

Faculty of Engineering and Information Technology
University of Technology Sydney

Image Co-saliency Detection and Co-segmentation from The Perspective of Commonalities

A thesis submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

by

Lina Li

March 2020

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This thesis is the result of me conducted jointly with Shanghai University as part of a collaborative Doctoral degree.

This research is supported by the Australian Government Research Training Program.

Signature of Candidate: _____
Production Note:
Signature removed
prior to publication.

Date: 2020.03.09

Acknowledgments

Foremost, I would like to express the deepest appreciation to my supervisor, Associate Prof. Jian Zhang, for the constant support and guidance for my Ph.D work. As a collaborative degree candidate, I would also like to thank Prof. Zhi Liu, my co-supervisor, for his patient guidance, and scientific advice. Without their generous support, this dissertation would not have been possible.

I would like to thank my labmates in Global Big Data Technology Center: Junjie Zhang, Yifan Zuo, Xiaoshui Huang, Muming Zhao, Hao Cheng, Lu Zhang, Yongshun Gong, and Zhibin Li and labmates in Shanghai University, for the help and discussions and for all the fun we have had in my PhD study period.

Last but not the least, many thanks would go to my parents and my sisters, for their support and confidence in me through all these years.

Contents

Certificate	i
Acknowledgment	ii
List of Figures	vi
List of Tables	xii
List of Publications	xiii
Abstract	xv
Chapter 1 Introduction	1
1.1 Background	1
1.2 Research Problems	7
1.2.1 Global consistency among all images	7
1.2.2 Co-saliency fusion and refinement	7
1.2.3 Simple image selection	8
1.2.4 Simple image guidance	8
1.2.5 Common features in CNNs	9
1.2.6 The simplification of the CNNs model	9
1.3 Research Contributions	10
1.4 Thesis Structure	11
Chapter 2 Relevant Theories and Related Work	12
2.1 Image Co-saliency Detection	12

2.1.1	Concepts	12
2.1.2	Image saliency detection	14
2.1.3	Image co-saliency detection	20
2.2	Image Co-segmentation	25
2.2.1	Concepts	25
2.2.2	Image segmentation	26
2.2.3	Image co-segmentation	31
2.3	Summary	35
 Chapter 3 Co-saliency Detection Based on Region-level Fusion and Pixel-level Refinement 36		
3.1	Motivations	36
3.2	Co-saliency Detection Model	39
3.2.1	Intra saliency	39
3.2.2	Region-level co-saliency	41
3.2.3	Pixel-level co-saliency	43
3.3	Experiments	45
3.4	Summary	54
 Chapter 4 Unsupervised Image Co-segmentation via Guidance of Simple Images 55		
4.1	Motivations and Significance	55
4.2	The Proposed Method	58
4.2.1	Image ranking	62
4.2.2	Simple image selection	67
4.2.3	Samples extraction	70
4.2.4	Complicated image segmentation	72
4.3	Experiments	74

4.3.1	Parameter setting	75
4.3.2	Results of the proposed method	75
4.3.3	Comparison with other methods	82
4.3.4	Discussion	85
4.4	Summary	86
 Chapter 5 Learn Image Object Co-segmentation with Multi-		
	scale Feature Fusion	87
5.1	Motivations and Significance	87
5.2	The Proposed Model	91
5.2.1	The reformed VGG network	92
5.2.2	Features extraction and fusion	93
5.2.3	Upsampling	94
5.3	Experiments	96
5.3.1	Experimental settings	96
5.3.2	Experimental results	97
5.4	Summary	99
 Chapter 6 Conclusions and Future Work 100		
6.1	Conclusions	100
6.2	Future Work	102
 Bibliography		103

List of Figures

1.1	An example of a saliency map. From the left to right: (a) Original images, (b) the saliency map generated by an image object saliency detection model, (c) ground truth	1
1.2	From the left to right: (a) the simple images and its segmentation results, (b) complicated images.	2
1.3	Illustrations of the definition of objects and common objects. .	3
1.4	The saliency maps of complicated images. From top to bottom: (a) Original images, (b) the saliency map generated by an image object single saliency detection model, (c) ground truth.	5
2.1	An example of the saliency map, and the shifted eye fixation results in one early work.	13
2.2	One general architecture of early works from (Itti, Koch & Niebur 1998).	15
2.3	An image is over segmented into regions in Saliency Tree (Liu, Zou & Le Meur 2014).	16
2.4	An overview of FCN.	19
2.5	The flowchart of a typical bottom-up image co-saliency detection model (Fu, Cao & Tu 2013).	21

2.6	The flowchart of an image co-saliency detection model based on fusion of saliency maps (Cao, Tao, Zhang, Fu & Feng 2014).	23
2.7	The flowchart of an image co-saliency detection model based on CNNs (Wei, Zhao, Bourahla, Li, Wu & Zhuang 2019).	25
2.8	A simple 2D segmentation example for a 3*3 image. The seeds are $O = v$ and $B = p$. The cost of each edge is reflected by the edges thickness.	27
2.9	Three examples of Grab-Cut. The user drags a rectangle loosely around an object. The object is then extracted automatically.	29
2.10	ASPP structure.	30
2.11	(a) Sparse feature extraction with standard convolution on a low resolution input feature map. (b) Dense feature extraction with atrous convolution with rate $r = 2$, applied on a high resolution input feature map.	31
2.12	The illustration of a deep CNNs for image object co-segmentation (Li, Jafari & Rother 2018).	35
3.1	Illustration of the proposed co-saliency model. From top to bottom: the image set (5 out of 36 images are shown), ultrametric contour maps, region segmentation results, intra saliency maps, initial region-level co-saliency maps, final region-level co-saliency maps, and pixel-level co-saliency maps.	40

3.2	Co-saliency maps generated for image pairs in the CP dataset. From top to bottom: original image pairs, binary ground truths, co-saliency maps generated using Li's model (Li & Ngan 2011), Fu's model (Batra, Kowdle, Parikh, Luo & Chen 2010), Liu's model (Liu, Zou, Li, Shen & Le Meur 2013) and our model, respectively.	46
3.3	Co-saliency maps generated for image pairs in the CP dataset. From top to bottom: original image pairs, binary ground truths, co-saliency maps generated using Li's model (Li & Ngan 2011), Fu's model (Batra et al. 2010), Liu's model (Liu et al. 2013) and our model, respectively.	47
3.4	Co-saliency maps generated for an image set in the iCoseg dataset. From top to bottom in each sub-figure: some original images in the image set, binary ground truths, co-saliency maps generated using Fu's model (Batra et al. 2010), Liu's model (Liu et al. 2013) and our model, respectively.	48
3.5	Co-saliency maps generated for an image set in the iCoseg dataset. From top to bottom in each sub-figure: some original images in the image set, binary ground truths, co-saliency maps generated using Fu's model (Batra et al. 2010), Liu's model (Liu et al. 2013) and our model, respectively.	49
3.6	Co-saliency maps generated for an image set in the iCoseg dataset. From top to bottom in each sub-figure: some original images in the image set, binary ground truths, co-saliency maps generated using Fu's model (Batra et al. 2010), Liu's model (Liu et al. 2013) and our model, respectively.	50

3.7	Co-saliency maps generated for an image set in the iCoseg dataset. From top to bottom in each sub-figure: some original images in the image set, binary ground truths, co-saliency maps generated using Fu's model (Batra et al. 2010), Liu's model (Liu et al. 2013) and our model, respectively.	51
3.8	Precision-recall curves for different saliency maps on the CP dataset.	52
3.9	Precision-recall curves for different saliency maps on the iCoseg dataset.	53
4.1	The flowchart. There are four steps in the proposed method: image ranking, simple image selection, samples extraction and complicated image segmentation.	59
4.2	The pre-processing. One image is over-segmented into hierarchical regions in three layers and is represented by two histograms: color histogram and dense SIFT histogram.	60
4.3	Region matching. For one region in layer 2 of image 1, its most similar region is searched through all the three layers in image 2.	62
4.4	The region matching results. The 1st row shows original images, the 1st column shows the regions hierarchically extracted from the three layers, and the 2nd-5th column show the corresponding best matched regions in other images.	63
4.5	Image Ranking. Images are ranked based on their saliency maps.	64

4.6	Image ranking results with different saliency models. The 1st, 2nd and 3rd row show the ranking results based on saliency maps generated using ST (Liu et al. 2014), HS (Yan, Xu, Shi & Jia 2013) and RBD (Zhu, Liang, Wei & Sun 2014), respectively.	67
4.7	Simple image selection. There are two boxes. Top box: the original ranked images and their corresponding single saliency maps. Bottom box: the selected simple images and their corresponding co-segmentation results.	68
4.8	Experimental results. In each image group, from top to bottom, RI: Ranked images, ISR: Initial segmentation results, and FSR: Final segmentation results.	76
4.9	Performance on iCoseg Dataset with Different Features.	77
4.10	Performance on iCoseg Dataset with Different Saliency Models	78
4.11	Experimental results. (a) the original images, from (b) to (d), the results of the proposed method, (Faktor & Irani 2013) and (Lee, Jang, Sim & Kim 2015), respectively.	80
4.12	Experimental results. (a) the original images, from (b) to (d), the results of the proposed method, (Faktor & Irani 2013) and (Lee et al. 2015), respectively.	81
4.13	Experimental results. (a) the original images, from (b) to (d), the results of the proposed method, (Faktor & Irani 2013) and (Lee et al. 2015), respectively.	82
4.14	Weighted F-measure on iCoseg Dataset	83
4.15	Intersection-over-Union(IoU) on iCoseg Dataset	83
4.16	F-measure on iCoseg Dataset	84

5.1	An overview of the proposed model. It can be viewed as three parts. The input images are first passed through the reformed VGG networks, which share exactly the same layer structure and parameters. Then, multi-scale [2,4,8,16] features are extracted at each layer (represented in colors and they are stacked according to the scales). They are fused both within each image (in grey) into intra-image features and across images (in colors) into inter-image features, and then intra-image and inter-image features are further fused at each scale, the features at different scales are represented in the corresponding colors. At last, the fused multi-scale features are summed up and upsampled to obtained the coarse co-segmentation results, and final co-segmentation results are obtained via Grabcut refinement.	88
5.2	Experimental results, objective evaluations.	97
5.3	Experimental results. (a) the original images, from (b) to (d), the results of (Li, Liu & Zhang 2018), the proposed model, and ground truth, respectively.	98
5.4	Experimental results. (a) the original images, from (b) to (d), the results of (Li, Liu & Zhang 2018), the proposed model, and ground truth, respectively.	98

List of Tables

4.1	Overall Performance on iCoseg Dataset	85
5.1	The implementation details of layers in 5.2.1.	92
5.2	The implementation details of the layers in Section 5.2.2. . . .	94
5.3	The implementation details of the layers in 5.2.3	95

List of Publications

Papers Published

- **Lina Li**, Zhi Liu, Wenbin Zou, Xiang Zhang, Olivier Le Meur(2014), Co-saliency detection based on region-level fusion and pixel-level refinement. *in* ‘Proceedings of the International conference on Multimedia and Expo (ICME’14)’, IEEE, pp. 1-6.
- **Lina Li**, Zhi Liu, Jian Zhang, (2018), Unsupervised image co-segmentation via guidance of simple images. *Neurocomputing*, vol. 275, pp. 1650-1661.
- **Lina Li**, Zhi Liu, Jian Zhang, Xiaofei Zhou, Learn Image Object Co-segmentation with Multi-scale Feature Fusion. *in* ‘Visual Communications and Image Processing (VCIP’19)’, IEEE, Accepted.
- Shuhua Luo, Zhi Liu, **Lina Li**, Xuemei Zou, Olivier Le Meur(2013), Efficient saliency detection using regional color and spatial information, *in* ‘4th European Workshop on Visual Information Processing (EUVIP’13)’, pp. 184-189.
- Zhi Liu, Wenbin Zou, **Lina Li**, Liquan Shen, Olivier Le Meur(2014), Co-saliency detection based on hierarchical segmentation, *IEEE Signal Processing Letters*, vol. 21, no. 1, pp. 88-92.

- Linwei Ye, Zhi Liu, **Lina Li**(2015), Evaluation on fusion of saliency and objectness for salient object segmentation, *in* ‘Proceedings of the 7th International Conference on Internet Multimedia Computing and Service(ICIMCS’15)’, ACM, p.18.
- Linwei Ye, Zhi Liu, **Lina Li**, Liquan Shen, Cong Bai, Yang Wang(2017), Salient object segmentation via effective integration of saliency and objectness. *IEEE Transactions on Multimedia*, vol. 19, no. 8. pp. 1742-1756.

Abstract

Image co-saliency detection and image co-segmentation aim to identify the common salient objects and extract them in a group of images.

Image co-saliency detection and image co-segmentation are important for many content-based applications such as image retrieval, image editing, and content aware image/video compression. The image co-saliency detection and image co-segmentation are very close works. The most important part in these two works is the definition of the commonality of the common objects. Usually, common objects share similar low-level features, such as appearances, including colours, textures shapes, etc. as well as the high-level semantic features.

In this thesis, we explore the commonalities of the common objects in a group of images from low-level features and high-level features, and the way to achieve the commonalities and finally segment the common objects. Three main works are introduced, including an image co-saliency detection model and two image co-segmentation methods.

Firstly, an image co-saliency detection model based on region-level fusion and pixel-level refinement is proposed. The commonalities between the common objects are defined by the appearance similarities on the regions from all the images. It discovers the regions that are salient in each individual image as well as salient in the whole image group. Extensive experiments

on two benchmark datasets demonstrate that the proposed co-saliency model consistently outperforms the state-of-the-art co-saliency models in both subjective and objective evaluation.

Secondly, an unsupervised images co-segmentation method via guidance of simple images is proposed. The commonalities are still defined by hand-crafted features on regions, colours and textures, but not calculated among regions from all the images. It takes advantages of the reliability of simple images, and successfully improves the performance. The experiments on the dataset demonstrate the outperformance and robustness of the proposed method.

Thirdly, a learned image co-segmentation model based on convolutional neural network with multi-scale feature fusion is proposed. The commonalities between objects are not defined by handcraft features but learned from the training data. When training a neural network with multiple input images simultaneously, the resource cost will increase rapidly with the inputs. To reduce the resource cost, reduced input size, less downsampling and dilation convolution are adopted in the proposed model. Experimental results on the public dataset demonstrate that the proposed model achieves a comparable performance to the state-of-the-art methods while the network has successfully gotten simplified and the resources cost is reduced.