Fast and Effective Superpixel Segmentation using Accurate Saliency Estimation

Felipe Belém, Isabela Borlido, Leonardo João, Benjamin Perret, Jean Cousty, Silvio Guimarães, Alexandre Falcão

Fast and Effective Superpixel Segmentation using Accurate Saliency Estimation*

Felipe Belém[1,3][0000−0002−6037−5977], Isabela Borlido[2][0000−0001−7288−2485], Leonardo João[1][0000−0003−4625−7840], Benjamin Perret[3][0000−0003−0933−8342], Jean Cousty[3][0000−0002−2163−9714], Silvio J. F. Guimarães[2][0000−0001−8522−2056], and Alexandre Falcão[1][0000−0002−2914−5380]

[1] University of Campinas, Campinas, Brazil
{felipe.belem,leonardo.joao,afalcao}ic.unicamp.br
[2] Pontifical Catholic University of Minas Gerais, Belo Horizonte, Brazil
isabela_borlido@hotmail.com,sjamil@pucminas.br
[3] Université Gustave Eiffel, Noisy-le-Grand, France
{felipe.belem,benjamin.perret,jean.cousty}@esiee.fr

Abstract. *Superpixels through Iterative CLEarcutting* (SICLE) is a recently proposed framework for superpixel segmentation. SICLE consists of three steps: (i) seed oversampling; (ii) superpixel generation; and (iii) seed removal; such that, after step (i), steps (ii) and (iii) are repeated until a desired number of superpixels is obtained. Such pipeline showed effective and efficient multiscale superpixel segmentation. Furthermore, if an object is desired, it is possible to improve delineation by providing its probable location, often called saliency. While classical methods estimate object saliency by contrast-based criteria, recent ones use deep-learning strategies for accurate estimation. SICLE shows robustness for low-quality saliency estimations, but it struggles to effectively take advantage of the high-quality ones. In this work, we propose a generalization of its path-cost function and seed removal criterion (steps (ii) and (iii), respectively), adapting SICLE to a given saliency map. By choice of a binary parameter, SICLE can take advantage of low- and high-quality saliency maps for better segmentation. Results show that, by exploiting the accurate information of the saliency map, our improved SICLE version surpasses state-of-the-art methods in traditional delineation metrics while requiring only two iterations for segmentation, being significantly faster than its predecessor and SLIC.

Keywords: Superpixel Segmentation · Object Saliency Map · Image Foresting Transform.

* The authors thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq – (Universal 407242/2021-0, PQ 303808/2018-7, 310075/2019-0), the Fundação de Amparo a Pesquisa do Estado de Minas Gerais – FAPEMIG – (PPM-00006-18), the Fundação de Amparo a Pesquisa do Estado de São Paulo – FAPESP – (2014/12236-1) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES – Finance code 001 (COFECUB 88887.191730/2018-00) for the financial support.

1 Introduction

Superpixels represent homogeneous regions that contain a perceptual meaning and provide more information than pixels. Although some authors raise compactness and regularity as indicators of high-quality superpixel segmentation, boundary adherence, efficiency, and controllable quantity of superpixels, are indispensable for any method [23,25]. From that, several applications such as object tracking [10], semantic segmentation [33], and image classification [21] exploit their properties.

Classical approaches differ in their strategy for superpixel generation and generally consider only color and spatial position to measure superpixel similarity, without any prior information. We may cite *Simple Linear Iterative Clustering* (SLIC) [1] as an example of a clustering-based method given its adapted K-means strategy in a 5-dimensional feature space. Conversely, graph-based approaches, such as *Entropy Rate Superpixels* (ERS) [16], have higher boundary adherence but they are often slow. Also, *Superpixel Hierarchy* (SH) [26] and *Waterpixels* [17] are effective hierarchical graph-based examples whose drawback is error propagation to coarser scales. Finally, *Dynamic and Iterative Spanning Forest* (DISF) [6] is a path-based method that applies oversampling and iteratively generates superpixels on refined seed sets.

Using local information without any prior or high-level knowledge may be insufficient to obtain a good delineation in images with complex characteristics, such as textured or noisy images [28]. More recent approaches circumvent this drawback with Deep Learning architectures [29,14] or by including high-level information, such as texture [31] and gradient mask [28]. However, they present moderate delineation, and their constraints are the regular grid shape on standard convolution operations for deep-learning-based methods, high computational time, and the lack of superpixel groundtruth [29].

Although using saliency in segmentation is not a novel strategy [12], it has not been thoroughly exploited for generating superpixels until recently. In [32], the authors proposed a SLIC-based algorithm that uses a saliency map based on the Fourier Transform for generating more superpixels in textured regions. The *Object-based DISF* (ODISF) [3] method is another example that extends DISF for incorporating object saliency maps. However, since the map's influence is not controllable in both, higher-quality saliency may not promote higher-quality delineation. Conversely, *Object-based ISF* (OISF) [4,5] overcomes this issue by allowing user control over the saliency influence during delineation, but it is slow and highly sensitive to incorrect estimations.

A recent proposal named *Superpixels through Iterative CLEarcutting* (SICLE) [8] generalizes ODISF for allowing user control over the number of iterations for segmentation, being more efficient than SLIC at generating superpixels in different experiments. The SICLE pipeline is composed of three steps: (i) seed oversampling; (ii) superpixel generation by the *Image Foresting Transform* (IFT) [11]; and (iii) object-based seed removal. After step (i), steps (ii) and (iii) are performed until obtaining a desired number of superpixels. By using object information only in the last step, SICLE delineation performance

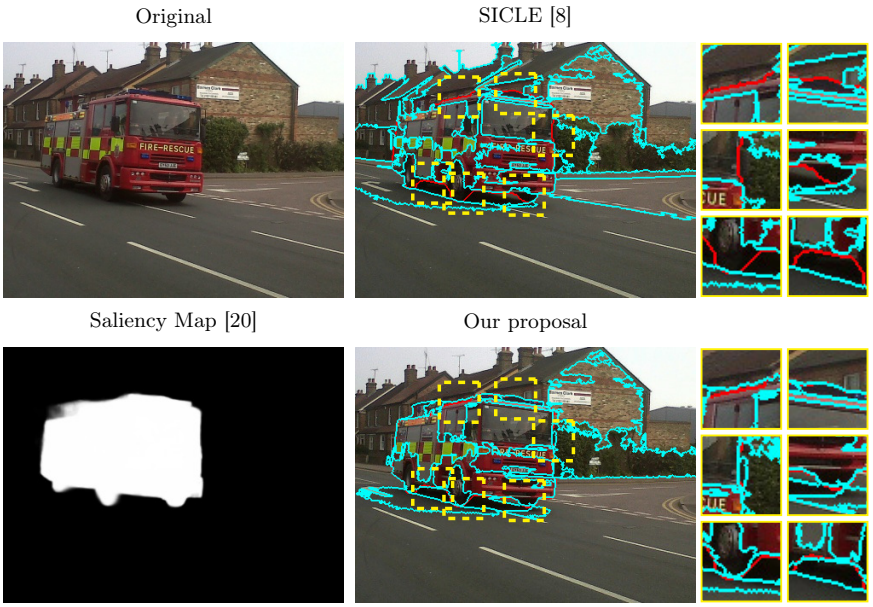

Fig. 1. Comparison between SICLE and our improved version for accurate maps considering 25 superpixels. Red lines indicate object boundaries, whereas cyan ones, superpixel borders. Yellow rectangles indicate delineation errors that our approach overcame.

is robust to incorrect estimations, contrasting with OISF. However, similarly to ODISF, it cannot improve its delineation performance for higher-quality saliency maps.

One may argue that an object-based method should exploit the prior object location information with respect to its quality. For high-quality information, the approximation to the object boundaries can assist its delineation. Conversely, although low-quality information poorly estimates the object boundaries, it presents valuable information on its location. Therefore, in this work, we propose a generalization of SICLE's path-cost function and object-based seed removal criteria (*i.e.*, steps (ii) and (iii), respectively) for exploiting low- and high-quality saliency maps to improve segmentation results by choice of a binary parameter. Experimental results show that our proposal, named SICLE$^{\alpha}$, is robust to low-quality maps and improves delineation in the case of high-quality ones, as exemplified in Figure 1. Moreover, by exploiting the accurate estimation of the object boundaries, SICLE$^{\alpha}$ achieves higher precision in step (iii), thus requiring only two iterations for effective object delineation. Given both, our method surpasses state-of-the-art methods in terms of efficiency and effectiveness, considering classical evaluation metrics.

This paper is organized as follows. The related definitions are presented in Section 2 and our proposal is described in Section 3. Section 4 presents the

experiments, with an ablation study, and qualitative and quantitative evaluation. Finally, the conclusion and future work are presented in Section 5.

2 Theoretical Background

In this section, we present the theoretical background for our work. In Section 2.1, we discuss basic concepts in image and graphs and, subsequently, in Section 2.2, we present the *Image Foresting Framework* (IFT) [11].

2.1 Image and Graphs

An *image* I is a pair $\langle \mathcal{P}, \mathbf{F} \rangle$ in which $\mathbf{F}(p) \in \mathbb{R}^m$ maps the *features* of every *picture element* (*i.e.*, pixel) $p \in \mathcal{P} \subset \mathbb{Z}^2$, given $m \in \mathbb{N}^*$. I is either *colored* or *grayscale* whenever $m > 1$ or $m = 1$, respectively. An *object saliency map* $O = \langle \mathcal{P}, \mathbf{O} \rangle$ is an instance of the latter in which $\mathbf{O}(p) \in [0, 1]$ maps p to its probability of belonging to an object of interest (*i.e.*, saliency). Finally, for a set of pixels $X \subseteq \mathcal{P}$, we may compute its *mean feature* $\overline{\mathbf{F}}(X) = \sum_{x \in X} \mathbf{F}(x)/|X|$ and *mean saliency* $\overline{\mathbf{O}}(X) = \sum_{x \in X} \mathbf{O}(x)/|X|$.

From I, we may build a *directed graph* (*i.e.*, digraph) $G = \langle \mathcal{V}, \mathcal{A} \rangle$ so that $\mathcal{V} \subseteq \mathcal{P}$ contains its *vertices* and $\mathcal{A} \subset \mathcal{V}^2$ its *arcs*. The existence of an arc $\langle x, y \rangle \in \mathcal{A}$ indicates that x is *adjacent* to y. Often, $\mathcal{A}$ is defined by the 8-adjacents of every pixel $x \in \mathcal{P}$, such that $\mathcal{A} = \left\{ \langle x, y \rangle : \|x - y\|_2 \leq \sqrt{2} \right\}$. In this work, $\mathcal{A}$ holds no self-loops nor parallel edges (*i.e.*, G is a *simple graph*).

A *(directed) path* $\rho = \langle v_1, \ldots, v_k \rangle$ is a sequence of $k \in \mathbb{N}^*$ distinct adjacent vertices (*i.e.*, $\langle v_i, v_i + 1 \rangle \in \mathcal{A}$ for $i < k$). If $k = 1$, ρ is said to be *trivial*, and *non-trivial* otherwise. In ρ, we term v_i as the *predecessor* of v_{i+1} and the *successor* v_{i-1} given $1 < i < k$. Moreover, we may exhibit the *root* v_1 and the *terminus* v_k of ρ either by $\rho_{v_1 \rightsquigarrow v_k}$ or by ρ_{v_k} whenever v_1 is irrelevant for the context. For instance, $\rho_y = \rho_x \odot \langle x, y \rangle$ denotes the path ρ_y resultant from concatenating ρ_x with $\langle x, y \rangle$.

2.2 Image Foresting Transform

The *Image Foresting Transform* (IFT) [11] is a framework whose effectiveness in object delineation has been reported in several works [13,6,3,9]. When a set of representative vertices (*i.e.*, *seeds*) $\mathcal{S} \subset \mathcal{V}$ is provided, the algorithm builds trees with optimum path-cost from their seed $s \in \mathcal{S}$ to any $p \in \mathcal{V} \setminus \mathcal{S}$ through path concatenation.

We can estimate the cost of an arc $\langle x, y \rangle \in \mathcal{A}$ by an *arc-cost function* $\mathbf{w}_*(x, y) \in \mathbb{R}$ and, likewise, the cost of any path ρ_x can be computed by an *path-cost function* $\mathbf{f}_*(\rho_x) \in \mathbb{R}_+$. As an example, the $\mathbf{f}_{max}$ (Equation 1) is commonly used in IFT-based methods due to its effectiveness [3,6] in delineating objects:

$$\begin{aligned} \mathbf{f}_{max}(\langle x \rangle) &= \begin{cases} 0, & \text{if} x \in \mathcal{S} \\ +\infty, & \text{otherwise} \end{cases} \\ \mathbf{f}_{max}(\rho_x \odot \langle x, y \rangle) &= \max\left\{ \mathbf{f}_{max}(\rho_x), \mathbf{w}_*(x, y) \right\} \end{aligned} \tag{1}$$

If $\mathbf{f}_*(\rho_x) \leq \mathbf{f}_*(\tau_x)$, considering $\tau_x \in \mathscr{P}$ to be any other path reaching x within the set $\mathscr{P}$ of all possible paths in G, then ρ_x is *optimum*.

The IFT minimizes a *cost map* $\mathbf{C}(x) = \min_{\rho_x \in \mathscr{P}} \{\mathbf{f}_*(\rho_x)\}$ by assigning an optimum path ρ_x from a seed to $x \in \mathcal{V} \setminus \mathcal{S}$. Simply put, the method builds trees whose paths to non-seed vertices are more closely connected to its seed than to any other using a generalization of the Dijkstra's shortest-path algorithm. Even if $\mathbf{f}_*$ is not smooth, it still exhibits properties suitable for segmentation [18]. While minimizes $\mathbf{C}$, the algorithm builds an acyclic map $\mathbf{P}$ (*i.e.*, *predecessor map*) which assigns x either to its predecessor defined in ρ_x or to a distinctive marker $\blacktriangle \notin \mathcal{V}$, whenever x is the root of ρ_x and, thus, of $\mathbf{P}$. As one may see, it is possible to map x to its root $\mathbf{R}(x) \in \mathcal{S}$ recursively through $\mathbf{P}$. Furthermore, by assuming $s = \mathbf{R}(x)$, we can map every vertex to its *optimum-path tree* $\mathbf{T}(s) \subset \mathcal{V}$ by $\mathbf{T}(s) = \{t : \mathbf{R}(t) = s\}$. In this work, every superpixel is an optimum-path tree rooted in a seed.

3 Methodology

In this section, we review *Superpixels through Iterative CLEarcutting* (SICLE) [8] alongside our proposed evolutions to better take advantage of high quality saliency estimations. Briefly, each section refers to a specific SICLE step, given that our contributions reside on the last ones: (i) seed oversampling; (ii) superpixel generation; and (iii) object-based seed removal (*i.e.*, Sections 3.1, 3.2 and 3.3, respectively).

3.1 Seed Oversampling

Being a seed-based method, the first SICLE step consists in selecting a set $\mathcal{S} \subset \mathcal{V}$ of N_0 initial seeds for generating N_f superpixels, given $N_0, N_f \in \mathbb{N}^*$. However, differently from most approaches, it *oversamples* (*i.e.*, $N_0 \gg N_f$) for improving the probability of selecting the seeds that promote accurate object delineation (*i.e.*, *relevant*). In this strategy, the aim is to remove the irrelevant ones until reaching N_f seeds in the final iteration (see Section 3.3). In this work, we argue no need for seed selection improvement given the reported loss of efficiency when object saliency maps are considered [7] and the seed relevance redundancy premise [8]. Consequently, the central strategy for delineation improvement relies on maintaining the relevant seeds throughout the iterations.

3.2 Superpixel Generation

For generating superpixels, SICLE uses the seed-restricted IFT version. Furthermore, although several path-cost and arc-cost functions have been proposed [9,24,6], it opts for the $\mathbf{f}_{max}$ and a root-based arc-cost estimation $\mathbf{w}_{root}(x, y) = \|\mathbf{F}(\mathbf{R}(x)) - \mathbf{F}(y))\|_2$ due to its reported effectiveness in superpixel delineation [8].

As one may note, such arc-cost function does not consider any object information. The authors in [8] justify such option in SICLE mainly on the existence of incorrect estimations in the map, deteriorating the object delineation. And although it resulted in a more robust performance for any map, it also led to the inability to improve its delineation for state-of-the-art estimators. First, let $\alpha \in \{0, 1\}$ be a user-defined "trustiness" switch of the saliency map's object boundary approximation. When $\alpha = 1$, the user judges that the map's borders are sufficiently accurate for assisting in delineation, due to its closeness to the object's boundaries. Otherwise, it may set $\alpha = 0$ for avoiding incorrect estimations within the map, preventing segmentation degradation, while still exploiting the object location information for improving the SICLE performance. For both cases, it is expected that the object is known beforehand and it was properly located by the saliency estimator. Then, to achieve each property when desired, we propose a generalization $\mathbf{w}_*^\alpha(x, y) = (\mathbf{w}_*(x, y))^{1+\alpha\cdot\|\mathbf{O}(\mathbf{R}(x))-\mathbf{O}(y)\|_1}$. Note that, aside from not requiring optimization, $\mathbf{w}_*^\alpha = \mathbf{w}_*$ when $\alpha = 0$ since it discards the influence of the saliency difference. Finally, in contrast to the arc-cost function proposed in [4], the magnitude of the saliency influence in $\mathbf{w}_*^\alpha$ is significantly smaller, leading to a lighter impact by eventual incorrect estimations.

3.3 Seed Removal

In SICLE, N_f superpixels are obtained after successively removing $N_0 - N_f$ seeds from $\mathcal{S}$, requiring at most $\Omega \in \mathbb{N}^* > 1$ iterations. At each iteration $i \in \mathbb{N} < \Omega$, $\mathbf{M}(i) = \max\{(N_0)^{1-\omega\cdot i}, N_f\}$, given $\omega = 1/(\Omega - 1)$, most irrelevant seeds are removed, while the remaining ones are perpetuated for testing their relevance in the next iteration $i + 1$.

For each seed $s \in \mathcal{S}$, its relevance $\mathbf{V}_*(s) \in \mathbb{R}_+$ is estimated based on the characteristics of its superpixel $\mathbf{T}(s)$ resultant from the last IFT execution. As an example, one may opt for a size- and contrast-based criterion $\mathbf{V}_{sc}(s)$ for accurate selection of relevant seeds irrespective of whether a map is provided [8]. First, we define the color gradient between two superpixels rooted in $s, t \in \mathcal{S}$ by $\mathbf{G}_\mathbf{F}(s, t) = \left\|\overline{\mathbf{F}}(\mathbf{T}(s)) - \overline{\mathbf{F}}(\mathbf{T}(t))\right\|_2$. Moreover, it is possible to define the *adjacents* of s by $\mathbf{A}(s) = \{t : \exists\ \langle x, y\rangle \in \mathcal{A}\}$ considering $t \in \mathcal{S}$, $x \in \mathbf{T}(s)$, $y \in \mathbf{T}(t)$ and $s \neq t$. Finally, from both, we can compute the relevance of s by $\mathbf{V}_{sc}(s) = \frac{|\mathbf{T}(s)|}{|\mathcal{V}|} \cdot \min_{\forall\ t\in\mathbf{A}(s)} \{\mathbf{G}_\mathbf{F}(s, t)\}$ using the aforementioned criterion.

However, when a saliency map is given, the seed relevance is linked to its object proximity: the farther the superpixel, the more irrelevant it is, even though it is considered relevant by its non-object-based criterion [8]. Thus, in SICLE, every criterion is subjected to an object-based weighting factor. Similarly to $\mathbf{G}_\mathbf{F}(s, t)$, we define $\mathbf{G}_\mathbf{O}(s, t) = \left\|\overline{\mathbf{O}}(\mathbf{T}(s)) - \overline{\mathbf{O}}(\mathbf{T}(t))\right\|_2$ as the saliency gradient between the superpixels of s and t. Then, the object-based relevance of s is measured by $\mathbf{V}_{obj}(s) = \mathbf{V}_*(s) \cdot \max\{\overline{\mathbf{O}}(\mathbf{T}(s)), \max_{\forall\ t\in\mathbf{A}(s)} \{\mathbf{G}_\mathbf{O}(s, t)\}\}$.

One can see that such function favors seeds near the object border depicted on the map, promoting competition in crucial regions for delineation. Moreover,

by favoring those within it, SICLE populates the regions incorrectly estimated as object parts, diminishing the influence of such error through competition. However, when a high-quality map is provided, not only are its borders more accurate but promoting competition in internal object borders minimally impacts its exterior boundaries. In such case, one may favor the seeds nearby the object rather than those within it. Thus, and similarly to $\mathbf{w}_*^\alpha$, we generalize SICLE's object-based seed relevance function for assessing such properties whenever one of them is requested: $\mathbf{V}_{obj}^\alpha = \mathbf{V}_*(s) \cdot \max\{(1-\alpha) \cdot \overline{\mathbf{O}}(\mathbf{T}(s)), \max_{\forall\, t \in \mathbf{A}(s)} \{\mathbf{G}_\mathbf{O}(s,t)\}\}$. Similarly to $\mathbf{w}_{obj}^\alpha$, when $\alpha = 0$, the tree's saliency is also considered as a relevant feature, resulting in $\mathbf{V}_{obj}^\alpha = \mathbf{V}_{obj}$. Otherwise, only the trees near the saliency borders are favored in the computation.

4 Experimental Results

In this section, we present the experimental framework for analyzing and evaluating the proposed method. We first describe the experimental setup in Section 4.1 and subsequently perform an ablation study in Section 4.2. Lastly, we present a quantitative and qualitative analysis in Section 4.3.

4.1 Experimental Setup

We selected three datasets for assessing the performance of all methods. Given that the most used segmentation evaluation dataset [2] is contour-driven, it is not applicable when a single object is desired. Conversely, our selection tries to assess different delineation difficulties for distinct objects, while offering a broad perspective on the methods' performance in their primary goal: generating superpixels. For handling different objects, we opt for the popular *Extended Complex Saliency Scene Dataset* (ECSSD) [22], which contains 1000 natural images with complex objects and backgrounds. On the other hand, the thin object legs in *Insects* [18] (130 images) offers a proper delineation challenge. Similarly, the *Liver* [24] dataset contains 40 CT slices of the human liver whose smooth boundaries are difficult to detect. We selected, by random, 30% and 70% of each dataset for optimization and testing, respectively. Finally, we considered the U^2-Net [20], fine-tuned with its default parameters, for generating the object saliency maps.

As baselines, we chose the following state-of-the-art methods based on their speed and accuracy: (i) SLIC [1] [4]; (ii) SH [26] [5]; (iii) ERS [16] [6]; (iv) OISF [5] [7]. By selecting such baselines, we assess the major properties for superpixel segmentation: speed and object delineation. Thus, although deep-learning-based methods with promising results have been proposed, more research is required for surpassing the performance of classical algorithms [27,30,34,15,8]. As initial

[4] https://www.epfl.ch/labs/ivrl/research/slic-superpixels/
[5] https://github.com/semiquark1/boruvka-superpixel
[6] https://github.com/mingyuliutw/EntropyRateSuperpixel
[7] https://github.com/LIDS-UNICAMP/OISF

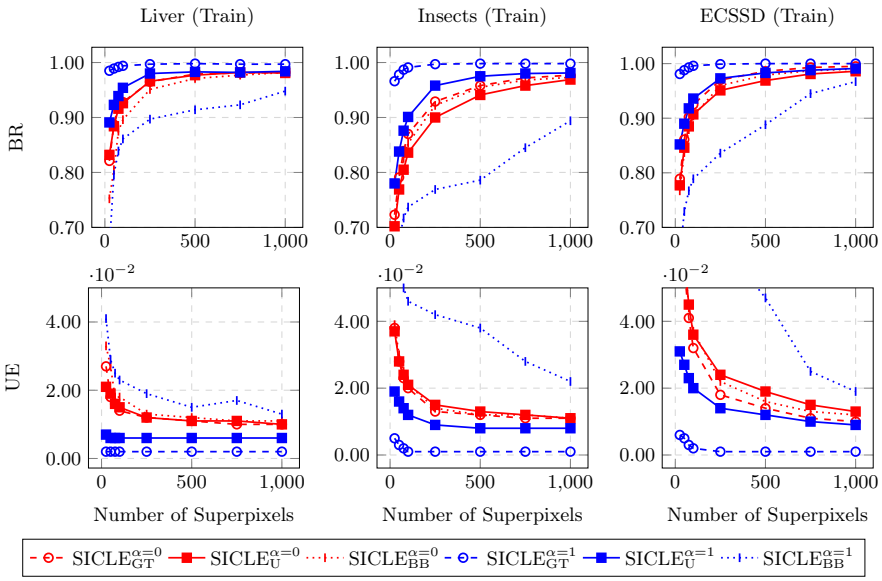

Fig. 2. Impacts of object saliency map quality on SICLE$^{\alpha}$.

setting for SICLE$^{\alpha}$, we used the default recommendation [8]:(i) random oversampling with $N_0 = 3000$; (ii) $\mathbf{f}_{max}$; (iii) $\mathbf{V}_* = \mathbf{V}_{sc}$; and (iv) $\Omega = 5$. Our code is publicly available online [8]. For measuring their performances, we used the *Boundary Recall* (BR) [23] and the *Under-segmentation Error* (UE) [19] due to their expressiveness [23]. While the former measures the ratio between object boundaries and superpixel borders (*i.e.*, higher is better), the latter estimates errors from superpixel "leakings" (*i.e.*, lower is better).

4.2 Ablation Study

We first analyzed the impacts of the saliency map quality on SICLE$^{\alpha}$, as shown in Figure 2. In this experiment, we considered a representative of a poor, a state-of-the-art, and an ideal estimator: (i) object's minimum bounding box (BB); (ii) U^2-Net (U); and (iii) ground-truth (GT); respectively. We highlight that the GT is only considered in this experiment. By setting $\alpha = 1$, our improved SICLE improves its segmentation proportionally to the saliency map quality (*i.e.*, the better the map, the better the delineation). Note that SICLE$^{\alpha}$ performance deteriorates when changing from GT to U maps, indicating that although highly accurate, the latter is not ideal. Still, by improving the saliency incorporation, our approach significantly improves when $\alpha = 1$, especially considering UE. Finally, by simply setting $\alpha = 0$ when a poor quality map is provided, SICLE$^{\alpha}$ becomes robust against saliency errors.

[8] https://github.com/LIDS-UNICAMP/SICLE

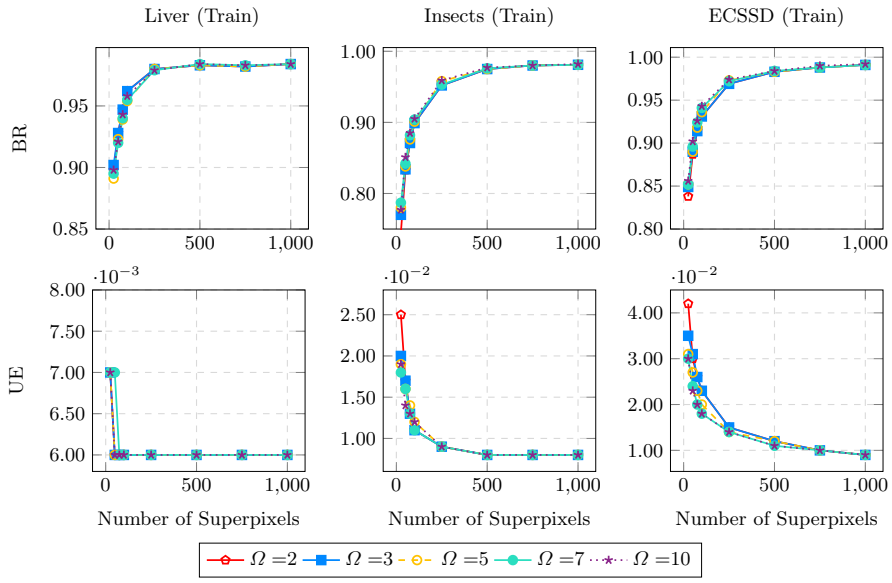

Fig. 3. Impacts of the maximum number of iterations on SICLE$^{\alpha}$ considering $\alpha = 1$ and the U^2-Net estimator.

Our second experiment analyses if the proposed method assists in reducing the number of iterations for segmentation. From the curves in Figure 3, we see that SICLE$^{\alpha}$ manages to achieve its top performance requiring only two iterations and increasing Ω does not lead to improvements when $\alpha = 1$. We argue that our improved seed relevance criteria accurately select N_f relevant seeds in only one iteration, requiring one more for promoting effective object delineation. For that, we set $\Omega = 2$ whenever $\alpha = 1$.

4.3 Quantitative and Qualitative Analysis

Our last experiment (Figure 4) compared our improved SICLE$^{\alpha}$ against the baselines. In terms of BR, SICLE$^{\alpha}$ managed to surpass all methods significantly, especially for $N_f = 200$ superpixels. Given its discrepant performance compared to OISF, we can argue that our approach best exploits the saliency information for segmentation. Moreover, SICLE$^{\alpha}$ presented better delineations in Insects than ERS, the best method known in such dataset. Regarding UE, our improvements reduced the superpixel leaking significantly, leading to on par results with OISF, which often presents the lowest values in several works [4,5,3].

Table 1 shows the average speed performance of all methods in the ECSSD dataset on a 64-bit Intel(R) Core(TM) i7-4790S PC with CPU speed of 3.20GHz. As one can see, even though SICLE$^{\alpha}$ is $\mathrm{O}\left(|\mathcal{V}| \log |\mathcal{V}|\right)$, it is the fastest method amongst all. For instance, SLIC and SH are $\mathrm{O}\left(|\mathcal{V}|\right)$, but they perform burdensome operations for obtaining a single segmentation. While the former executes

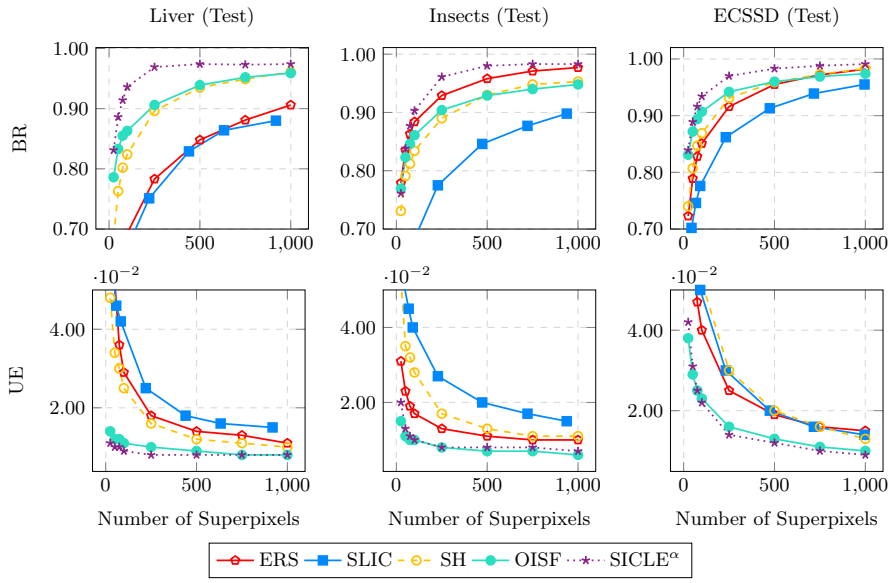

Fig. 4. Performance comparison between our approach, considering $\alpha = 1$ against state-of-the-art methods.

N_0	SLIC	ERS	SH	OISF	SICLE$^{\alpha}$
25	0.537±0.028	0.913±0.085	0.758±0.028	1.987±0.367	**0.279±0.037**
100	0.540±0.029	0.952±0.093	0.756±0.026	1.252±0.236	**0.279±0.030**
750	0.541±0.029	1.027±0.109	0.756±0.027	0.849±0.163	**0.296±0.037**

Table 1. Average speed performance (in seconds) on the ECSSD dataset considering $\alpha = 1$ for SICLE$^{\alpha}$. The best value for each N_f is depicted in bold.

a strict number of iterations (*e.g.*, 10), the latter computes the whole hierarchy. In contrast, SICLE$^{\alpha}$ surpasses both speed and delineation by benefitting from our improvements, leading to only two iterations. Lastly, it is straightforward to obtain an object-based multiscale segmentation on the fly from SICLE$^{\alpha}$ [8].

Finally, the superior performance of SICLE$^{\alpha}$ can be exemplified by Figure 5. Note that, by setting $\alpha = 1$, our improved method can correct the errors for $\alpha = 0$, achieving top object delineation and surpassing all baselines. As indicated by the yellow rectangles, SICLE$^{\alpha}$ best exploits the saliency information and best approximates the object borders, especially when compared to other object-based methods like OISF.

5 Conclusion and Future Work

This work proposes SICLE$^{\alpha}$, an improved version of the state-of-the-art object-based method *Superpixels through Iterative CLEarcutting* (SICLE) by generalizing its path-cost function and seed removal criterion. Our proposal may promote

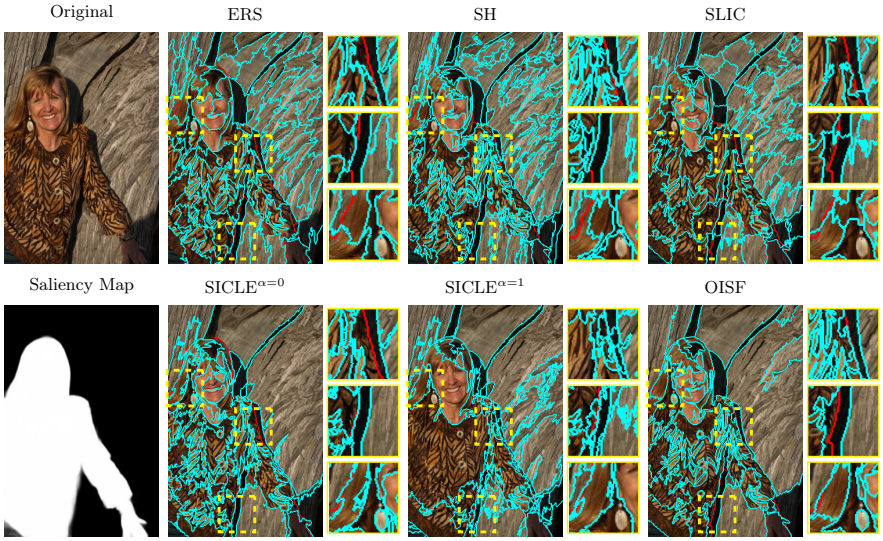

Fig. 5. Comparison between SICLE$^{\alpha}$ and state-of-the-art methods considering 100 superpixels. Red lines indicate object boundaries, whereas cyan ones, superpixel borders. Yellow rectangles indicate delineation errors that our approach overcame.

robustness for low-quality saliency maps or may improve its effectiveness and efficiency in delineation for high-quality ones through a single and intuitive parameter. Results show that SICLE$^{\alpha}$ surpasses state-of-the-art methods regarding popular metrics while being the fastest one in all datasets considered. For future work, we intend to extend SICLE$^{\alpha}$ for video supervoxel segmentation and study its performance for interactive image segmentation.

References

1. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: SLIC superpixels compared to state-of-the-art superpixel methods. Transactions on Pattern Analysis and Machine Intelligence **34**(11), 2274–2282 (2012)
2. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. Transactions on Pattern Analysis and Machine Intelligence **33**(5), 898–916 (2011)
3. Belém, F., Cousty, J., Perret, B., Guimarães, S., Falcão, A.: Towards a simple and efficient object-based superpixel delineation framework. In: 34th Conference on Graphics, Patterns and Images (SIBGRAPI). pp. 346–353 (2021)
4. Belém, F., Guimarães, S., Falcão, A.: Superpixel segmentation by object-based iterative spanning forest. In: 23rd Iberoamerican Congress on Pattern Recognition. pp. 334–341 (2018)
5. Belém, F., Guimarães, S., Falcão, A.: Superpixel generation by the iterative spanning forest using object information. In: 33rd Conference on Graphics, Patterns and Images (SIBGRAPI). pp. 22–28 (2020), workshop of Thesis and Dissertations

6. Belém, F., Guimarães, S., Falcão, A.: Superpixel segmentation using dynamic and iterative spanning forest. Signal Processing Letters **27**, 1440–1444 (2020)
7. Belém, F., Melo, L., Guimarães, S., Falcão, A.: The importance of object-based seed sampling for superpixel segmentation. In: 32nd Conference on Graphics, Patterns and Images (SIBGRAPI). pp. 108–115 (2019)
8. Belém, F., Perret, B., Cousty, J., Guimarães, S., Falcão, A.: Efficient multiscale object-based superpixel framework. arXiv preprint pp. 1–19 (2022)
9. Borlido, I., Belém, F., Miranda, P., Falcão, A., Patrocínio, Z., Guimarães, S.: Towards interactive image segmentation by dynamic and iterative spanning forest. In: Discrete Geometry and Mathematical Morphology. pp. 351–364 (2021)
10. Conze, P.H., Tilquin, F., Lamard, M., Heitz, F., Quellec, G.: Unsupervised learning-based long-term superpixel tracking. Image and Vision Computing **89**, 289–301 (2019)
11. Falcão, A., Stolfi, J., Lotufo, R.: The image foresting transform: Theory, algorithms, and applications. Transactions on Pattern Analysis and Machine Intelligence **26**(1), 19–29 (2004)
12. Fehri, A., Velasco-Forero, S., Meyer, F.: Prior-based hierarchical segmentation highlighting structures of interest. Mathematical Morphology-Theory and Applications **3**(1), 29–44 (2019)
13. Galvão, F., Falcão, A., Chowdhury, A.: RISF: recursive iterative spanning forest for superpixel segmentation. In: 31st Conference on Graphics, Patterns and Images (SIBGRAPI). pp. 408–415 (2018)
14. Jampani, V., Sun, D., Liu, M., Yang, M., Kautz, J.: Superpixel sampling networks. In: 18th European Conference on Computer Vision (ECCV). pp. 352–368 (2018)
15. Kang, X., Zhu, L., Ming, A.: Dynamic random walk for superpixel segmentation. IEEE Transactions on Image Processing **29**, 3871–3884 (2020)
16. Liu, M., Tuzel, O., Ramalingam, S., Chellappa, R.: Entropy rate superpixel segmentation. In: 24th Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2097–2104 (2011)
17. Machairas, V., Faessel, M., Cárdenas-Peña, D., Chabardes, T., Walter, T., Decenciere, E.: Waterpixels. IEEE Transactions on Image Processing **24**(11), 3707–3716 (2015)
18. Mansilla, L., Miranda, P.: Oriented image foresting transform segmentation: Connectivity constraints with adjustable width. In: 29th Conference on Graphics, Patterns and Images (SIBGRAPI). pp. 289–296 (2016)
19. Neubert, P., Protzel, P.: Superpixel benchmark and comparison. In: Forum Bildverarbeitung. vol. 6, pp. 1–12 (2012)
20. Qin, X., Zhang, Z., Huang, C., Dehghan, M., Zaiane, O., Jagersand, M.: U2-net: Going deeper with nested u-structure for salient object detection. Pattern Recognition **106**, 107404 (2020)
21. Sellars, P., Aviles-Rivero, A.I., Schönlieb, C.B.: Superpixel contracted graph-based learning for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing **58**(6), 4180–4193 (2020)
22. Shi, J., Yan, Q., Xu, L., Jia, J.: Hierarchical image saliency detection on extended cssd. Transactions on Pattern Analysis and Machine Intelligence **38**(4), 717–729 (2015)
23. Stutz, D., Hermans, A., Leibe, B.: Superpixels: An evaluation of the state-of-the-art. Computer Vision and Image Understanding **166**, 1–27 (2018)
24. Vargas-Muñoz, J., Chowdhury, A., Alexandre, E., Galvão, F., Miranda, P., Falcão, A.: An iterative spanning forest framework for superpixel segmentation. Transactions on Image Processing **28**(7), 3477–3489 (2019)

25. Wang, M., Liu, X., Gao, Y., Ma, X., Soomro, N.Q.: Superpixel segmentation: A benchmark. Signal Processing: Image Communication **56**, 28–39 (2017)
26. Wei, X., Yang, Q., Gong, Y., Ahuja, N., Yang, M.: Superpixel hierarchy. Transactions on Image Processing **27**(10), 4838–4849 (2018)
27. Wu, J., Liu, C., Li, B.: Texture-aware and structure-preserving superpixel segmentation. Computers and Graphics **94**, 152–163 (2021)
28. Wu, J., Liu, C., Li, B.: Texture-aware and structure-preserving superpixel segmentation. Computers & Graphics **94**, 152–163 (2021)
29. Yang, F., Sun, Q., Jin, H., Zhou, Z.: Superpixel segmentation with fully convolutional networks. In: 33rd Conference on Computer Vision and Pattern Recognition (CVPR) (2020)
30. Yu, Y., Yang, Y., Liu, K.: Edge-aware superpixel segmentation with unsupervised convolutional neural networks. In: 28th International Conference on Image Processing (ICIP). pp. 1504–1508 (2021)
31. Yuan, Y., Zhang, W., Yu, H., Zhu, Z.: Superpixels with content-adaptive criteria. IEEE Transactions on Image Processing **30**, 7702–7716 (2021)
32. Zhang, J., Aviles-Rivero, A.I., Heydecker, D., Zhuang, X., Chan, R., Schönlieb, C.B.: Dynamic spectral residual superpixels. Pattern Recognition **112**, 107705 (2021)
33. Zhao, W., Fu, Y., Wei, X., Wang, H.: An improved image semantic segmentation method based on superpixels and conditional random fields. Applied Sciences **8**(5), 837 (2018)
34. Zhu, L., She, Q., Zhang, B., Lu, Y., Lu, Z., Li, D., Hu, J.: Learning the superpixel in a non-iterative and lifelong manner. In: 34th Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1225–1234 (2021)