$ works well on a range of images. We then set a rescaling step for the search radius using (3) as $\mathcal{G}(x) := \exp(\mathcal{SR}(x) - \overline{\mathcal{SR}})$, where $\overline{\mathcal{SR}}$ denotes the mean of the structure function on the image grid. We then propose to have the distance computations depending on our function, which reads:

```
1  #pseudocode for  computing distances
2  input m (compactness), number of superpixels
3  while e_r ≤  outset
4      for  x ∈ X, 1 ≤ i ≤ k:
5          if |x − (F(S_i^(t)))_1| ≤ 2SG(F(S_i^(t))):
6              #distance computation
7              Compute: d((x, I(x)), F(S_i^(t)))
8          else:
9              d((x, I(x)), F(S_i^(t))) = ∞
10 S_i^(t+1) = min_{1≤j≤n} d((x, I(x)), F(S_j^(t)))
11 Compute Residual Error e_r
12 Increase t=t+1
```

By incorporating our proposed function, which we use as a measure of object density, into the distance computation we hope to address the two drawbacks mentioned in the introduction. Firstly, by doing a dynamic adjustment of the search range based on our function $\mathcal{G}$, one can connect uniform regions, and so avoid segmenting large, homogeneous regions into unnecessarily small superpixels. This effect is illustrated in Fig. 2; for example, in the second column where our approach was able to keep the sky in a same region, and the yellow car. Secondly, our approach focuses on segmenting fine details by capturing relevant structures; this is visible in the owl's eyes and head in Fig. 2. Furthermore, we remark that our approach does not introduce additional parameters to fine-tuning.

We now turn to explain our second major modification, which concerns the seeding initialisation.

3.2. Seeding initialisation: a new strategy

In this section, we describe our new seeding initialisation. Our main motivation is that we can use the object density measure $\mathcal{G}$ defined above to help seed clusters in object-rich parts of the image, which we expect to contain more distinct regions. In this way, we obtain greater resolution at fewer iterations, and improve the focus on relevant and interesting regions.

We remark to the reader that SLIC initialisation is based on sampling pixels at the image grid. For comparison purposes, we start by defining the SLIC initilisation, which reads as follows.

```
1  #pseudocode for seeding initialisation SLIC
2  Set: Initialise cluster centers as
3  C_i = [x_i, y_i, l_i, a_i, b_i] by sampling at regular grid
4  step: S = √(N/k)
5  #where N is the size of the image
6  Move cluster centers to the lowest gradient
7  position in a 3 × 3 neighborhood
```

Our proposed approach, which incorporates $\mathcal{G}$ into this seeding, can be described informally as follows. We first set as an initial point the pixel with the lowest value in $\mathcal{G}$, and then we increase the values near to the initial point such that its neighbours are unlikely to be selected as another initial point. In this way, we guarantee that the distance between seed points is comparable to the search range, which will help reduce redundant searches. This process is illustrated in Fig. 3 for two initial seeds.

The hedging described above is carried out in two stages as follows.

- Points adjacent to the initial point are made unselectable, by setting the value of $\mathcal{G}$ at these points excessively large.
- Points in the proximity of the initial point are made less likely, but not impossible, to select, again by altering $\mathcal{G}$. The influence range and to what extent are under consideration.

The advantage of these changes is that the density of area is limited twice compared with original method. The overall procedure of our method, which suitably sets the initialisation points according to our structure measure $\mathcal{G}$, is described formally as follows.

```
1  #pseudocode for seeding initialisation OURS
2  Set r = max(G)/min(G)
3  While Enough Seeds:
4      Set range=sqrt(NumOfPixels/NumOfSuperpixels)
5      for each Superpixel center C_j:
6          Initialise S_i with the lowest value
7          in G
8          Set the adjacent neighbours of S_i
9          G = ∞
10         for all points S_j with d(S_i, S_j) < range
11             G(S_j) = G(S_j) * sqrt(r)
12             Smooth region with g_τ, τ = 13/2
```

An output example is displayed in Fig. 4. Subfigure (a) shows an initialisation comparison between our approach and SLIC, and we see that our approach gives more importance to the ostrich than the background. In subfigure (b), we evaluate possible choices for τ, and display outputs for $\tau = 4, 6, 8$. In practice, we found that $\tau = 13/2$ works for a range of images. Finally, subfigure (c) shows a visual illustration of the advantage of our technique. By integrating our density function into the distance computation one can avoid unnecessary over-segmentation of large uniform regions (i.e. the background) and aim the attention at structure-rich parts (i.e. the ostrich). A closer look at the zoom-in regions one can see that our technique was able to preserve fine details in the face whilst SLIC lost relevant areas such as the left eye.

4. Experimental results

In this section, we describe in detail the experiments that we ran to evaluate our approach.

4.1. Evaluation protocol

Dataset description We evaluate our proposed approach on a publicly available dataset, the Berkeley Segmentation Dataset [40], which provides ground truth of the images for quantitative analysis. Moreover, to further support our technique, we included two interesting cases to study where fine-details are particularly difficult to retrieve: (1) out-of-the-focus images using the dataset of that [41] and (2) wide angle images using the dataset from [42]. All the measurements and comparisons in this section are taken from these datasets.

Comparison methodology We compare our approach to the SOTA methods on superpixels. For this, we design a two-part evaluation scheme. For the first part, we compared our approach against SLIC [9]; this comparison therefore demonstrate that our carefully design solution achieves better performance than the top reference in clustering-based methods. For the second part, we
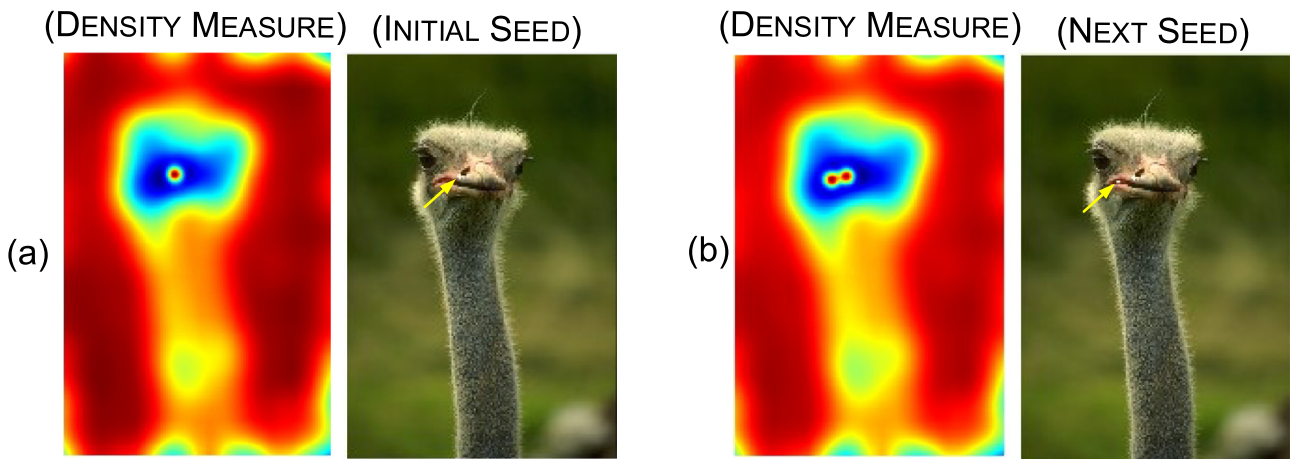
Fig. 3. Illustration of our initialisation strategy, which incorporates the object density measure $\mathcal{G}$,for two initial seeds. From left to right, first seed and second one.
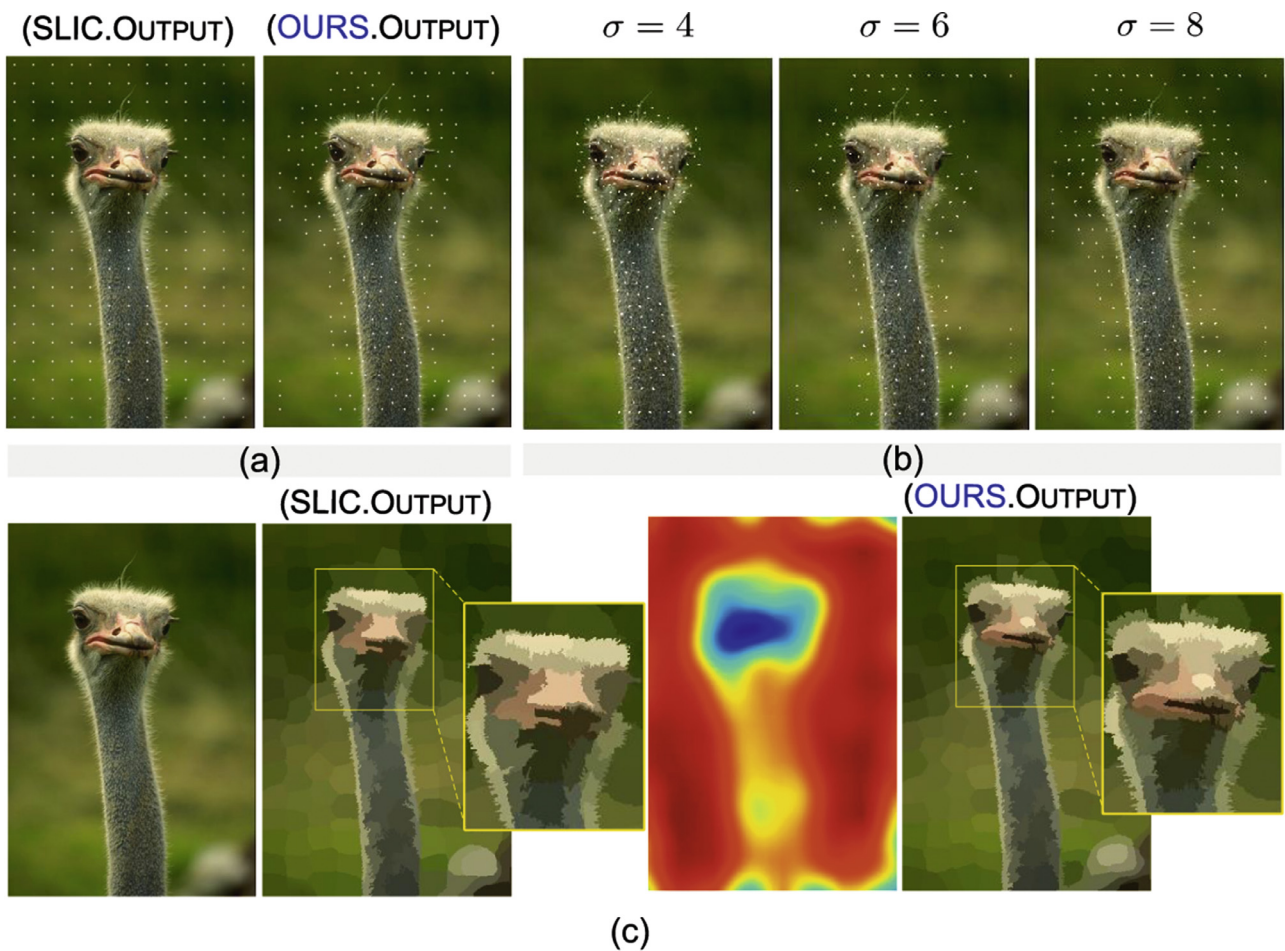

Fig. 4. Visual comparison of the seeding initialisation of SLIC vs ours.(a) One can see that we seek to focus on relevant areas (i.e. other than background). (b) The effect of τ in our seeding strategy. (c) Visual illustration of the positive repercussion of our technique over SLIC. Our approach seeks to focus on structure-rich parts (see the face) whilst avoid over-segmenting uniform areas such as the background.

compared to state-of-the-art techniques: QS [13], TP [14],TPS [12], LRW [19], SNIC [33] and dSLIC [15]. We compare our approach qualitatively by visual comparisons and quantitatively by computing three metrics: Under-segmentation Error (UE), Boundary Recall (BR) and Boundary Precision (BP), which definitions are described next.

We assume that we are given an image, along with a *ground truth* $\Gamma = \{g_i\}_{i=1}^{M}$,representing the true regions of the image.

Boundary recall measures the proportion of the boundary of the true regions in the ground truth which are close to a boundary in the segmentation. To quantify the notion of being close to a boundary, we recall the following definition.
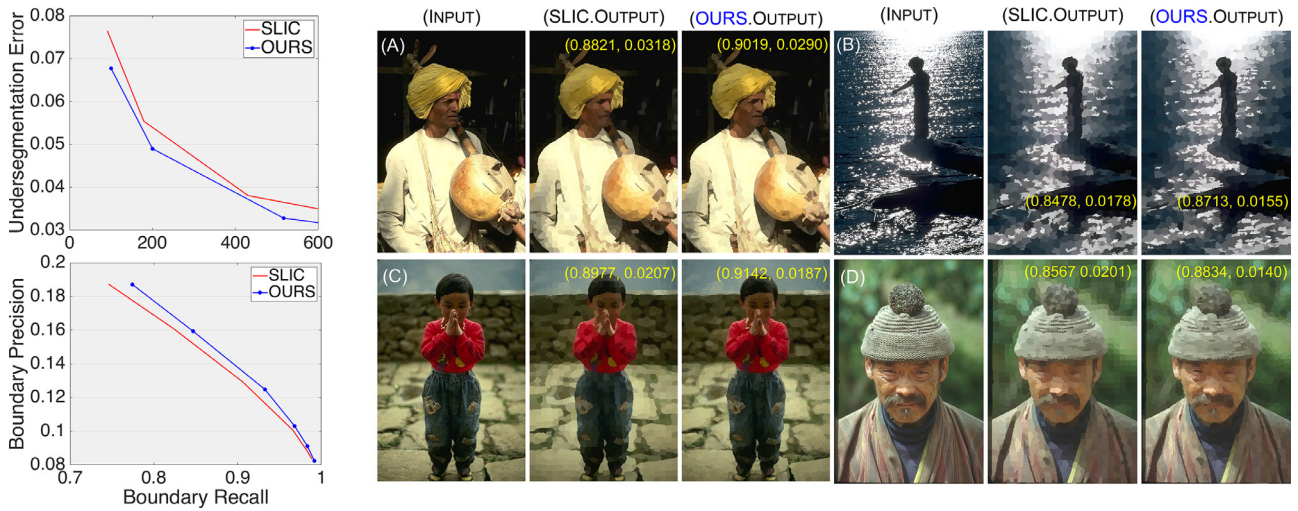
Fig. 5. From left to right. Quantitative comparison of our approach vs SLIC using three metrics: UE, BR and BP. Our approach reported the best scores metric-wise. Four visual outputs comparisons of SLIC vs our approach. In a closer look, one can see that our approach achieves better connection of structures and keeps fine-details. For example, see (A), (C) and (D) the faces and (B) the hand. The numerical results in parenthesis denote (Undersegmentation Error, Boundary Recall).
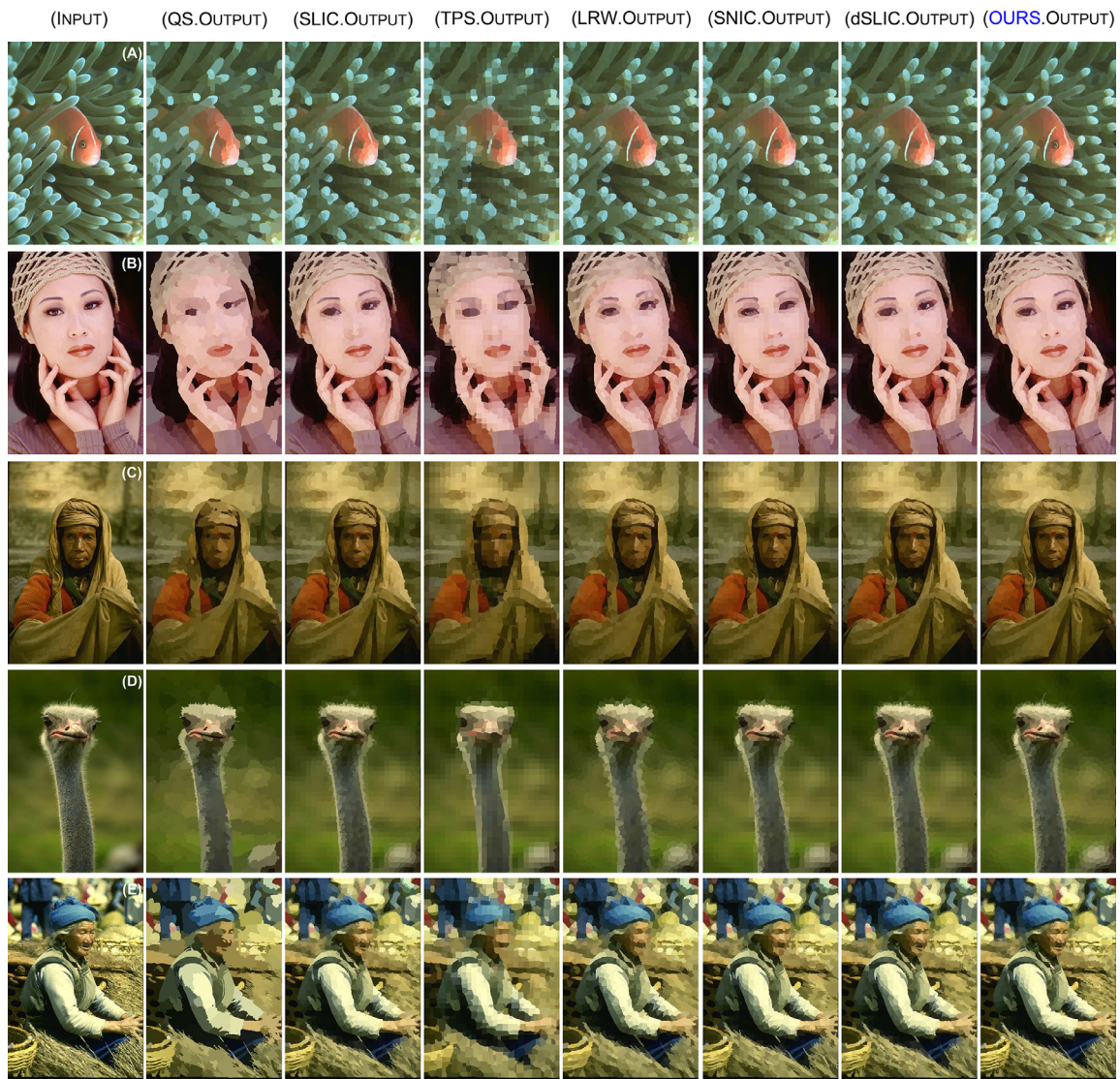

Fig. 6. Superpixel outputs comparisons of our approach vs different methods from the body of literature: QS [13], SLIC [9] TP [14], TPS [12], LRW [19], SNIC [33] and dSLIC [15]. A closer inspection, one can see that our approach offers better superpixels segmentation. For example, (A), (B), and (C) the eyes; (D) the ostrich's boundary and (E) the eyes and basket.
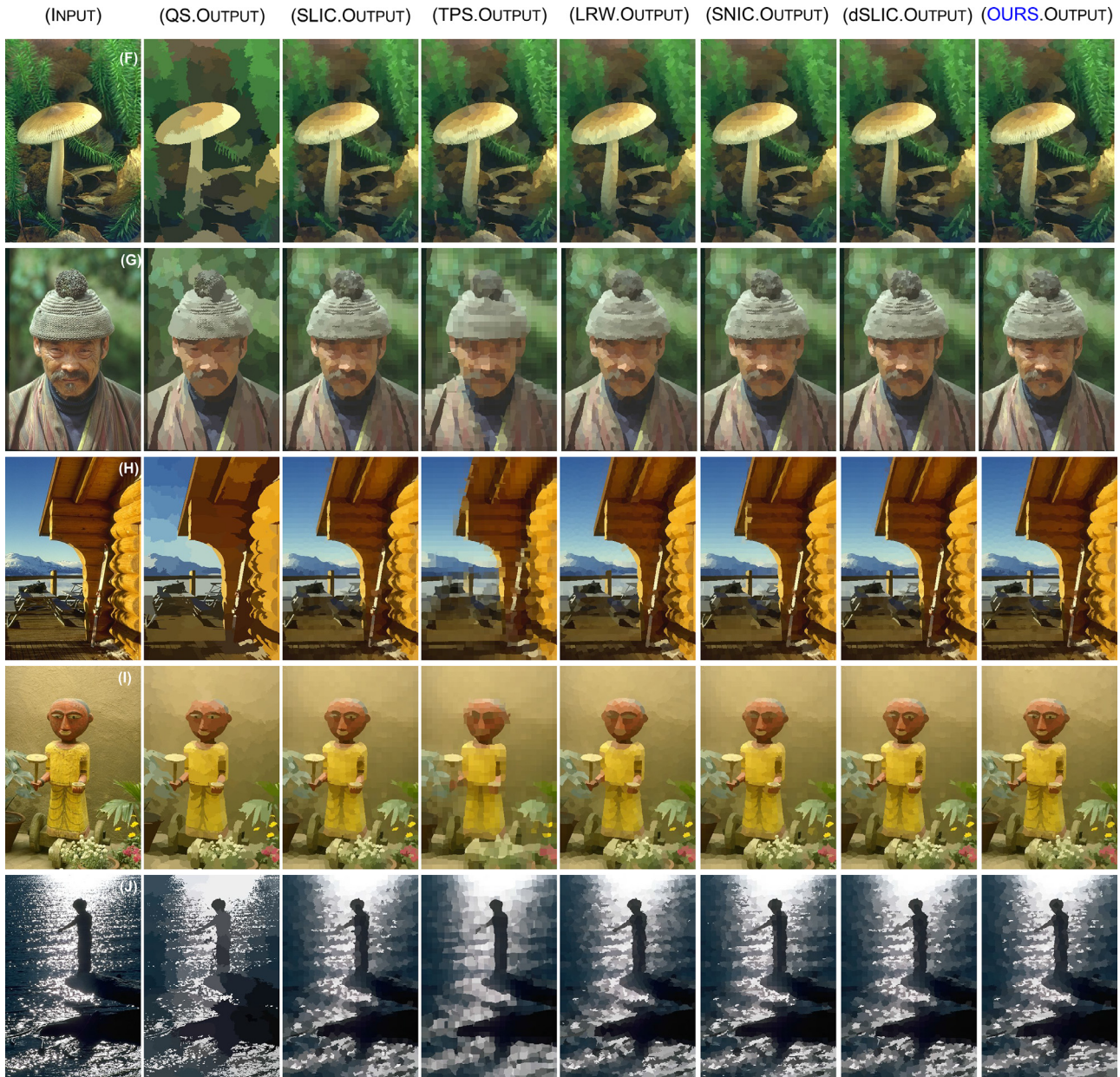
(INPUT)   (QS.OUTPUT)   (SLIC.OUTPUT)   (TPS.OUTPUT)   (LRW.OUTPUT)   (SNIC.OUTPUT)   (dSLIC.OUTPUT)   (OURS.OUTPUT)

Fig. 7. Superpixel segmentation outputs of our approach vs QS [13], SLIC [9] TP [14], TPS [12], LRW [19], SNIC [33] and dSLIC [15]. Visual assessment shows that the proposed algorithm performs better than the compared approaches. Examples are: (F) the leaves; (G) and (I) the face; (H) the house boundaries and (J) the hand.

Given a subset E of the edge set, we define the distance $d(e, E) = \inf\{|e - f| : f \in E\}$, where $|\cdot|$ denotes the l_2 norm of the difference, measured in pixels. We then define the *Boundary Recall* as

Definition 2 (Boundary Recall). Given a ground truth $\Gamma = \{g_i\}_{i=1}^M$ and a segmentation $\mathcal{S} = \{s_j\}_{j=1}^k$, we write $\partial\Gamma$ for the union of the edge boundaries ∂g_i, and similarly write $\partial\mathcal{S}$ for $\{s_j\}_{j=1}^k$. We define the boundary recall by

$$\partial(\Gamma, \mathcal{S}) = \frac{\#\{e \in \partial\Gamma : d(e, \partial\mathcal{S}) \leq 2\}}{\#\partial\Gamma}.$$

In words, the boundary recall is the proportion of true edges which are close to a superpixel edge.

Undersegmentation error Intuitively, this measures the size of all superpixels which *spill* across boundaries of the ground-truth.

Definition 3. For a ground truth $\Gamma = \{g_i\}_{i=1}^M$, we fix thresholds $B_i, i = 1, \ldots, M$. Given segmentation $\mathcal{S} = \{s_i\}_{i=1}^k$ of the image, the under-segmentation error is given by

$$U = \frac{1}{N}\left[\sum_{i=1}^M \left(\sum_{\#s_j \cap g_i \geq B_i} \#s_j\right) - N\right]$$

Observe that, since $\mathcal{S}$ is a partition of the image, we can rewrite

$$U = \frac{1}{N}\sum_{i=1}^M \left[\left(\sum_{\#s_j \cap g_i \geq B_i} \#s_j\right) - \#g_i\right]$$

Hence, the undersegmentation error U measures how wasteful the coverings of the true regions g_i by the superpixel regions s_j are. We usually take B_i to be a fixed proportion of $\#g_i$

Parameter selection For all compared approaches QS [13], TP [14], TPS [12], SLIC [9], LRW [19], SNIC [33] and dSLIC [15],
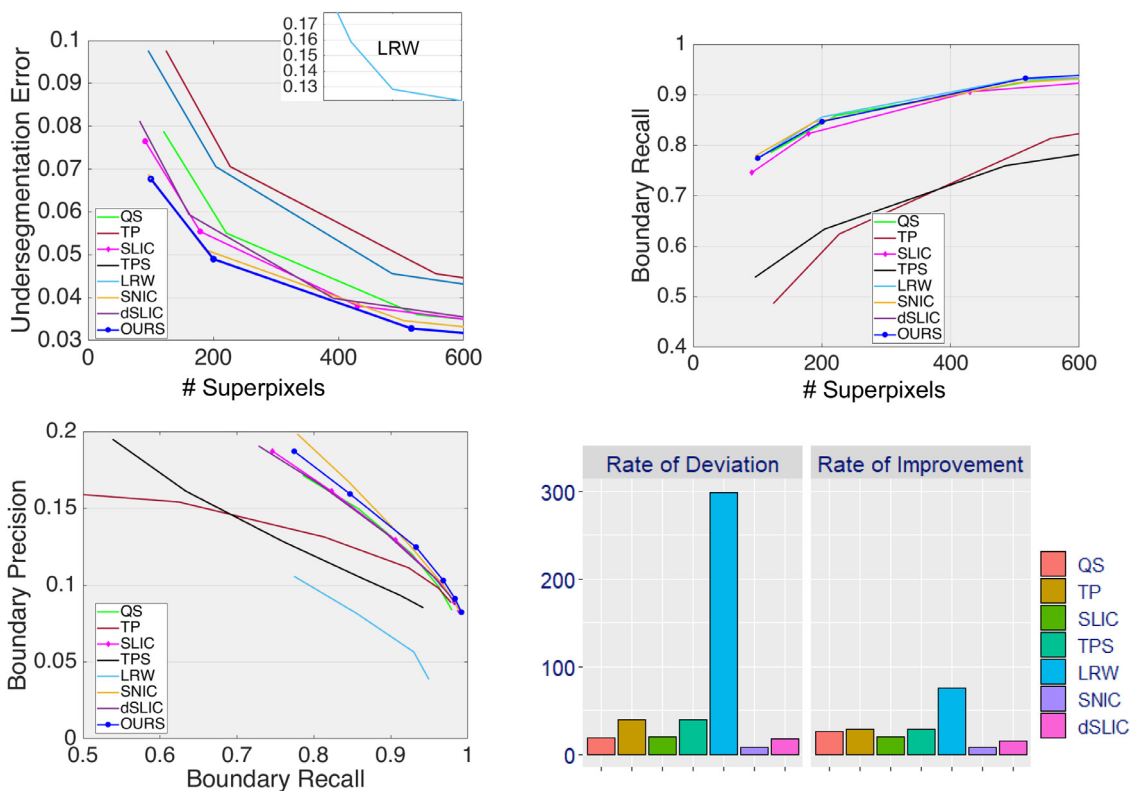
Fig. 8. Metric-wise comparison of our approach vs SOTA techniques using UE, BR and BP. In a closer look, we can see that our approach, overall, offers the lowest UE and the highest BR. Finally, the good boundary adherence to the true edges is reflected in the last plot, in which our approach overall gets the best trade-off between those metrics. These results are further supported by the rate of deviation and rate of improvement.

we set the parameters as suggested in the corresponding work. We also used the codes released from each corresponding author. For our approach, we set the $m = 10$ since it offers a good trade-off between shape uniformity and boundary adherence (see Supplementary Material Section 2 for further description on m). We performed the evaluation using up to a range of number of superpixels up to 600.

The experiments reported in this section were under the same conditions in a Matlab CPU-based implementation. We used an Intel Core i7 with 16 GB RAM.

4.2. Results and discussion

We divide this section in two parts, following the comparison methodology scheme presented in previous sections.

▷**Is our Approach better than SLIC?** As SLIC approach remains a top reference, and is the basis of our approach, we start by evaluating our approach against it. Results are displayed in Fig. 5. In a closer look, at the right side, of this figure, one can see that our approach yields to a better segmentation of the structures, keeping fine details of the objects. Moreover, it avoids unnecessary oversegmentation on uniform areas. These positive properties of our approach can be seen, for example, in (B) the proper recovery of the hand; in (C) the hair, eyebrows and the lines patterns in the jumper that are correctly clustered; in (D) where our approach successfully capture the eyes and moustache, and in (A) with better preservation of the face structure including the nose and lips.

To further support of our visual results, we ran a quantitative analysis based on three metrics UE, BR and BP. The results are displayed at the left side of Fig. 5. The top part shows a comparison in terms of UE, where the results reflect conformity to the true boundaries. We can observe that our approach achieves the lowest UE for all superpixels counts. The same positive effect was found in

terms of precision versus recall, in which our approach displayed the best performance. Overall, our proposed technique outperforms SLIC at all superpixels counts and reduces the error rate by 20%. This improvement is translated to our approach to be the best in terms of producing superpixels that respect the object boundaries.

▷**Is our approach better than other Superpixel approaches?**. As the second part of our evaluation, we compare our approach against SOTA models: QS [13], TP [14], TPS [12], LRW [19], SNIC [33] and dSLIC [15]. We selected for our comparison approaches coming from different perspectives: graph-based, path-based, density-based and clustering-based approaches. Results are displayed in Figs. 6, 7 and 8.

We first present a visual comparison of a selection of images from the Berkeley dataset in Figs. 6 and Fig. 7. By visual inspection, one can see that QS and TPS are the ones that perform more poorly than the other compared approaches. They fail to provide good boundaries of the structures and they do not preserve relevant details. Examples can be seen in (A), (B), (E) and (G) the eyes; (C), (E) and (F) preservation of fine details. LRW offer a better edge adherence to the objects than QS and TPS but also fails to preserve relevant objects, for example (G) the eyes and moustache.

In contrast to those approaches, SLIC and dSLIC performs better than QS, TPS and LRW. One can observe that SLIC and dSLIC readily compete in terms of having better boundary adherence to the structures and grouping correctly majority of the objects. However, in particular, SLIC still produces outputs with more superpixels than necessary in homogeneous parts of the structures; see, for instance, in (A) the fish eye, (B) the nose and (J) the hand. Although dSLIC performs slightly better than SLIC, it still fails to capture fine details.

Among those approaches, SNIC approach display more robustness in terms of grouping structures correctly than the previous
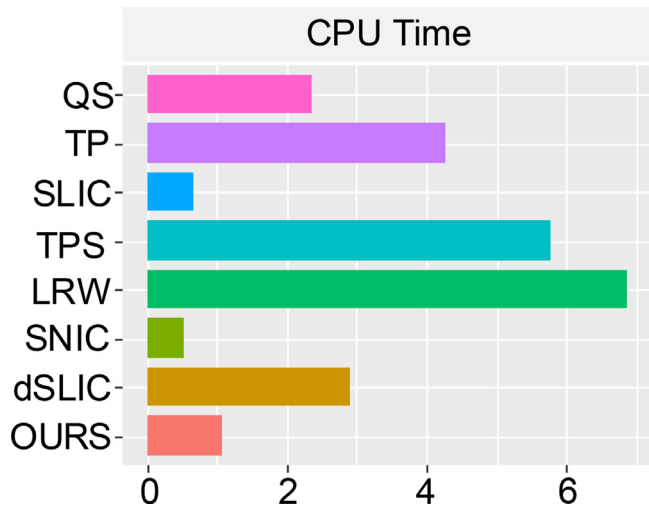
Fig. 9. CPU time averaged comparison of our approach vs the body of literature. One can see that our approach improvement comes at a negligible cost in runtime in comparison with the fastest approaches SLIC and SNIC.

approaches. However, like SLIC it also tends to generate more superpixels than needed in uniform regions so that the final outputs do not capture fine details. Examples are (G), (E) and (C) face details; (I) the eyes and head and (F) the leaves.

These major drawbacks are mitigated by our model. Our algorithmic approach shows the best boundary adherence and regularity. This is visible in the leaves in image (F), in which our approach is able to better capture the structure, in (I) on the lips where our approach is able to capture the correct geometry, and (A) the fish eyes, where our approach is the only one that correctly segments the inner part. These positive properties of our approach are prevalent in all images. More examples include preservation of the geometry such as in (J) the hand and (G) face, in which our model is the only one able to correctly segment these fine details.

To further support our visual evaluation, we show a metric-wise comparison in Fig. 8. We start by evaluating the approaches in terms of UE, which is displayed at the left side of this figure. Close observation shows that QS, TPS and TP perform poorly, and in particular, LRW that reported the highest Undersegmentation Error. dSLIC and SLIC show quite low UE, and SNIC ranks the second best. Overall, our approach shows substantial improvement over the compared approaches reporting the lowest UE for all superpixels counts. A similar effect is exhibited in terms of Boundary
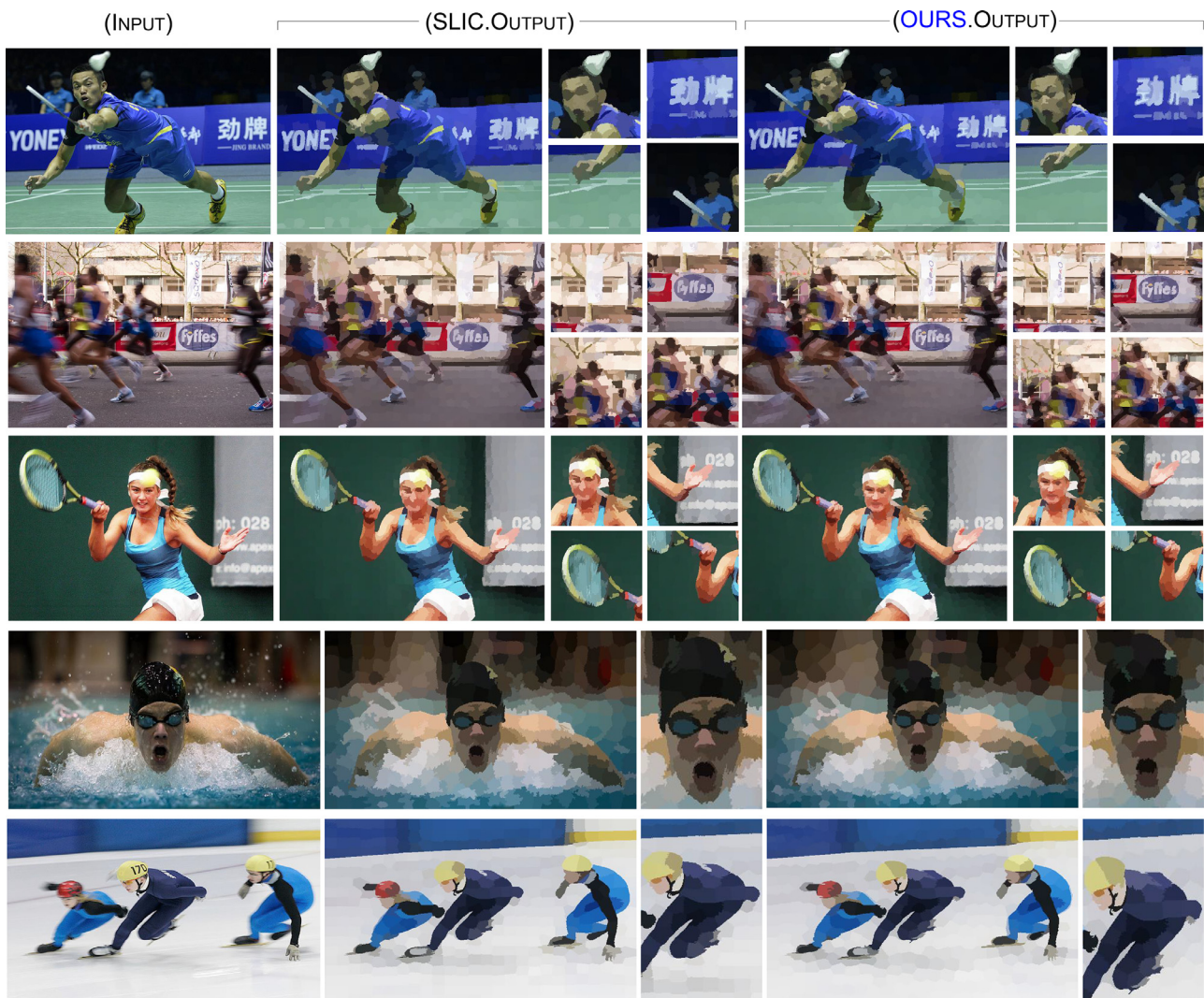

Fig. 10. Visual comparison of our approach and SLIC using out-of-the-focus sport images. Zoom-in views display relevant areas, in which our approach successfully retrieve fine detail and blurry structures.

Fig. 11. Visual comparison against SLIC using out-of-the-focus images. The selected frames show challenges cases when objects are highly blurry. Overall, our technique gives a better superpixel segmentation by preserving objects structures and fine details.

Recall. TP and TPS perform poorly while the other compared approaches reported better BR. Our approach readily competes with the other compared schemes and the overall BR of our approach was reported to be the best. The same effect is observed in terms of BP-vs-BR, which reflects that our approach overall adheres better to the truth boundaries. To have a better sense of the improvement, we finally report the rate of deviation and rate of improvement of all compared techniques with respect to ours, which results are displayed at the bottom left side of Fig. 8. With respect to the rate of deviation our technique displays a clear difference ranging from 299% to 8% whilst in terms of rate of improvement a range of 75% to 10%. In particular, the worst performance was reported by LRW in which we improved substantially with 299% and 75% in the rate of deviation and improvement. Our closest competitor is SNIC with which we reduce the error rate by more than 10%. This improvement comes at a negligible cost time which can be seen in Figure 9.

▷**How is the Computational Performance?** Finally, we evaluate our approach vs the SOTA models in terms of CPU performance in seconds. Results are displayed at Fig. 9, using the average time across all images and over the range of [80, 2500] superpixels. From this plot, we observe that TP, TPS and LRW require high computational time whilst QS and dSLIC sightly improve in this regard. Finally, SLIC, SNIC and OURS provide more feasible runtimes that are appropriate for a pre-processing task. We remark that our negligible computational load (∼5%) is justified by the substantially improved results over SLIC, SNIC with the same number of superpixels. Overall, the computational demand of our technique is of practical interest for several domains, where decreasing the computational load by reducing the number of primitives in the image domain is needed, including remote sensing, medical applications and video analysis.

▷**Case Study: Out-of-the Focuss and Wide Field of View Images** To further support our technique, we included two interesting yet challenging cases. The first one is superpixels with out-of-the focus images. Whilst the second case is when using wide field of view images.

We start evaluating our approach using out-of-focus images and the visual outputs of our technique and SLIC are displayed in Figs. 10 and 11. The study of this case is relevant to assess the effectiveness of our technique in retrieving fine details and rich-structures with blurry and non blurry objects in the image. A common scenario of this case is in sport events due to the fast actions in the scene. In Fig. 10, we displayed a set of outputs and in a closer look, one can see that our technique was able to preserve relevant structures. We provide zoom-in snapshots of interesting cases. For example see the face details in all image, in which our technique segmented correctly the eyes, mouth and nose. Another example is the letters in the background from the first, second and third images. We also included a set of different images with different scenarios in Fig. 11, in which previous positive effects are also observed. For example at the first image the left eye is correctly retrieved with our technique whilst SLIC failed to segmented correctly. We observed that blurry parts are also well-segmented by our technique, and in particular, also in keeping fine details. Overall, the outputs displayed in Figs. 10 and 11 suggest that our proposed method produces visually more pleasant results than SLIC.

Our second case of study is when using wide angle images. The challenging when segmenting is due to the structure of the objects become highly distorted with respect to the distance between the camera point and the object position in the scene. This makes image segmentation harder than rectilinear images. We then investigate if our technique is still efficient in terms of keeping the object shape and fine details. With this purpose, we ran our technique and SLIC over a set of image samples from [42], and the results are reported in Fig. 12. A closer look at the outputs, one can observe that our technique fulfill its purpose of avoiding over-segmenting homogeneous parts such as the sky and focus on the rich parts for example buildings or people. Visual comparison of this effect can be seen as pointed out by the red arrows. Overall, our technique shows potentials for the two investigated cases by generating better output segmentations in terms of retrieving fine details and preserving the objects shape.
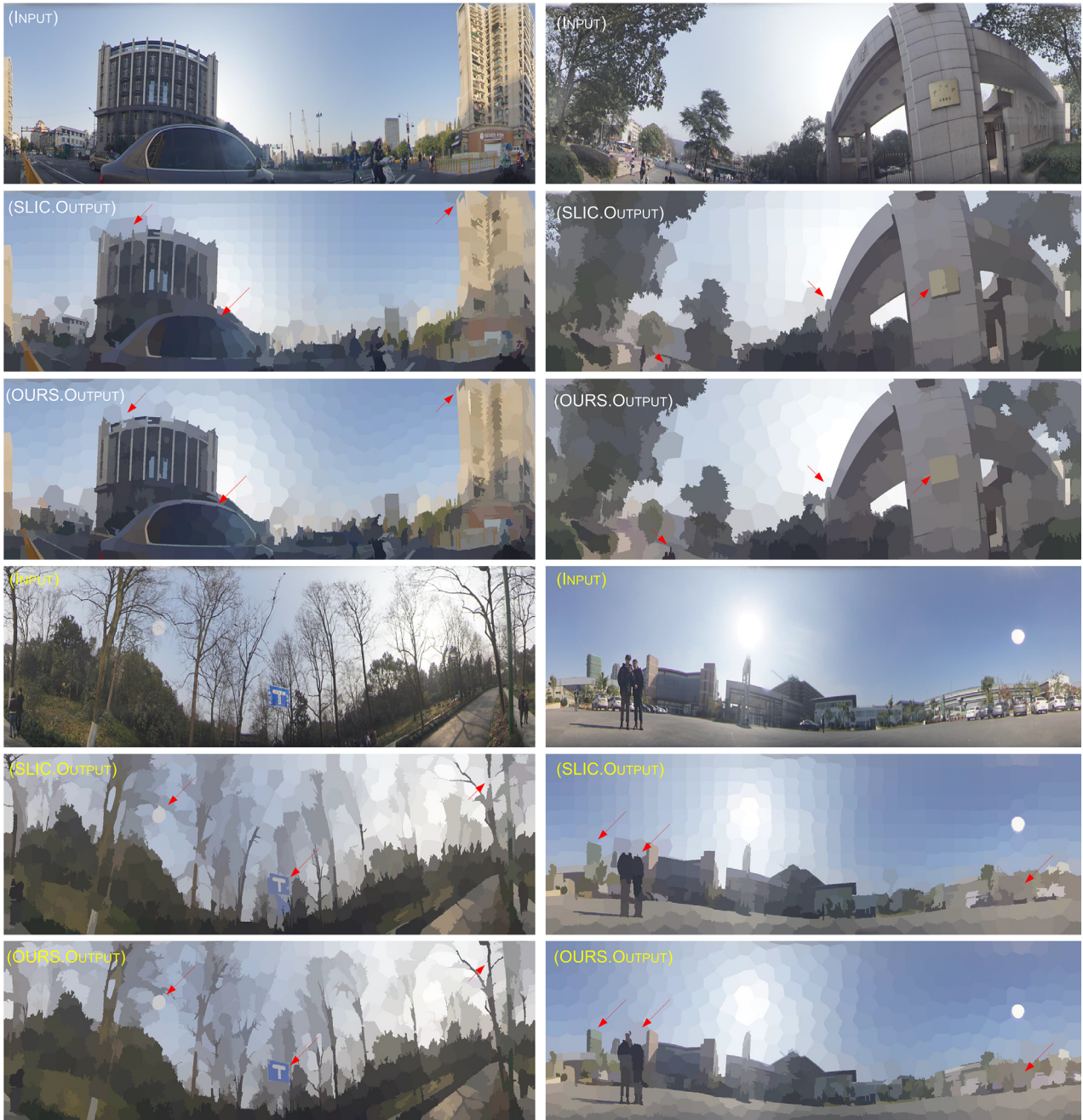
Fig. 12. Visual assessment of our techique vs SLIC using wide angle images, in which the objects are distorted and the segmentation task becomes more challenging. Red arrows show interesting areas, in which our technique improves over SLIC.

5. Conclusion

In this work, we proposed a new superpixel approach that builds upon SLIC technique. Our approach incorporates the notion of spectral residual as a proxy for object density and a novel seeding strategy. We demonstrated that our approach seeds clusters advantageously and modifies the local search radius. This leads to better segmentation, with a comparable computational load, to other state-of-the-art algorithms. These modifications leads to major advantages in terms of avoiding the over-segmentation of uniform areas in terms of fine-detail preservation, with a comparable computational load to other state-of-the-art algorithms.

We offer the following two remarks on how future work may reduce the computational load of our technique. First, we observe that our slightly higher computational load is justified by the substantially improved results over SLIC and SNIC with the same number of superpixels. This suggests that, by reducing the number of superpixels in our approach, we could achieve better segmentation than SLIC and SNIC while also reducing computational load. Secondly, we remark that the full saliency measure $\mathcal{SR}(x)$ contains more information that is strictly necessary for our technique, since it is only used as a *rough* measure of object density, while also adding substantially to the computational load. We therefore suggest that a density measure $\mathcal{G}$ could instead be constructed from a *down-sampled* version of the image so that the computational load could be significantly reduced, while maintaining excellent segmentation. For instance, down-sampling by a factor of 2 will reduce the computational load of computing $\mathcal{SR}$ by a factor of 4

and, provided that the superpixel resolution is much lower than the resolution of the true image, have minimal effect on the initial seeding. We also note that the map $\mathcal{SR}$ used here is certainly not the only possible technique for determining object density; future research may consider using other techniques, possibly based on machine learning, to replace our object density measure $\mathcal{G}$. Furthermore, future work will include performance evaluation over a range of applications, where superpixels makes significant positive effect in decreasing the computational cost. For example, in remote sensing, video analysis and medical image analysis.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

References

[1] P. Sellars, A. Aviles-Rivero, C.-B. Schönlieb, Superpixel contracted graph-based learning for hyperspectral image classification, arXiv preprint arXiv:1903.06548 (2019).

[2] C. Shi, C.-M. Pun, Multiscale superpixel-based hyperspectral image classification using recurrent neural networks with stacked autoencoders, IEEE Trans. Multimed. (TMM) 22 (2) (2019) 487–501.

[3] M. Menze, A. Geiger, Object scene flow for autonomous vehicles, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3061–3070.

[4] P. Liu, M. Lyu, I. King, J. Xu, Selflow: self-supervised learning of optical flow, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 4571–4580.

[5] R. Giraud, V.-T. Ta, N. Papadakis, Superpixel-based color transfer, in: IEEE International Conference on Image Processing (ICIP), 2017, pp. 700–704.

[6] J. Chen, J. Hou, Y. Ni, L.-P. Chau, Accurate light field depth estimation with superpixel regularization over partially occluded regions, IEEE Trans. Image Process. (TIP) 10 (2018) 4889–4900.

[7] D. Yeo, J. Son, B. Han, J. Hee Han, Superpixel-based tracking-by-segmentation using Markov chains, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1812–1821.

[8] X. Ren, J. Malik, Learning a classification model for segmentation, in: IEEE International Conference on Computer Vision, 2003, p. 10.

[9] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Süsstrunk, Slic superpixels compared to state-of-the-art superpixel methods, IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) 34 (11) (2012) 2274–2282.

[10] D. Stutz, A. Hermans, B. Leibe, Superpixels: an evaluation of the state-of-the-art, Comput. Vis. Image Underst. 166 (2018) 1–27.

[11] A. Humayun, F. Li, J.M. Rehg, The middle child problem: Revisiting parametric min-cut and seeds for object proposals, in: IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1600–1608.

[12] D. Tang, H. Fu, X. Cao, Topology preserved regular superpixel, in: IEEE International Conference on Multimedia and Expo, 2012, pp. 765–768.

[13] A. Vedaldi, S. Soatto, Quick shift and kernel methods for mode seeking, in: European Conference on Computer Vision, Springer, 2008, pp. 705–718.

[14] A. Levinshtein, A. Stere, K.N. Kutulakos, D.J. Fleet, S.J. Dickinson, K. Siddiqi, Turbopixels: fast superpixels using geometric flows, IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) 31 (12) (2009) 2290–2297.

[15] G. Maierhofer, D. Heydecker, A.I. Aviles-Rivero, S.M. Alsaleh, C.-B. Schonlieb, Peekaboo-where are the objects? structure adjusting superpixels, in: IEEE International Conference on Image Processing (ICIP), 2018, pp. 3693–3697.

[16] S. Lloyd, Least squares quantization in PCM, IEEE Trans. Inf. Theory (1982) 129–137.

[17] X. Hou, L. Zhang, Saliency detection: a spectral residual approach, in: 2007 IEEE Conference on Computer Vision and Pattern Recognition, Ieee, 2007, pp. 1–8.

[18] J. Shi, J. Malik, Normalized cuts and image segmentation, Dep. Pap. (CIS) (2000) 107.

[19] J. Shen, Y. Du, W. Wang, X. Li, Lazy random walks for superpixel segmentation, IEEE Trans. Image Process. 4 (2014) 1451–1462.

[20] M.-Y. Liu, O. Tuzel, S. Ramalingam, R. Chellappa, Entropy rate superpixel segmentation, in: CVPR 2011, IEEE, 2011, pp. 2097–2104.

[21] P.F. Felzenszwalb, D.P. Huttenlocher, Efficient graph-based image segmentation, Int. J. Comput. Vis. 59 (2) (2004) 167–181.

[22] A.P. Moore, S.J. Prince, J. Warrell, U. Mohammed, G. Jones, Superpixel lattices, in: 2008 IEEE Conference on Computer Vision and Pattern Recognition, Citeseer, 2008, pp. 1–8.

[23] H. Fu, X. Cao, D. Tang, Y. Han, D. Xu, Regularity preserved superpixels and supervoxels, IEEE Trans. Multimed. 16 (4) (2014) 1165–1175.

[24] M.A. Condori, F.A. Cappabianco, A.X. Falcão, P.A. Miranda, An extension of the differential image foresting transform and its application to superpixel generation, J. Vis. Commun. Image Represent 71 (2020) 102748.

[25] D. Comaniciu, P. Meer, Mean shift: a robust approach toward feature space analysis, IEEE Trans. Pattern Anal. Mach. Intell. 24 (5) (2002) 603–619.

[26] L. Vincent, P. Soille, Watersheds in digital spaces: an efficient algorithm based on immersion simulations, IEEE Trans. Pattern Anal. Mach. Intell. 13 (6) (1991) 583–598.

[27] S. Wolf, A. Bailoni, C. Pape, N. Rahaman, A. Kreshuk, U. Köthe, F.A. Hamprecht, The mutex watershed and its objective: efficient, parameter-free image partitioning, arXiv preprint arXiv:1904.12654 (2019).

[28] W. He, C. Li, Y. Guo, Z. Wei, B. Guo, A two-stage gradient ascent-based superpixel framework for adaptive image segmentation, Appl. Sci. 9 (12) (2019) 2421.

[29] J. Wang, X. Wang, Vcells: simple and efficient superpixels using edge-weighted centroidal voronoi tessellations, IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) 34 (6) (2012) 1241–1247.

[30] J. Papon, A. Abramov, M. Schoeler, F. Worgotter, Voxel cloud connectivity segmentation-supervoxels for point clouds, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 2027–2034.

[31] P. Neubert, P. Protzel, Compact watershed and preemptive SLIC: on improving trade-offs of superpixel segmentation algorithms, in: International Conference on Pattern Recognition, 2014, pp. 996–1001.

[32] Z. Li, J. Chen, Superpixel segmentation using linear spectral clustering, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1356–1363.

[33] R. Achanta, S. Susstrunk, Superpixels and polygons using simple non-iterative clustering, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 4651–4660.

[34] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, IEEE Trans. Pattern Anal. Mach. Intell. 40 (4) (2017) 834–848.

[35] D. Marmanis, K. Schindler, J.D. Wegner, S. Galliani, M. Datcu, U. Stilla, Classification with an edge: improving semantic image segmentation with boundary detection, ISPRS J. Photogramm. Remote Sens. 135 (2018) 158–172.

[36] X. Liu, Z. Deng, Y. Yang, Recent progress in semantic image segmentation, Artif. Intell. Rev. 52 (2) (2019) 1089–1106.

[37] C. Yang, L. Zhang, H. Lu, X. Ruan, M.-H. Yang, Saliency detection via graph-based manifold ranking, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3166–3173.

[38] G. Li, Y. Yu, Deep contrast learning for salient object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 478–487.

[39] N. Tong, H. Lu, X. Ruan, M.-H. Yang, Salient object detection via bootstrap learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1884–1892.

[40] P. Arbelaez, M. Maire, C. Fowlkes, J. Malik, Contour detection and hierarchical image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2010) 898–916.

[41] J. Shi, L. Xu, J. Jia, Discriminative blur detection features, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 2965–2972.

[42] K. Yang, X. Hu, L.M. Bergasa, E. Romera, K. Wang, Pass: panoramic annular semantic segmentation, IEEE Trans. Intell. Transp. Syst. 21 (2019) 4171–4185.

Jianchao Zhang is an undergrad student at City University of Hong Kong in Mathematics. He was a visitor at the University of Cambridge in the summer of 2019. His areas of interest include computational mathematics and image processing.

Angelica I. Aviles received the Ph.D. degree (2017) from the Polytechnic University of Catalonia, Spain. She is currently a Research Associated in the Department of Pure Mathematics & Mathematical Statistics (DPMMS), University of Cambridge,

UK. Her research lies at the intersection of computational mathematics, computer vision and machine learning for applications to large-scale real world problems. Her central research is to develop new data- driven algorithmic techniques that allow computers to gain high-level understanding from vast amounts of data, this, with the aim of aiding the decisions of users from multiple-disciplines. This line of research has allowed her to gain expertise with a wide range of various data types including medical imaging, computational photography, computer graphics and remote sensing to new a few. She is an IEEE, SIAM, CVF and ACM member.

Daniel Heydecker received the B.A. and master's degrees in mathematics from the University of Cambridge in 2017, where he is currently pursuing the Ph.D. degree with the Cambridge Centre for Analysis, Department of Pure Mathematics and Mathematical Statistics (DPMMS), with a focus on the probability of systems with many degrees of freedom.

Xiaosheng Zhuang received his bachelor's degree and master's degree in mathematics from Sun Yat-Sen (Zhongshan) University, China, in 2003 and 2005, respectively. He received his Ph.D. in applied mathematics from University of Alberta, Canada, in 2010. He was a Postdoctoral Fellow at Universität Osnabrck in 2011 and Technische Universitt Berlin in 2012. His research interest includes directional multiscale representation systems, image/signal processing, and compressed sensing

Raymond Chan graduated with First Class Honors from the Department of Mathematics at The Chinese University of Hong Kong in 1980. Uncertain of what to do next, he stayed in the Department as a full-time instructor after graduation. He started his graduate study in 1981 with a full fellowship from the Courant Institute of Mathematical Sciences at New York University. He obtained the M.Sc. and Ph.D. degrees in Applied Mathematics there in 1984 and 1985 respectively under the supervision of Professor Olof Widlund. Chan began his career as a tenure-track Assistant Professor at the University of Massachusetts at Amherst in 1985. With heart and mind always in Hong Kong, he came back to Hong Kong in 1986, first at The University of Hong Kong (1986-92) and then at The Hong Kong University of Science and Technology (1993) before joining his Alma Mater in 1993. He was the Associate Director of the Institute of Mathematical Sciences (1996-98), the Associate Dean of Science (2004–2009) and the Head of the Mathematics Department (2012–2018). He has retired from the university in 2019 as an Emeritus Professor and is now the Dean of College of Science at City University of Hong Kong. Chan has published 140 journal papers and has been in the ISI Science Citation List of Top Highly-Cited Mathematicians in the world (2001 List). He won a Leslie Fox Prize for Numerical Analysis in 1989 at Cambridge, United Kingdom; a Feng Kang Prize of Scientific Computing in 1997 in Beijing, China; a Morningside Award in 1998 in Beijing, China; and 2011 Higher Education Outstanding Scientific Research Output Awards (First Prize) from the Ministry of Education in China. He was elected a SIAM Fellow in 2013 and a SIAM Council Member for 2015-20.

Carola-Bibiane Schönlieb received the B.S. degree from the Institute for Mathematics, University ofSalzburg, Salzburg, Austria, in 2004, and the Ph.D.degree from the University of Cambridge, Cam-bridge, U.K., in 2009.From 2004 to 2005, she held a teaching positionin Salzburg. She held a post-doctoral position atthe University of Gttingen, Gttingen, Germany,for one year. She became a Lecturer with theDepartment of Applied Mathematics and Theoret-ical Physics (DAMTP), University of Cambridge,in 2010, where she was promoted to a Reader in 2015 and a Professorin 2018. Since 2011, she has been a fellow of the Jesus College, University ofCambridge, where she is currently a Professor of applied mathematics with DAMTP, the Head of the Cambridge Image Analysis Group, the Directorof the Cantab Capital Institute for Mathematics of Information, and the Director of the Engineering and Physical Sciences Research Council Center for Mathematical and Statistical Analysis of Multimodal Clinical Imaging. Herresearch interests include variational methods, partial differential equations,and machine learning for image analysis, image processing, and inverseimaging problems.