Dynamic Spectral Residual Superpixels

Jianchao Zhang[1*] , Angelica I. Aviles-Rivero[2*], Daniel Heydecker[2*] ,
Xiosheng Zhuang[1], Raymond Chan[1] and Carola-Bibiane Schönlieb[2]
[1] City University of Hong Kong (CityU), Hong Kong.
[2] DPMMS and DAMTP, University of Cambridge, UK.
{jzhang537-c@my.cityu.edu.hk}, {ai323,dh489,cbs31}@cam.ac.uk,
{xzhuang7,rchan.sci}@cityu.edu.hk

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, while 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.

Keywords: Superpixels, K-means, Spectral Residual, Segmentation

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, 7] and tracking [8,

*Equal Contribution.

Figure 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.

9]. 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 [10, 11, 12].

Since the pioneering work of Ren and Malik [10], the community has devoted to develop different algorithmic approaches to improve over [10]. These approaches can be roughly divided in: graph-based methods e.g. [10, 13], path-based approaches e.g. [14], density-based models e.g. [15], contour models e.g. [16] and clustering methods e.g. [11, 17].

Out of all of the approaches reported in the literature, the Simple Linear Iterative Clustering (SLIC) [11] is perhaps the most popular method that offers a good performance whilst demanding low computational cost, by building on Lloyd's algorithm [18] for k-means. The central idea of SLIC is to perform the superpixels partition based on an iterative scheme that search for similarities between points ensuring at each step that 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, we propose a new algorithmic approach, exhibited in Fig. 1, that improves upon the SLIC approach, motivated by the drawbacks discussed above. We show that our approach outperforms 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. [19] 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 mitigates 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 and the partition is based on the similarity of the nodes, e.g. colour, including [10, 20, 21, 22, 23, 24]. 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 [15]. However, there is not 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 the well-known Dijkstra algorithm, as reported in [14, 25]. 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 *gradient ascent* and other geometric methods [26, 27, 16]. For an extensive review of the literature, we refer the reader to [12].

In particular, in this work we focus on, probably, the most popular superpixel category, which is clustering based approaches. The basis of this perspective builds on Lloyd's algorithm [18] 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. [11], 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 [28, 29, 30, 31]. Most recently, in [32] 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 [17] propose 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.

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. [19], which inspires our current approach. More recent works include techniques based on graphs [24] or machine-learning [33, 34].

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 A and height B, as a map $I : \mathcal{X} \to \Omega$, where $\mathcal{X} = [\mathcal{A}] \times [\mathcal{B}]$ is a rectangular domain, and $\Omega \subset \mathbb{R}^3$ 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 color space). *The Lab color space L_{ab} describes mathematically all perceivable colors in the three dimensions l for lightness and a and b for the color 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 $\mathbf{p}_1, \mathbf{p}_2 \in \mathcal{X} \times D, \mathbf{p}_i = [\mathbf{x}_i, \mathbf{l}_i]^T, \mathbf{x}_i \in \mathcal{X}, \mathbf{l}_i \in D$ define:

Figure 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.

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

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 [17] 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 [19] as a better measure of object detection.

We briefly review the Fourier analysis leading to the definition of the spectral residual in [19]. 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

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

where $\mathfrak{R}$ denotes the real part; we also write $\mathcal{P}(f) = \mathfrak{I}(\mathfrak{F}I)(f)$ for the imaginary part, or phase spectrum. The key insight of [19] 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 nonnegative, and the convolution with a Gaussian kernel g_σ ensures that the final result is smooth. For our purposes, we found that in practice $\sigma = 20$ is an excellent value. 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:

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

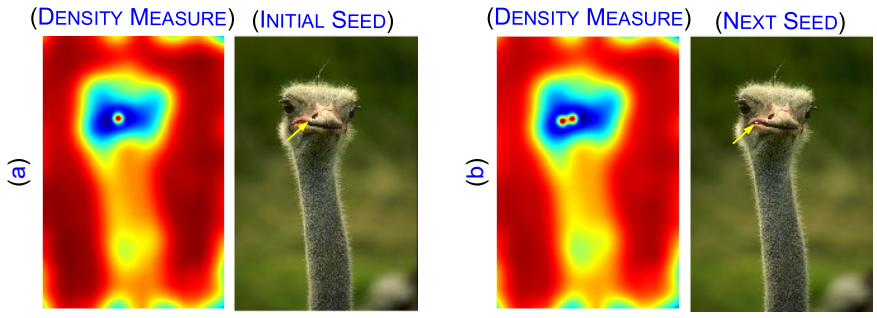

Figure 3: Illustration of our initialisation strategy, which incorporates the object density measure $\mathcal{G}$, for two initial seeds. From left to right, first seed and second one.

By incorporating our proposed function, which we use as a measure of object density, into the distance computation one obtains have two major advantages. 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 the images into unnecessary small superpixels. This effect is illustrated in Fig. 2, for example, at the second column where the 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.

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.

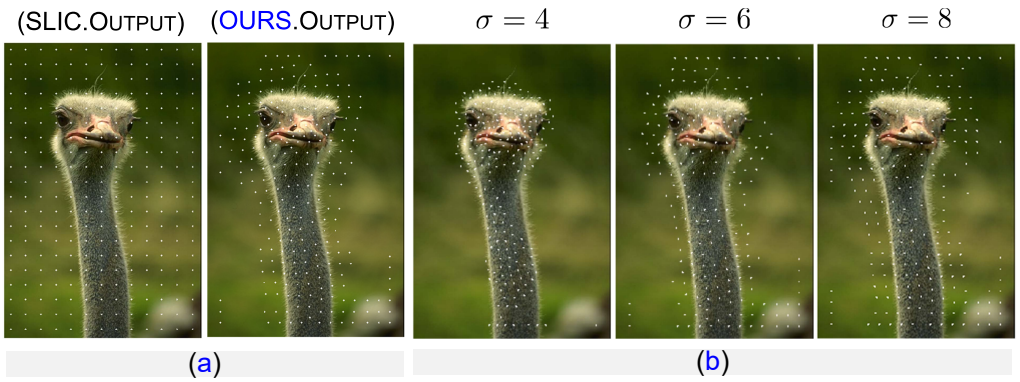

Figure 4: Visual comparison of the seeding initialisation of SLIC vs ours.(a) One can see that we seek to focus on relevant areas (i.e. other than background). (b) The effect of τ in our seeding strategy.

```
#pseudocode for seeding initialisation SLIC
Set: Initialise cluster centers as
C_i = [x_i, y_i, l_i, a_i, b_i] by sampling at regular grid
step: S = sqrt(N/k)
#where N is the size of the image
Move cluster centers to the lowest gradient
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.

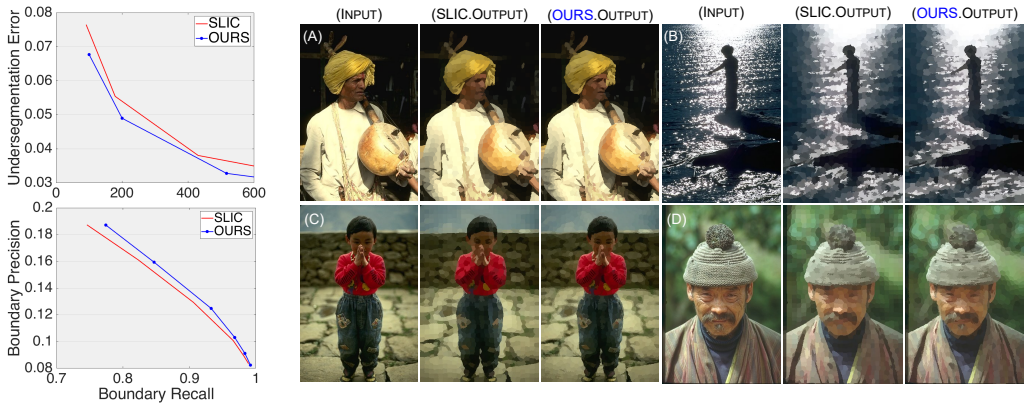

Figure 5: From left to right. Quantitative comparison of our approach vs SLIC using three metrics: UE, BR and BP. Our approach reported the best scores metric-wise. Four visual outputs comparisons of SLIC vs our approach. In a closer look, one can see that our approach achieves better connection of structures and keeps fine-details. For example, see (A), (C) and (D) the faces and (B) the hand.

The 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.

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

An output example is displayed in Fig. 4. Subfigure (a) shows a 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 practise, we found that the $\tau = 13/2$ works for a range of images.

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 [35], which provides ground truth of the images for quantitative analysis.

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 [11]; 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 compared to state-of-the-art techniques: QS [15], TP [16] ,TPS [14], LRW [20], SNIC [32] and dSLIC [17]. 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.

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

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* $B_i, i = 1, .., M$*. 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 [15], TP [16], TPS [14], SLIC [11], LRW [20], SNIC [32] and dSLIC [17], we set the parameters as suggested in the corresponding work. We also used the codes realease 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 CPU-based implementation. We used an Intel Core i7 with 16GB RAM.

4.2. Results and Discussion

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

▷ **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. 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 [15], TP [16], TPS [14], LRW [20], SNIC [32] and dSLIC [17]. 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 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

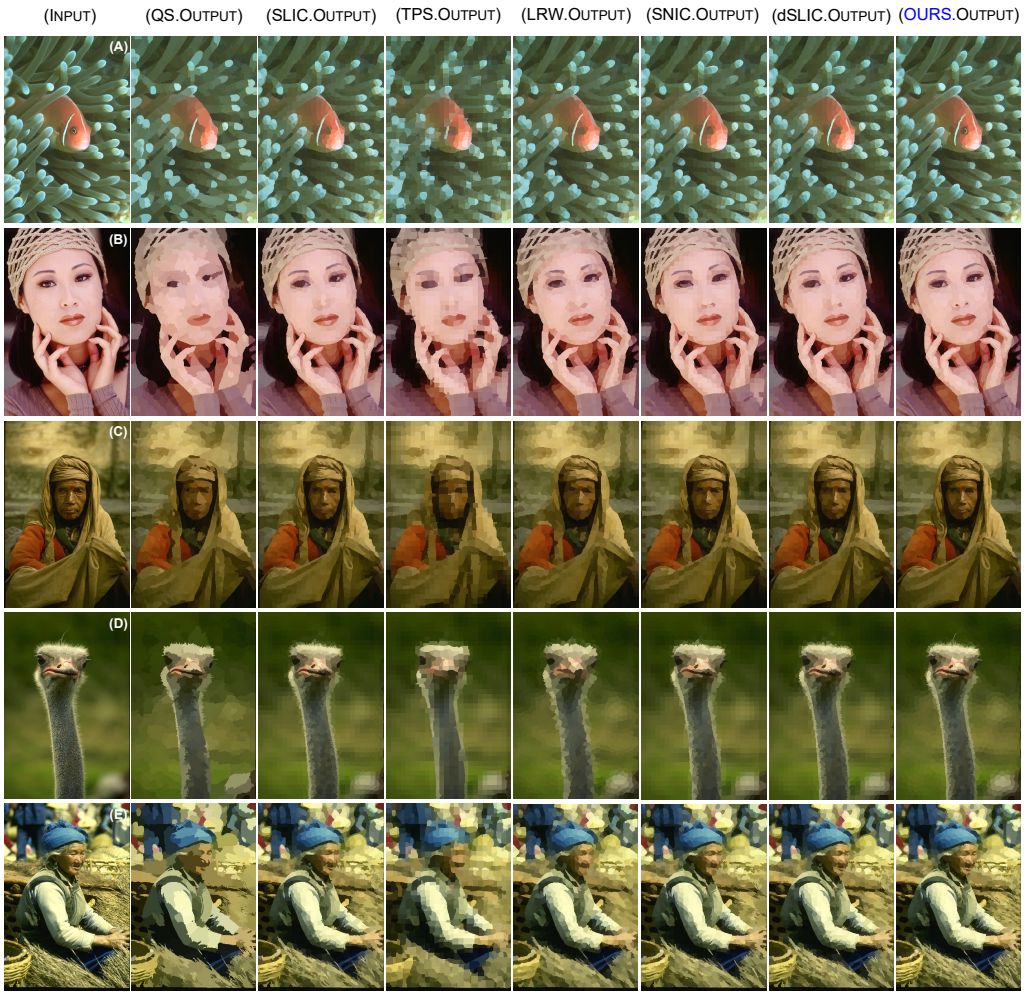

Figure 6: Superpixel outputs comparisons of our approach vs different methods from the body of literature: QS [15], SLIC [11] TP [16], TPS [14], LRW [20], SNIC [32] and dSLIC [17]. A closer inspection, one can see that our approach offers better superpixels segmentation. For example, (A), (B), and (C) the eyes; (D) the ostrich's boundary and (E) the eyes and basket.

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

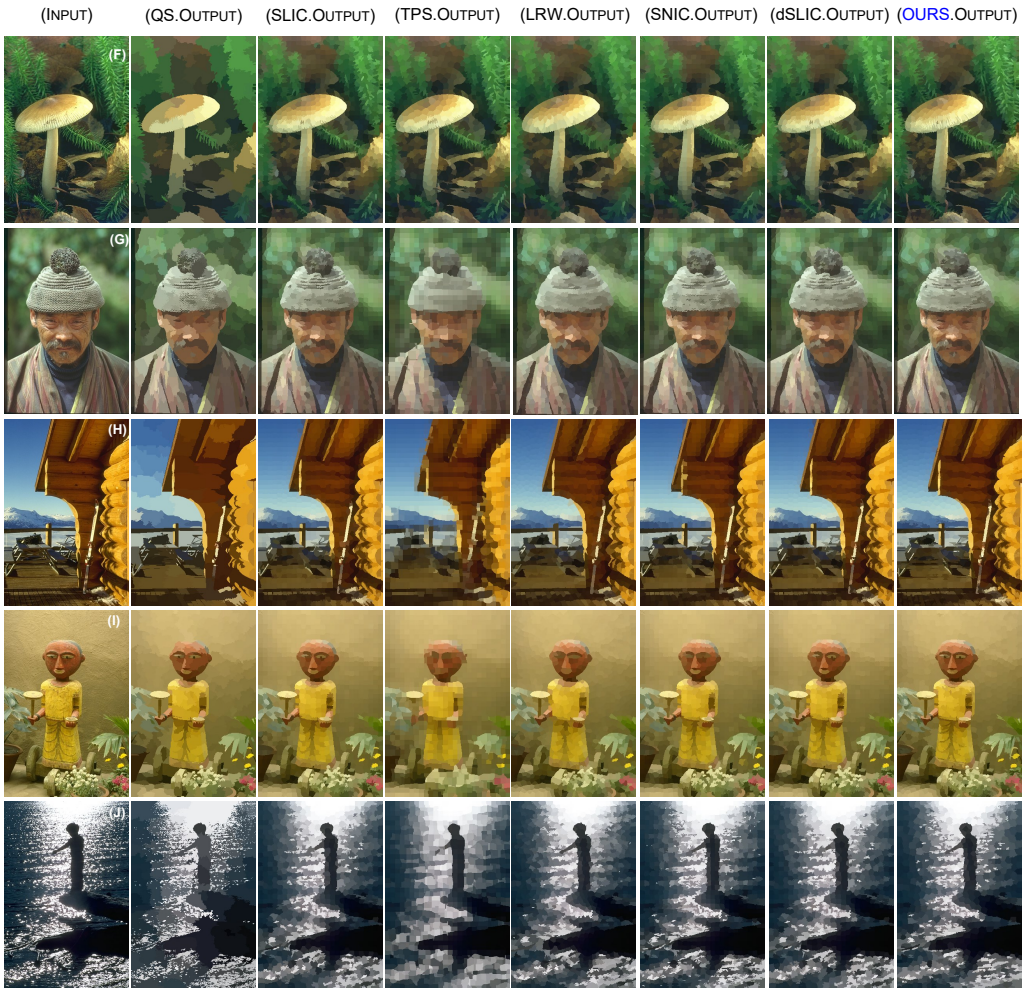

Figure 7: Superpixel segmentation outputs of our approach vs QS [15], SLIC [11] TP [16], TPS [14], LRW [20], SNIC [32] and dSLIC [17]. Visual assessment shows that the proposed algorithm performs better than the compared approaches. Examples are: (F) the leaves; (G) and (I) the face; (H) the house boundaries and (J) the hand.

terms of grouping structures correctly than the previous 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.

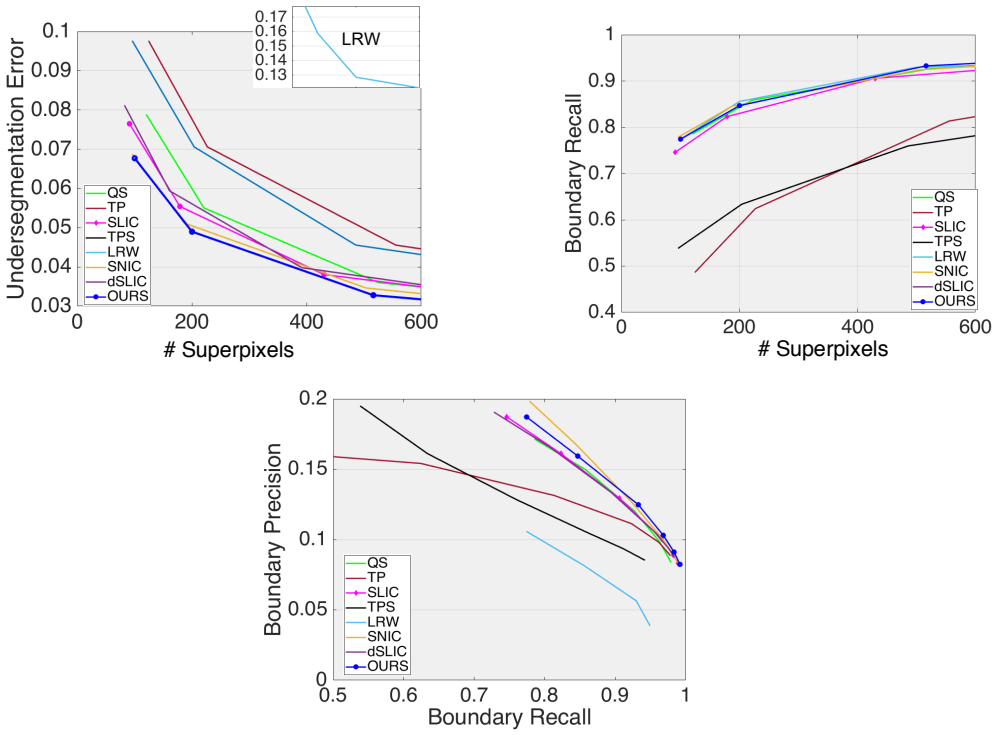

Figure 8: Metric-wise comparison of our approach vs SOTA techniques using UE, BR and BP. In a closer look, we can see that our approach, overall, offers the lowest UE and the highest BR. Finally, the good boundary adherence to the true edges is reflected in the last plot, in which our approach overall gets the best trade-off between those metrics.

These 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 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.

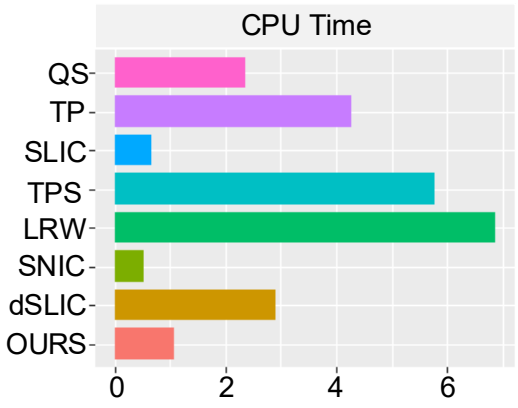

Figure 9: CPU time averaged comparison of our approach vs the body of literature. One can see that our approach improvement comes at a negligible cost in runtime in comparison with the fastest approaches SLIC and SNIC.

▷ **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 Figure 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 preprocessing task. We remark that our slightly higher computational load is justified by the substantially improved results over SLIC, SNIC with the same number of superpixels

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 modify 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 over-segmenting uniform areas into more superpixels than necessary and preserving fine-details in the image, with a comparable computational load to other state-of-the-art algorithms.

Future work includes to reduce the sightly increment in computational time. Since our technique has slightly higher run-time than other approaches, we offer the following two remarks on how this may be mitigated. First, we observe that our slightly higher computational load is justified by the substantially improved results over SLIC, 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, 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 adding to the computational load. We therefore suggest that, if the density measure $\mathcal{G}$ were constructed instead from a downsampled version of the image, that the computational load could be significantly reduced, while maintaining excellent segmentations.

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 Transactions on Multimedia (TMM) (2019).

[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] F. Liu, C. Shen, G. Lin, Deep convolutional neural fields for depth estimation from a single image, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 5162–5170.

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

[8] S. Wang, H. Lu, F. Yang, M.-H. Yang, Superpixel tracking, in: International Conference on Computer Vision (CVPR), 2011, pp. 1323–1330.

[9] 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.

[10] X. Ren, J. Malik, Learning a classification model for segmentation, in: null, 2003, p. 10.

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

[12] D. Stutz, A. Hermans, B. Leibe, Superpixels: An evaluation of the state-of-the-art, Computer Vision and Image Understanding 166 (2018) 1–27.

[13] 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.

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

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

[16] A. Levinshtein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, K. Siddiqi, Turbopixels: Fast superpixels using geometric flows, IEEE

Transactions on Pattern Analysis and Machine Intelligence (PAMI) (2009).

[17] 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.

[18] S. Lloyd, Least squares quantization in pcm, IEEE Transactions on Information Theory (1982) 129–137.

[19] 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.

[20] J. Shen, Y. Du, W. Wang, X. Li, Lazy random walks for superpixel segmentation, IEEE Transactions on Image Processing (2014) 1451–1462.

[21] J. Shi, J. Malik, Normalized cuts and image segmentation, Departmental Papers (CIS) (2000) 107.

[22] P. F. Felzenszwalb, D. P. Huttenlocher, Efficient graph-based image segmentation, International Journal of Computer Vision (2004).

[23] 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.

[24] 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.

[25] H. Fu, X. Cao, D. Tang, Y. Han, D. Xu, Regularity preserved superpixels and supervoxels, IEEE Transactions on Multimedia 16 (2014) 1165–1175.

[26] D. Comaniciu, P. Meer, Mean shift: A robust approach toward feature space analysis, IEEE Transactions on Pattern Analysis & Machine Intelligence (2002) 603–619.

[27] L. Vincent, P. Soille, Watersheds in digital spaces: an efficient algorithm based on immersion simulations, IEEE Transactions on Pattern Analysis & Machine Intelligence (1991) 583–598.

[28] J. Wang, X. Wang, Vcells: Simple and efficient superpixels using edge-weighted centroidal voronoi tessellations, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) (2012).

[29] 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.

[30] 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.

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

[32] 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.

[33] 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.

[34] 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.

[35] P. Arbelaez, M. Maire, C. Fowlkes, J. Malik, Contour detection and hierarchical image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence (2010) 898–916.