Superpixel Sizes using Topology Preserved Regular Superpixel Algorithm and their Impacts on Interactive Segmentation

Kok Luong Goh[1], Soo See Chai[2], Giap Weng Ng[3], Muzaffar Hamzah[4]
Faculty of Science and Technology, i-CATS University College, Kuching, Malaysia[1]
Faculty of Computer Science and Information Technology, University Malaysia Sarawak, Kuching, Malaysia[2]
Faculty of Computing and Informatics, University Malaysia Sabah, Kota Kinabalu, Malaysia[1, 3, 4]

***Abstract*—Interactive Image Segmentation is a type of semi-automated segmentation that uses user input to extract the object of interest. It is possible to speed up and improve the end result of segmentation by using pre-processing steps. The use of superpixels is an example of a pre-processing step. A superpixel is a collection of pixels with similar properties such as texture and colour. Previous research was conducted to assess the impact of the number of superpixels (based on SEEDS superpixel aglorithms) required to achieve the best segmentation results. The study, however, only examined one type of input (strokes) and a small number of images. As a result, the goal of this study is to extend previous work by performing interactive segmentation with input strokes and a combination of bounding box and strokes on images from Grabcut image data sets generated by Topology preserved regular superpixel (TPRS). Based on our findings, an image with 1000 to 2500 superpixels and a combination of bounding box and strokes will help the interactive segmentation algorithm produce a good segmentation result. Finally, the size of the superpixels would influence the final segmentation results as well as the input type.**

***Keywords*—Image segmentation; superpixel; input type; interactive segmentation**

I. Introduction

Computer vision is an artificial intelligence subfield that teaches computers to understand and interpret their visual environment. Image processing is a subset of computer vision that entails enhancing or extracting useful information from images. Image segmentation, a subprocess of image processing, on the other hand, is a technique that allows humans to extract objects of interest from images. Image segmentation can be done manually, semi-automatically, or fully automatically.

In automated segmentation, there is no user involvement. Semi-automated segmentation, also known as interactive segmentation, requires little to no user intervention during the segmentation process. The ultimate goal of image segmentation is to fully automate the process. However, because of the image complexity, automated segmentation continues to face significant challenges in producing satisfactory results. As a result, for better image segmentation results, semi-automated or interactive image segmentation methods are preferred.

Traditionally, image segmentation is performed using information from each individual pixel in the image. This process, however, consumes a significant amount of processing power. Ren and Malik [1] proposed superpixel as a solution to this issue. A superpixel is a group of pixels with similar properties, such as texture or colour. The introduction of the superpixel altered the segmentation processing steps.

Since the introduction of superpixels, many interaction segmentation methods have used them as part of the pre-processing phase [2-5]. However, different interactive segmentation algorithms employ different superpixel algorithms of varying sizes. For instance, [6] and [7] used the SLIC superpixel algorithm [8] with 1000 and 2000 superpixels superpixel per image, respectively. In addition, [9] used a meant-shift superpixel algorithm with 100 superpixels per image. As can be seen, no standard method exists in interactive segmentation for determining the size of a superpixel. As a result, a study [10] was conducted using MSRM interactive segmentation [11] and the SEEDS superpixel algorithm[12] to assess the effect of the number of superpixels required to achieve the best segmentation results. According to the study, 500 superpixels per image were the optimal size for achieving a good segmentation result. The study, however, only looked at one type of input (strokes) and a small number of images. Therefore, the purpose of this paper is to expand on the previous study on the following:

- Using all images from the Grabcut dataset [13] (50 images).
- Using different superpixel algorithm e.g., Topology preserved regular superpixel (TPRS) [14].
- Using various input types, such as bounding box and strokes.

The study's findings will be compared to state-of-the-art interactive algorithms on a variety of metrics.

The study's findings will provide information on the optimal size of the superpixel for TPRS based on the input types utilised by the MSRM interactive segmentation algorithms. In addition, it will serve as an extra guideline when TPRS superpixels are used with other segmentation algorithms.

In the following section, interactive segmentation will be discussed, followed by an overview of the various user input types used in interactive segmentation, maximal similarity-based region merging, and topology preserved regular superpixel algorithms.

II. Literature Review

Interactive image segmentation has been used in a variety of applications. For example, a tool for medical volume images has been created (SmartPaint [15] and MRI for orthopaedic surgery[16]. In the field of remote sensing, a segmentation tool for lithological boundary detection has also been developed [17]. It has also proven to be very useful in agriculture, assisting farmers in detecting crop diseases [18].

As stated in the preceding paragraph, the user will provide guidance in order for the segmentation system to extract the object of interest in interactive segmentation. The general process of interactive segmentation is summarized below:

Step 1: The user will provide information about the context and the object of interest.

Step 2: The segmentation system will generate a segmentation result based on the user's input.

Step 3: The user will evaluate the results, and if the user is satisfied, the process will be completed. Otherwise, the user will provide more background and object of interest information until the system produces a satisfactory segmented result.

Existing work [19] distinguishes between interactive and semi-automatic segmentation by involving the user in both the initialization and post-processing stages of the segmentation process iteratively, whereas semi-automatic segmentation only involves the user in the initialization stage. This study combines the two terms and defines interactive segmentation as any segmentation that requires user input. This study, on the other hand, will concentrate on the involvement of user input during the initialization stage.

Various input types are used in interactive segmentation to provide information about the background and object of interest. Some examples of these input formats are as follows:

- Strokes [20-24]: The user must apply stroke(s) to the image's object of interest and background.
- Seed point [25-29]: The seed points must be placed on the image's background and object of interest by the user.
- Bounding box [13, 30-34]: The user must position the bounding box around the object of interest within the image.

The following section explains Maximal Similarity-based Region Merging (MSRM) and Topology Preserved regular superpixels (TPRS).

A. Maximal Similarity-based Region Merging (MSRM)

Maximal Similarity-based Region Merging algorithm [11] is based on region merging. The image is first converted into superpixels using mean shift segmentation. The contour of the object is then extracted based on the labelling of non-marked regions as region of interest or background. Fig. 1(a) shows the superpixels of the image with strokes on the background and object of interest, and (b) shows the segmentation result.

(a) (b)

Fig. 1. The Algorithm's Segmentation Process [11]: (a) Superpixel Strokes Entered by users (b): The Segmentation Outcome.

B. Topology Preserved Regular Superpixel (TPRS)

Topology preserved regular superpixel (TPRS) [14] is a path-based method that partitions an image into superpixels by connecting seed points via pixel path. It begins by arranging initial seeds on a lattice grid and associating them with appropriate pixels on the boundary map. It then relocates each seed to the pixel with the highest locally maximal edge magnitudes, taking into account both the distance term and the probability term. Finally, it finds the local optimal path between vertical and horizontal seed pairs.

III. Experimental Setting

To determine the best superpixel size, TPRS will generate a test image with five superpixel sizes: 500, 1000, 1500, 2000, and 2500. Aside from that, each superpixel image will be paired with the following user input type:

- s1: 1 foreground stroke and 3 background strokes.
- s2: multiple foreground and background strokes.
- m2: 1 background bounding box and 2 foreground strokes.
- m3: 1 background bounding box and 2 foreground strokes.

The user input type and superpixel image as well as test image will be fed into the MSRM interactive segmentation algorithm in order to generate segmentation results. Fig. 2 shows some of test image with ground truth (see (a) and (b)). Aside from that, s1 and s2 input strokes from [35] were chosen, as shown in Fig. 2(c) and (d). Fig. 2(e) and (f) show the use of a bounding box with two and three foreground strokes, respectively. Fig. 3 shows the superpixel image generated by TPRS with sizes of 500 in (a) and 2500 in (b).

To assess the efficacy of superpixel algorithms in interactive segmentation, the error rate, F-score, and Jaccard indexes used by [31, 36-38]. Error rate (ERR) is the percentage of pixels placed in an incorrect region which is shown as below equation:

$$ERR = 1 - \left(\frac{TP+TN}{TP+TN+TP+FN}\right) \quad (1)$$

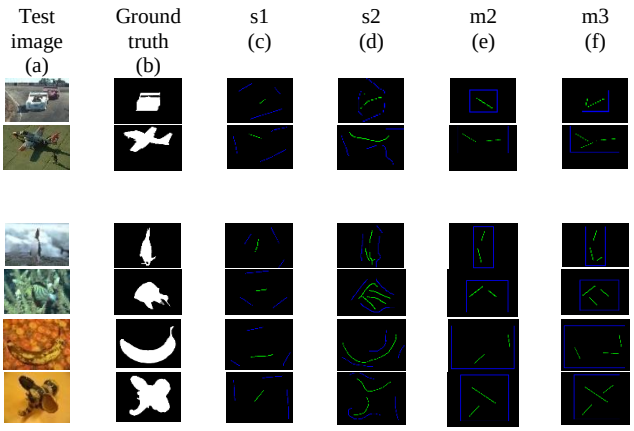

Fig. 2. (a) Test Image. (b) Ground Truth. (c) Simple Stroke. (d) Complex Stroke. (e) Bounding Box with Two Foreground Strokes. (f) Bounding Box with Three Foreground Strokes.

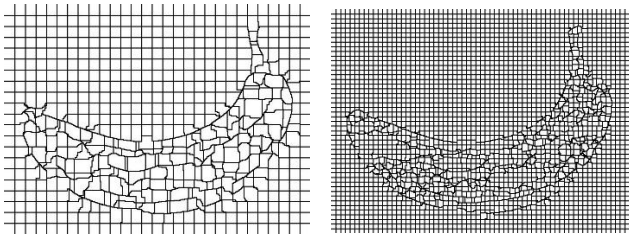

Fig. 3. Superpixel Image of Banana1 from the Dataset which had been Generated by TPRS. (a) 500 (b) 2500.

However, error rate takes into account the percentage of pixels that accurately map to the background information. As a result, the F-score and Jaccard Index are included.

F-score is equivalent to Dice Coefficient. The F-score is also known as the F1-Score or F-Measure. It is equal to 2 * the Area of Overlap divided by the total number of pixels in both images.

$$P = \frac{TP}{TP+FP} \quad (2)$$

$$R = \frac{TP}{TP+FN} \quad (3)$$

$$F = 2 * \left(\frac{P*R}{P+R}\right) \quad (4)$$

The Jaccard index, also known as the Intersection over Union (IoU) metric. It is the ratio of the number of pixels that are shared by X and Y to the total number of pixels in X and Y. In this case, X and Y represent the segmented image and ground truth, respectively. The Jaccard index/ IoU formulation is depicted as follows:

$$J/IOU = \frac{TP}{TP+FP+FN} \quad (5)$$

IV. Results and Discussion

To have better understanding of the result generated from various size as well as input type, the table below (see Table I) had divided the result based on the superpixel size, s and user input type, t. following by Error rate, e, f-score, f and Jaccard index, j. The error rate, F-score, and Jaccard index are also depicted graphically in Fig. 4 and 5.

TABLE I. Overall Image Segmentation Result based on Three Metrics: Error Rate, e; F, f-score; and Jaccard Index, J Performed by MSRM by using TPRS Superpixel Algorithms that Generated Different Numbers of Superpixels, S Ranging between 500, 1000, 1500, 2000, and 2500 by using Different Types of Input Types: M2 (Bounding Box with Two Foreground Strokes), M3 (Bounding Box with Three Foreground Strokes), S1 (Simple Strokes) and s2 (Complex Strokes)

s	t	e ↓	p ↑	r ↑	f ↑	j ↑
500	m2	0.037	0.943	0.842	0.882	0.803
500	m3	0.030	0.939	0.893	0.913	0.845
1000	m2	0.034	0.947	0.856	0.890	0.816
1000	m3	0.025	0.943	0.919	0.929	0.872
1500	m2	0.036	0.955	0.833	0.878	0.802
1500	m3	0.027	0.954	0.886	0.915	0.851
2000	m2	0.030	0.950	0.867	0.896	0.830
2000	m3	0.025	0.950	0.904	0.923	0.863
2500	m2	0.032	0.954	0.842	0.881	0.811
2500	m3	0.024	0.952	0.898	0.921	0.863
500	s1	0.059	0.911	0.776	0.805	0.712
500	s2	0.036	0.907	0.901	0.900	0.825
1000	s1	0.064	0.935	0.745	0.789	0.695
1000	s2	0.029	0.941	0.914	**0.924**	**0.864**
1500	s1	0.066	0.928	0.740	0.778	0.685
1500	s2	0.030	0.949	0.901	0.921	0.859
2000	s1	0.068	0.953	0.704	0.759	0.668
2000	s2	0.029	0.943	0.902	0.918	0.856
2500	s1	0.064	0.943	0.699	0.751	0.658
2500	s2	**0.028**	0.958	0.891	0.919	0.857

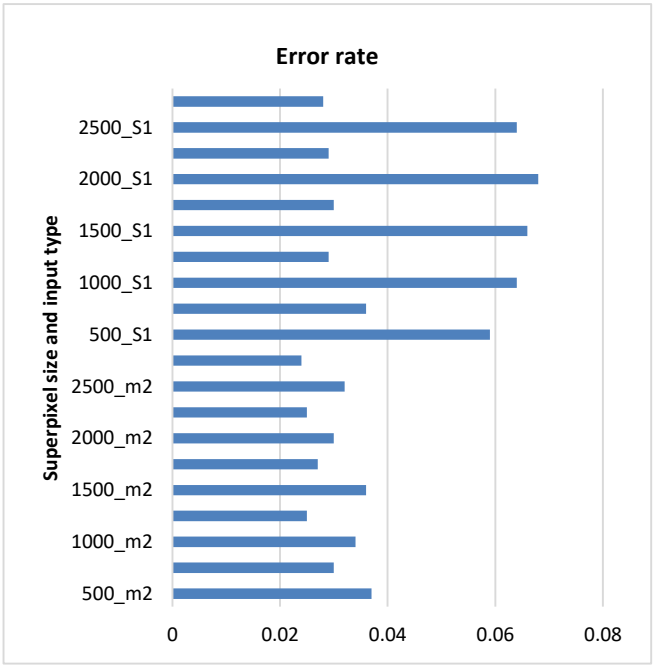

Fig. 4. Segmentation Results based on Error Rate using Various user Input T and Superpixel Size.

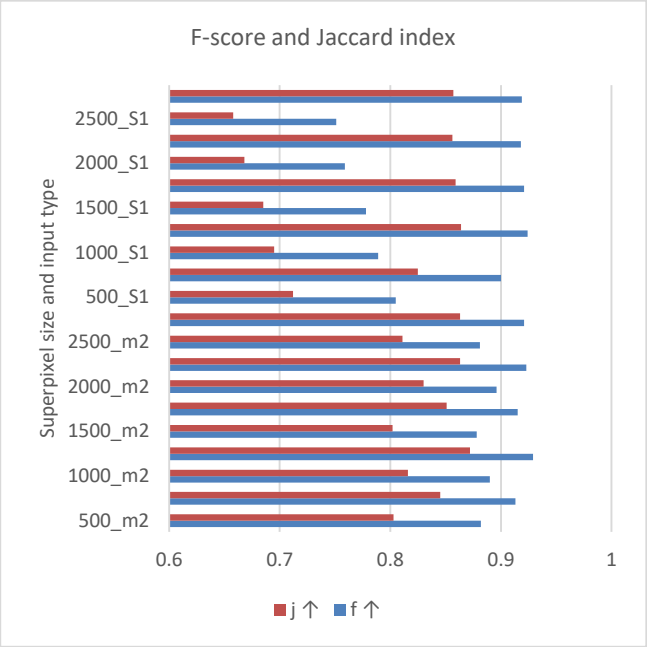

Fig. 5. Segmentation Results based on F-score and Jaccard Index using Various user Input Types and Superpixel Size.

From the Table I, it can summarize the finding of research study as below:

- In term of input type, bounding box input type had performed better than strokes with error rate 0.024 and 0.028, f-score 0.929 and 0.924, and Jaccard index 0.872, 0.864, respectively.
- Regardless of input type, superpixel size of 1000 had outperformed than other superpixel size in term of Jaccard index and F-score. On the other hand, if using error rate as a metric, size of 2500 is much better than size of 1000. However, the difference between only 0.001. Also, if comparing using f-score (0.005-0.008) and Jaccard index (0.007-0.009). It can conclude that size of 1000 should be an optimum size for TPRS.
- The increasing of input strokes for foreground had improved the segmentation results regardless of bounding box or strokes.
- The difference between the highest and lowest result for each category (s1, s2, m2, m3) of each matric are 0.006 and 0.009, error rate, 0.016 and 0.054, f-score and 0.024 and 0.054, Jaccard index.
- Overall, the difference between the highest and lowest result of all are 0.044, error rate, 0.178, f-score, and 0.214, Jaccard index.

The individual image results based on the 1000 and 2500 superpixels for strokes and bounding box categories shown in Table II. These two sizes from two categories were chosen due to the fact that they had achieved the best result among others. From the Table II, it is shown that 153077 had achieved high error rate (>0.1) if using bounding box while 189080 and stone2 if using strokes.

TABLE II. INDIVIDUAL IMAGE RESULT (ERROR RATE, E) BASED ON THE 1000 AND 2500 SUPERPIXELS FOR STROKES AND BOUNDING BOX CATEGORIES

filename	1000_m3	2500_m3	1000_s2	2500_s2
106024	0.023	0.019	0.016	0.016
124084	0.030	0.026	0.030	0.025
153077	0.142	0.131	0.058	0.089
153093	0.030	0.043	0.033	0.034
181079	0.030	0.017	0.030	0.054
189080	0.020	0.020	0.135	0.131
208001	0.012	0.009	0.010	0.012
209070	0.045	0.051	0.030	0.041
21077	0.028	0.033	0.016	0.020
227092	0.013	0.024	0.013	0.032
24077	0.030	0.035	0.024	0.047
271008	0.030	0.030	0.015	0.026
304074	0.031	0.058	0.044	0.019
326038	0.031	0.027	0.060	0.054
37073	0.032	0.047	0.020	0.030
376043	0.026	0.029	0.017	0.021
388016	0.017	0.014	0.059	0.018
65019	0.010	0.014	0.006	0.018
69020	0.052	0.038	0.066	0.045
86016	0.006	0.007	0.032	0.007
banana1	0.037	0.031	0.028	0.015
banana2	0.035	0.030	0.018	0.057
banana3	0.025	0.028	0.024	0.029
book	0.029	0.073	0.024	0.026
bool	0.028	0.028	0.013	0.039
bush	0.033	0.020	0.016	0.024
ceramic	0.071	0.035	0.018	0.026
cross	0.013	0.009	0.009	0.020
doll	0.017	0.004	0.003	0.018
elefant	0.040	0.010	0.010	0.015
flower	0.009	0.034	0.049	0.011
fullmoon	0.003	0.018	0.024	0.003
grave	0.016	0.010	0.008	0.009
llama	0.016	0.007	0.038	0.064
memorial	0.020	0.009	0.016	0.022
music	0.014	0.014	0.011	0.010
person1	0.013	0.012	0.020	0.010
person2	0.009	0.026	0.007	0.011
person3	0.010	0.006	0.015	0.010
person4	0.018	0.015	0.007	0.020
person5	0.008	0.047	0.017	0.018
person6	0.017	0.006	0.005	0.024
person7	0.008	0.007	0.012	0.006
person8	0.016	0.007	0.007	0.017
scissors	0.045	0.034	0.017	0.064
sheep	0.005	0.029	0.030	0.004
stone1	0.009	0.026	0.021	0.007
stone2	0.007	0.037	0.178	0.007
teddy	0.016	0.013	0.049	0.034
tennis	0.033	0.018	0.037	0.030

TABLE III. PERFORMANCE COMPARISON WITH OTHER STATE-OF-ART INTERACTIVE SEGMENTATION ALGORITHMS BASED ON ERROR RATE, E AND F-SCORE, F

Reference algorithms	e	Reference algorithms	f
Extreme points [37]	0.023	GSC(reported in [39])	0.966
Diffusive likelihood [38]	0.023	BNQ [39]	0.947
2500_m3	0.024	Region-based nonparametric [40]	0.942
1000_m3	0.025	RW (reported in [39]	0.937
Probabilistic diffusion [41]	0.027	supercut average [33]	0.9356
1000_m3	0.025	NHO (reported in [40]	0.934
2500_s2	0.028	IGC (reported in [39]	0.933
1000_s2	0.029	Densecut [34]	0.932
Label propagation [42]	0.0321	1000_m3	0.929
Object Extraction from Bounding Box Prior [43]	0.032	LS (reported in [39])	0.926
Xia (reported in [43]	0.033	1000_s2	0.924
SBT with AT(reported in [43]	0.033	onecut(reported in [32]	0.923
RW with AT(reported in [44]	0.033	2500_m3	0.921
NHO (reported in [38]	0.034	SAL [45]	0.9207
MGC (reported in [38]	0.026	GrabCut(GMM) (reported in [34])	0.909
DEEPGC (reported in [37]	0.034	2500s2	0.919
PLL (reported in [38]	0.0349	grabcut (reported in [32]	0.916
TAM (reported in [38]	0.0364	milcut avg [33]	0.9136
Milcut (reported in[37])	0.036	ppbc (reported [32])	0.91
BoxPrior (reported in [37]	0.037	gbmr (reported in [40]	0.91
		RC (reported in [45])	0.9094
		RWR(reported in [45]	0.9057
		LC (reported in [45]	0.8839
		loosecut [32]	0.882
		ABS (reported in [45])	0.7155

TABLE IV. PERFORMANCE COMPARISON WITH OTHER STATE-OF-ART INTERACTIVE SEGMENTATION ALGORITHMS BASED ON JACCARD INDEX, J

Reference algorithms	j	Reference algorithms	j
grabcut (reported in [46])	0.91	SGML [47]	0.81
ppbc (reported in [46])	0.91	PD (reported in [47])	0.8
nc_cut(reported in [46])	0.91	onecut (reported in [46])	0.8
1000_m3	0.872	milcut (reported in [46])	0.8
1000_s2	0.864	RTPG (reported in [48])	0.76
2500_m3	0.863	SD (reported in [48])	0.76
2500_s2	0.857	NRW (reported in [47])	0.72
LPD [48]	0.85	TPG (reported in [47])	0.7
nc_cut0 (reported in [46])	0.85	LC (reported in [47])	0.68
FGML [47]	0.82	SMRW (reported in [47])	0.62

When compared to state-of-the-art interactive segmentation algorithms, 2500 m3 achieved 0.024, which is 0.001 higher than the best reference algorithm in the error rate category, which is a training-based algorithm (see Table III). It did, however, outperforms other training-based algorithms like Label propagation [43] (0.0321), Xia [44] (0.033), SBT with AT [44] (0.033,) RW with AT [45] (0.033), and DEEPGC [38] (0.034). Apart from this, in terms of the F-score (see Table III) and Jaccard index (see Table IV), there is a 4% difference between the best reference algorithm and our algorithm.

V. CONCLUSION AND FUTURE DIRECTION

The TPRS superpixels algorithm was used in this paper to generate various sizes of superpixels images on interactive segmentation algorithms, MSRM using strokes and a combination of bounding box with strokes as user input. The entire test images from the Grabcut dataset were used in the testing. Evaluation matrices such as error rate, e, f-score, f, and Jaccard index, J were used to evaluate the generated results. Overall, the interactive segmentation algorithm was able to achieve an optimal segmentation result by bounding the box with three strokes and using superpixel sizes of 1000 and 2500. This discovery will be useful for researchers who want to use superpixels, particularly on TPRS, with appropriate settings on the number of superpixels in an image on the interactive segmentation algorithm. The segmentation results, on the other hand, were compared to the results of the previous study, which determined that 500 is the optimal number of superpixels to use in interactive segmentation. The difference could be due to the number of test images used as well as the input types. Finally, we make several recommendations for future research:

- Comparative study of various types of superpixel algorithms with diverse input types on different interactive segmentation algorithms.
- Previous studies [49, 50] has found that image complexity can affect segmentation results. Existing interactive image segmentation algorithms have not addressed the relationship between image complexity and input type, as well as superpixel and size, and this can be investigated further.

ACKNOWLEDGMENT

The authors would like to thank Faculty of Computer Science and Information Technology, University of Malaysia Sarawak and Faculty of Science and Technology, i-CATS University College, for providing the facilities for this study.

REFERENCES

[1] Ren, X. and J. Malik. Learning a classification model for segmentation. in Computer Vision, IEEE International Conference on. 2003. IEEE Computer Society.

[2] Zhou, C., et al., An efficient two-stage region merging method for interactive image segmentation. Computers & Electrical Engineering, 2016. 54: p. 220-229.

[3] Ding, J.-J., et al. Real-time interactive image segmentation using improved superpixels. in 2015 IEEE International Conference on Digital Signal Processing (DSP). 2015. IEEE.

[4] Panagiotakis, C., et al., Interactive image segmentation based on synthetic graph coordinates. Pattern Recognition, 2013. 46(11): p. 2940-2952.

[5] Wu, J., et al. Milcut: A sweeping line multiple instance learning paradigm for interactive image segmentation. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014.

[6] Lu, H., et al., Spectral segmentation via midlevel cues integrating geodesic and intensity. IEEE transactions on cybernetics, 2013. 43(6): p. 2170-2178.

[7] Gueziri, H.-E., M.J. McGuffin, and C. Laporte, A generalized graph reduction framework for interactive segmentation of large images. Computer Vision and Image Understanding, 2016. 150: p. 44-57.

[8] Achanta, R., et al., Slic superpixels. 2010.

[9] Jian, M. and C. Jung, Interactive image segmentation using adaptive constraint propagation. IEEE transactions on image processing, 2016. 25(3): p. 1301-1311.

[10] Goh, K.L., et al. Sizes of Superpixels and their Effect on Interactive Segmentation. in 2021 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET). 2021. IEEE.

[11] Ning, J., et al., Interactive image segmentation by maximal similarity based region merging. Pattern Recognition, 2010. 43(2): p. 445-456.

[12] Bergh, M.V.d., et al. Seeds: Superpixels extracted via energy-driven sampling. in European conference on computer vision. 2012. Springer.

[13] Rother, C., V. Kolmogorov, and A. Blake, " GrabCut" interactive foreground extraction using iterated graph cuts. ACM transactions on graphics (TOG), 2004. 23(3): p. 309-314.

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

[15] Malmberg, F., et al., SmartPaint: a tool for interactive segmentation of medical volume images. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 2017. 5(1): p. 36-44.

[16] Ozdemir, F., et al., Interactive segmentation in MRI for orthopedic surgery planning: bone tissue. International Journal of Computer Assisted Radiology and Surgery, 2017. 12(6): p. 1031-1039.

[17] Yathunanthan Vasuki, E.-J.H., Peter Kovesi, Steven Micklethwaite, An interactive image segmentation method for lithological boundary detection: A rapid mapping tool for geologists. Computers & Geosciences, 2017. 100: p. 27-40.

[18] Ma, J., et al., A segmentation method for greenhouse vegetable foliar disease spots images using color information and region growing. Computers and Electronics in Agriculture, 2017. 142: p. 110-117.

[19] Heckel, F., et al., Interactive 3D medical image segmentation with energy-minimizing implicit functions. Computers & Graphics, 2011. 35(2): p. 275-287.

[20] Zhang, J., J. Zheng, and J. Cai. A diffusion approach to seeded image segmentation. in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2010. IEEE.

[21] Kim, T.H., K.M. Lee, and S.U. Lee. Generative image segmentation using random walks with restart. in European conference on computer vision. 2008. Springer.

[22] Bai, X. and G. Sapiro. A geodesic framework for fast interactive image and video segmentation and matting. in 2007 IEEE 11th International Conference on Computer Vision. 2007. IEEE.

[23] Wang, T., H. Wang, and L. Fan, A weakly supervised geodesic level set framework for interactive image segmentation. Neurocomputing, 2015. 168: p. 55-64.

[24] Ding, Z., et al., Adaptive fusion with multi-scale features for interactive image segmentation. Applied Intelligence, 2021: p. 1-12.

[25] Adams, R. and L. Bischof, Seeded region growing. IEEE Transactions on pattern analysis and machine intelligence, 1994. 16(6): p. 641-647.

[26] Meena, S., K. Palaniappan, and G. Seetharaman. User driven sparse point-based image segmentation. in 2016 IEEE International Conference on Image Processing (ICIP). 2016. IEEE.

[27] Song, G., H. Myeong, and K.M. Lee. Seednet: Automatic seed generation with deep reinforcement learning for robust interactive segmentation. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[28] Xu, N., et al. Deep interactive object selection. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

[29] Li, Z., Q. Chen, and V. Koltun. Interactive image segmentation with latent diversity. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

[30] Li, K. and W. Tao, Adaptive optimal shape prior for easy interactive object segmentation. IEEE Transactions on Multimedia, 2015. 17(7): p. 994-1005.

[31] Tang, M., et al. Grabcut in one cut. in Proceedings of the IEEE International Conference on Computer Vision. 2013.

[32] Yu, H., et al. Loosecut: Interactive image segmentation with loosely bounded boxes. in 2017 IEEE International Conference on Image Processing (ICIP). 2017. IEEE.

[33] Wu, S., M. Nakao, and T. Matsuda, SuperCut: Superpixel based foreground extraction with loose bounding boxes in one cutting. IEEE Signal Processing Letters, 2017. 24(12): p. 1803-1807.

[34] Cheng, M.-M., et al. Densecut: Densely connected crfs for realtime grabcut. in Computer Graphics Forum. 2015. Wiley Online Library.

[35] Andrade, F. and E.V. Carrera. Supervised evaluation of seed-based interactive image segmentation algorithms. in 2015 20th symposium on signal processing, images and computer vision (STSIVA). 2015. IEEE.

[36] Taha, A.A. and A. Hanbury, Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC medical imaging, 2015. 15(1): p. 1-28.

[37] Maninis, K.-K., et al. Deep extreme cut: From extreme points to object segmentation. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

[38] Wang, T., et al., Diffusive likelihood for interactive image segmentation. Pattern Recognition, 2018. 79: p. 440-451.

[39] Chen, D.-J., H.-T. Chen, and L.-W. Chang, Toward a unified scheme for fast interactive segmentation. Journal of Visual Communication and Image Representation, 2018. 55: p. 393-403.

[40] Wang, D., et al., Region-based nonparametric model for interactive image segmentation. IEEE Access, 2019. 7: p. 111124-111134.

[41] Wang, T., et al., Probabilistic diffusion for interactive image segmentation. IEEE Transactions on Image Processing, 2018. 28(1): p. 330-342.

[42] Breve, F., Interactive image segmentation using label propagation through complex networks. Expert Systems With Applications, 2019. 123: p. 18-33.

[43] Dai, L., et al. Object extraction from bounding box prior with double sparse reconstruction. in Proceedings of the IEEE International Conference on Computer Vision Workshops. 2015.

[44] Nguyen, T.N.A., et al., Robust interactive image segmentation using convex active contours. IEEE Transactions on Image Processing, 2012. 21(8): p. 3734-3743.

[45] Oh, C., B. Ham, and K. Sohn, Robust interactive image segmentation using structure-aware labeling. Expert Systems with Applications, 2017. 79: p. 90-100.

[46] Xian, M., et al., Neutro-connectedness cut. IEEE Transactions on Image Processing, 2016. 25(10): p. 4691-4703.

[47] Wang, T., et al., Global Manifold Learning for Interactive Image Segmentation. IEEE Transactions on Multimedia, 2021. 23: p. 3239-3249.

[48] Wang, T., et al., Interactive Image Segmentation Based on Label Pair Diffusion. IEEE Transactions on Industrial Informatics, 2020. 17(1): p. 135-146.

[49] Chai, S.S. and K.L. Goh, Evaluation of Different User Input Types in Interactive Segmentation. Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 2017. 9(2-10): p. 91-99.

[50] Goh, K.L., et al., A Comparative Study of Interactive Segmentation with Different Number of Strokes on Complex Images. International Journal on Advanced Science, Engineering and Information Technology, 2020. 10: p. 178-184.