SEMASUPERPIXEL: A MULTI-CHANNEL PROBABILITY-DRIVEN SUPERPIXEL SEGMENTATION METHOD

Xuehui Wang[1†] *Qingyun Zhao*[1†] *Lei Fan*[2] *Yuzhi Zhao*[3] *Tiantian Wang*[4] *Qiong Yan*[4] *Long Chen*[1⋆]

[1] Sun Yat-sen University, China [2]Northwestern University, USA
[3]City University of Hong Kong, Hong Kong, China [4]SenseTime Research, China

ABSTRACT

Superpixel, an efficient image segmentation approach, aggregates a group of similar pixels into the same cluster. Existing superpixel algorithms still mainly focus on the color information while ignoring the semantic distribution knowledge. In this paper, we propose a semantic information-driven method that introduces multi-channel semantic probabilities into a superpixel segmentation method. By conducting statistical analysis on the semantic output and then formulating the distance measure, the prior knowledge of the semantic with a dynamic confidence value could be utilized by our method during the global update effectively. Extensive experimental evaluations show that our method achieves a leading segmentation quality and convergence speed, compared to other five state-of-the-art algorithms, as measured by boundary recall, under-segmentation error and explained variation.

Index Terms— Superpixel, Image segmentation, Image processing

1. INTRODUCTION

As a significant component of image segmentation, superpixel segmentation merges neighboring pixels sharing similar texture, color or other characteristics into a superpixel [1]. The boundary of the superpixel usually adheres to the contour of the object. Due to its characters and effectiveness, superpixel segmentation is widely adopted into many visual tasks.

To promote the accuracy of superpixel segmentation, many methods have been proposed. The graph-based image segmentation method [2] is a classical greedy clustering algorithm, which is simple to implement and fast to process. The CIS [3] formulates the superpixel partitioning problem in an energy minimization framework and produces more regular superpixel. The SLIC [4] constructs a distance metric and performs the segmentation process locally and iteratively. By employing the hill climbing strategy on hierarchical blocks, the SEEDS algorithm [5] adjusts the color histogram and boundary regularity through an energy function. The depth information is utilized in the work [6], which proposes an over-segmentation algorithm to produce results with three-dimensional spatial information and achieves better performance. Both of them only integrate color information and spatial information among different pixels. This results in the issue that superpixel usually adheres to boundaries where colors of pixels have massive changes and lacks of semantic characteristic of pixels. In some cases, the color knowledge is not reliable or discriminative enough.

Many researchers also investigate to combine the strong capacity of neural networks into superpixel segmentation. [7] proposed a method for graph-based superpixel segmentation by exploring affinities between pixels and devising a segmentation-aware loss to solve the problem of back-propagating. [8] employs a fully convolutional network to produce superpixel on a ragular image grid that is predicted by an initialization strategy commonly used in traditional superpixel algorithms. But there also exists an inevasible problem that it is difficult to define what is a good superpixel so that paired datas to supervise the network are absent.

To solve the aforementioned drawbacks, we propose a energy optimization-based superpixel segmentation method that considers both the color information, spatial information, and the semantic prior knowledge from deep learning networks. The main contributions of this paper are:

- Instead of using semantic label, the proposed method uses the multi-channel semantic probability to assist the segmentation process effectively.
- To better derive the prior knowledge, we conduct the statistical analysis on the multi-channel probability to choose a suitable dynamic confidence value.
- The experiment results demonstrate our superiority in quantitative, qualitative and convergence evaluations.
- We provide a meaningful direction to handle some semantic-based tasks since good superpixel can facilitate the semantic segmentation and accurate semantic information can promote the quality of the superpixel. They constitute a positive cycle.

2. PROPOSED METHOD

According to the initial definition of superpixel in [1], the following properties should be guaranteed for an ideal superpixel segmentation: (a) Pixels in the same superpixel should

† contributes equally to this paper. ⋆ is the corresponding author.

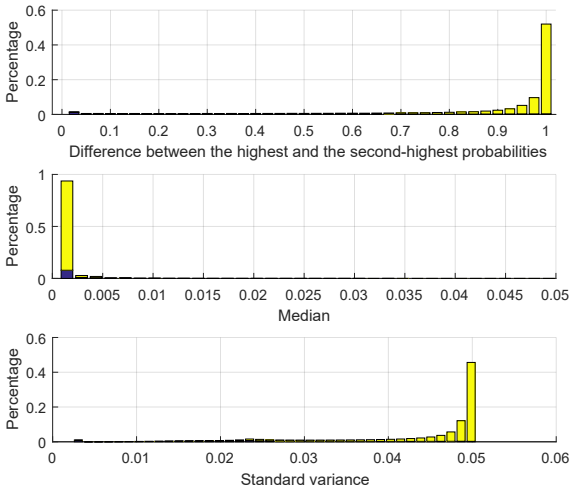


Fig. 1. The statistical analysis to semantic multi-channel probabilities with three different ways.

belong to the same object in the real world. The foundation to support this opinion is that different colors are usually the key factors to distinguish different objects, so there exists high possibilities that the same color means the same object. (b) The shape of the superpixel should be as compact and regular as possible. (c) Superpixels should maintain the original color properties and the result of superpixel segmentation should preserve the color distribution as soon as possible.

In this part, we describe two approaches of adopting semantic information into the superpixel segmentation. We obtain the semantic information from DeepLab v3+ [9] without any re-training. Note that arbitrary semantic segmentation method can be used here.

2.1. Simple Linear Iterative Clustering

The energy optimization-based superpixel segmentation method first partition the input image into regular grids and then iteratively exchanges pixels located at the boundary between adjacent superpixels with regard to the energy. With a color image $\mathcal{I}$, the superpixel segmentation $S = \{S_1, \ldots, S_k\}$ with k denotes the number of superpixel has the following properties:

$$S_i \cap S_j = \emptyset \quad and \quad \mathcal{I} = \bigcup_{i}^{k} S_i. \tag{1}$$

We utilize the SLIC [4] as our basic approach. The energy function during boundary update between pixels p_i and p_j is defined as $E_{slic}(p_i, p_j) = d_c(p_i, p_j) + \alpha d_s(p_i, p_j)$ with $d_c(.,.)$ and $d_s(.,.)$ stand for the Euclidean measure in the CIELab color space and the image spatial coordinate, and α is a constant balancing factor. The lower of E_{slic} between pixels is considered higher proximity.

2.2. Semantic Label-assisted SLIC

The straightforward idea to explicitly involve semantic information is to combine the semantic label difference with the original function. Let $l_i \in |\mathcal{L}|$ denotes the semantic label for pixel p_i with $|\mathcal{L}|$ the size of predicted categories. The equation for optimization with semantic labels is formulated as

$$D^1_{sem}(p_i, p_j) = D_{slic}(p_i, p_j) + 0 \cdot (l_i = l_j) + \gamma \cdot (l_i \neq l_j), \tag{2}$$

where γ is a constant. Thus, measured pixels are considered higher proximity if they share the same semantic category.

2.3. SemaSuperpixel

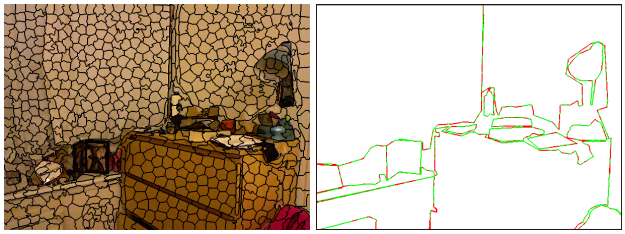
Due to the boundary of objects remains as the major error during semantic segmentation, employing the semantic output directly is not promising for superpixel segmentation. So we utilize the input vector to softmax layers as multi-channel probabilities. The index of each channel denotes a semantic class. Compared to one category label, the vector contains much more abundant information, which is defined as the semantic descriptor in this paper. The semantic descriptor for pixel p_i is defined as

$$\mathcal{X}_i = \{x_1, x_2, \ldots, x_{|\mathcal{L}|} | \sum_{k}^{|\mathcal{L}|} x_k = 1\}, \tag{3}$$

where x is the probability of each semantic category after the softmax layer.

We choose to calculate the distance between their semantic descriptors. As the semantic result still contains apparent mistakes, we multiple the distance with a dynamic confidence value. For the mechanism of neural networks, it is impossible to explain the value of multi-channel outputs based on the input color image. We choose to analyze the relation between semantic descriptors and the ground truth. As demonstrated in Fig. 1, we measured the difference between the highest and the second-highest probabilities, the standard deviation and the median in semantic descriptors. It is shown in Fig. 2 that the difference between the highest and the second-highest probabilities has higher discriminative ability and is easier to handle during implementation. The final energy function with semantic information is formulated as

$$D^2_{sem}(p_i, p_j) = D_{slic}(p_i, p_j) + \rho \cdot \omega(\mathcal{X}_i, \mathcal{X}_j) \cdot \sum_{k}^{|\mathcal{L}|} (x_{i,k} - x_{j,k})^2, \tag{4}$$



(a) (b)

Fig. 2. A qualitative result on the NYU2 dataset. (a) is the superpixel segmentation result and (b) is the accuracy compared to the ground truth.

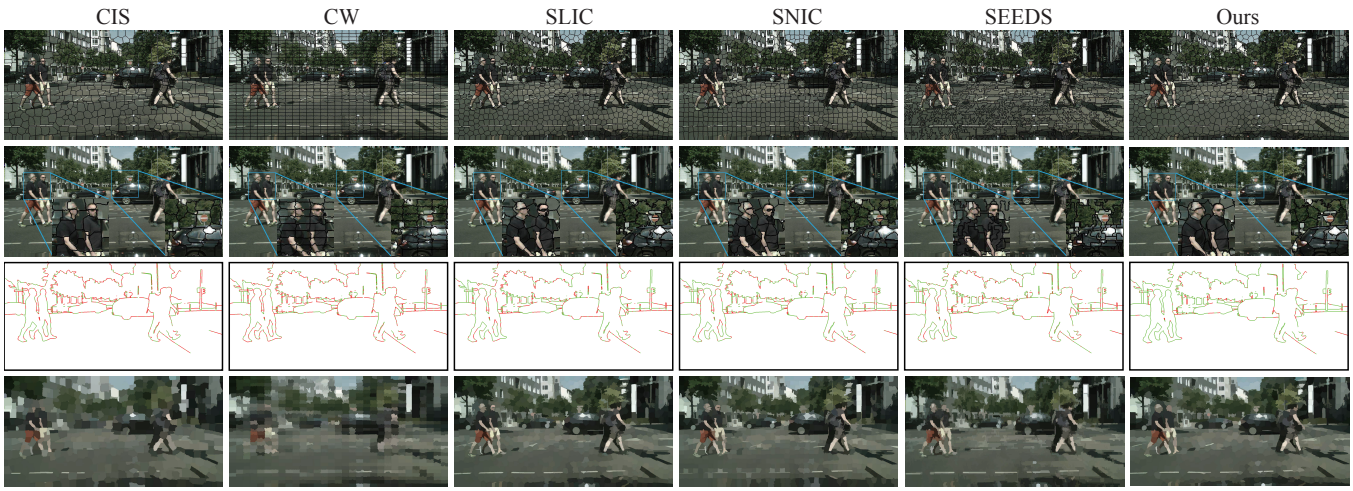


Fig. 3. From the top to bottom are the superpixel segmentation result, the boundary details, the accuracy compared to the ground truth and the reconstructed image created by superpixels.

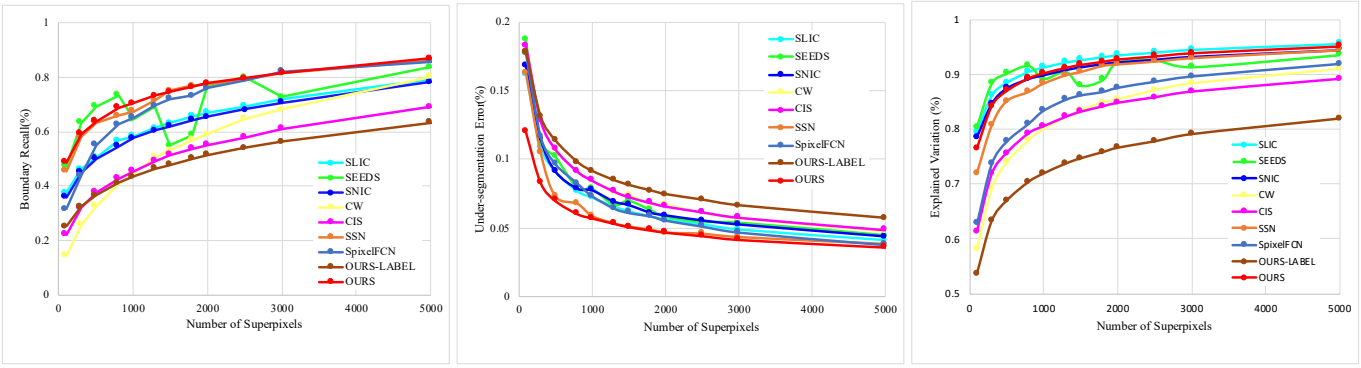


Fig. 4. From the left to right: Boundary Recall, Undersegmentation Error, Explained Variation. We evaluate six methods including our approach on three metrics. We treat $k = 5000$ as the upper limit and $k = 100$ as the lower limit.

where $\omega(\mathcal{X}_i, \mathcal{X}_j)$ is defined as the difference between the highest and the second-highest probabilities.

3. EXPERIMENT

We compared our method with other five traditional superpixel methods including the graph-based CIS [3], the watershed-based CW [10], the energy-based SEEDS [5], the SLIC [4] and the SNIC [11] on the Cityscapes dataset [12] and the NYU2 dataset [13]. We also demonstrate that using semantic label described in Sec 2.2 directly is unsuitable. We name it Ours-label.

WSS [14] is a new method proposed in 2019, but we can not have a comparision with it for they do not release their official codes. We also compare our method with some deep learning-based methods [8, 15] though these method have great demands for pair datas to train a network. How to define a CORRECT superpixel is unclear, which also indicates that the groundtruth used to supervise the network is objectively uncharted. As a fundamental technique for high-level tasks, superpixel should not be invested much resources for training and colloecting datas.

3.1. Qualitative Evaluation

We present the result when $k = 1000$ on the Cityscape dataset and qualitative details on the NYU2 dataset at different number of k . We obtain better performance in an intricate situation, such like the indoor images displayed in Fig. 2. Each superpixel produced by our method is a part of a plane rather than two or more planes, which agrees with the definition of the superpixel.

In Fig. 3, the green line indicates the correct segmentation while the red line marks the missing boundary pixels. We have better performance in boundary adherence than others. This improvement appears because of the richer semantic information which designates the movement range of pixels in the semantic space during the boundary update.

We also have distinct preponderance in fine-grained segmentation and the reconstruction of images. From Fig. 3, our method divides pixels belonging to different categories into

diverse superpixel more carefully. This also induces pixels with the same color to aggregate together, which makes the color restoration closer to the original image. In summary, our method regards the semantic prior information as a coarse factor while the color and spatial distance as a refined factor. Combining these two factors facilitates this results.

3.2. Quantitative Evaluation

We choose the Boundary Recall(BRecall), the Undersegmentation Error(UE), the Explained Variation(EV) proposed in [16] to evaluate our algorithm. We also use the AMR, the AUE, and the AUV described in [16] to reflect the performance without relying on the number of superpixels. Please refer to [16] for details of these metrics.

Boundary Recall. BRecall measures the overlapping condition between the superpixel boundary and the groundtruth. From Fig. 4, our method maintains the optimal result with almost all of different numbers of superpixels. Compared with the second highest SEEDS, our method surpasses it in most cases and is more stable. Especially, we give superior results in textureless and low-illumination area.

Undersegmentation Error. UE measures the proportion that the pixel set is assigned to different label against the major pixels in a superpixel. In Fig 4, our algorithm achieves an absolute leading effect on this metric. Especially, when the k is small, the result has been improved nearly 31% against the second best method. With respect to other methods, our algorithm decreases the number of "leakage" as much as possible thanks to the semantic energy. This ability reduces the heavy dependence of color energy.

Explained Variation. EV judges the quality of the superpixel to preserve the original image knowledge but is not relevant to the groudtruth. In Fig 4, our method get a comparable result to the best method SLIC. Extreme high EV also represents a fact that a superpixel fuses multiple colors and loses specific information of the color distribution. We utilize semantic prior knowledge to prevent from causing pixels belonging to the same object be mis-partitioned into neighboring superpixel, which reduce EV a little to a modest value.

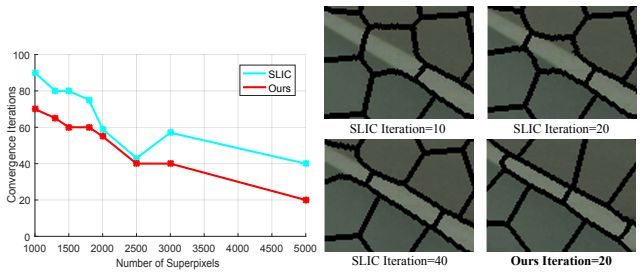
More Metrics. The AMR, AUE and AUV are adopted to evaluate the overall performance independent of k by calculating the average over a large given interval $k_{inter} = [100, 5000]$. As shown in Table 1, our method presents strong results on these metrics. The AMR decreases by 15.2% compared to the second best algorithm SEEDS. The AUE significantly outperforms the SLIC up to 21.6%, with only a little deficiency on AUV. Thus, we derive the most outstanding outcome.

3.3. Convergence Speed

We compare the convergence speed with the SLIC that we build on. This ensures parameters are consistent, i.e., numbers of superpixels, the compactness and the color-space. We define the current state at the iteration i as $C_i = BRecall_i + (1-UE_i)+EV_i$. i is the number of times it has been iterated. For the optimal C, we treat $C_i \in [0.999C, C]$ as convergence state. As demonstrated in the Fig 5, our approach accelerates the convergence speed with an average rate of 22.44%. The acceleration comes from delimiting a rough range when clustering pixels.

Table 1. The metrics of AMR, AUE, AUV to evaluate the performance of the methods independent of k. * are deep learning-based methods.

Algorithm	AMR↓	AUE↓	AUV↓
CW [10]	0.508	0.086	0.199
CIS [3]	0.519	0.087	0.197
SLIC [4]	0.394	0.075	**0.091**
SNIC [11]	0.406	0.078	0.104
SEEDS [5]	0.322	0.081	0.103
SSN* [15]	0.293	0.066	0.119
SpixelFCN* [8]	0.337	0.076	0.171
Ours-label	0.543	0.094	0.279
Ours	**0.273**	**0.058**	0.099



(a) Convergence speed (b) Details in different iterations using the SLIC and Ours

Fig. 5. The convergence speed comparison with the SLIC.

4. CONCLUSION

In this paper, we focus on the problem that generates superpixels combined with the semantic information. We measure the pros and cons of different definitions of the semantic confidence in the form of statistical analysis. We primarily discuss the positive impact that the semantic information with the prior knowledge could bring to the superpixel segmentation, and we propose a modified method combined with the semantic information based on the work of [4] to show the superiority. By adding the semantic term, we not only reduce required iterations during the segmentation but also outperform other methods both quantitatively and qualitatively on two datasets. Experimental results confirm that the semantic information has a promising and facilitating role in the superpixel segmentation.

5. REFERENCES

[1] Xiaofeng Ren and Jitendra Malik, "Learning a classification model for segmentation," in *Proc. 9th Int'l. Conf. Computer Vision*, 2003, vol. 1, pp. 10–17.

[2] Pedro F Felzenszwalb and Daniel P Huttenlocher, "Efficient graph-based image segmentation," *IJCV*, vol. 59, no. 2, pp. 167–181, 2004.

[3] Olga Veksler, Yuri Boykov, and Paria Mehrani, "Superpixels and supervoxels in an energy optimization framework," in *ECCV*. Springer, 2010, pp. 211–224.

[4] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, Sabine Susstrunk, et al., "Slic superpixels compared to state-of-the-art superpixel methods," *T-PAMI*, vol. 34, no. 11, pp. 2274–2282, 2012.

[5] Michael Van den Bergh, Xavier Boix, Gemma Roig, Benjamin de Capitani, and Luc Van Gool, "Seeds: Superpixels extracted via energy-driven sampling," in *ECCV*. Springer, 2012, pp. 13–26.

[6] David Weikersdorfer, David Gossow, and Michael Beetz, "Depth-adaptive superpixels," in *ICPR*. IEEE, 2012, pp. 2087–2090.

[7] Wei-Chih Tu, Ming-Yu Liu, Varun Jampani, Deqing Sun, Shao-Yi Chien, Ming-Hsuan Yang, and Jan Kautz, "Learning superpixels with segmentation-aware affinity loss," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 568–576.

[8] Fengting Yang, Qian Sun, Hailin Jin, and Zihan Zhou, "Superpixel segmentation with fully convolutional networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 13964–13973.

[9] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *ECCV*, 2018, pp. 801–818.

[10] Peer Neubert and Peter Protzel, "Compact watershed and preemptive slic: On improving trade-offs of superpixel segmentation algorithms," in *ICPR*. IEEE, 2014, pp. 996–1001.

[11] Radhakrishna Achanta and Sabine Susstrunk, "Superpixels and polygons using simple non-iterative clustering," in *CVPR*, 2017, pp. 4651–4660.

[12] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele, "The cityscapes dataset for semantic urban scene understanding," in *CVPR*, 2016.

[13] Pushmeet Kohli Nathan Silberman, Derek Hoiem and Rob Fergus, "Indoor segmentation and support inference from rgbd images," in *ECCV*, 2012.

[14] Xin Qian, Xuemei Li, and Caiming Zhang, "Weighted superpixel segmentation," *Vis. Comput.*, vol. 35, no. 6–8, pp. 985–996, jun 2019.

[15] Varun Jampani, Deqing Sun, Ming-Yu Liu, Ming-Hsuan Yang, and Jan Kautz, "Superpixel sampling network," in *The European Conference on Computer Vision (ECCV)*, September 2018.

[16] David Stutz, Alexander Hermans, and Bastian Leibe, "Superpixels: An evaluation of the state-of-the-art," *CVIU*, vol. 166, pp. 1–27, 2018.