UNIVERSITY OF TECHNOLOGY SYDNEY
Faculty of Engineering and Information Technology

# LEARNING FOR OBJECT LOCALIZATION WITH IMPERFECT DATA

by

**Xiaolin Zhang**

A THESIS SUBMITTED
IN FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

**Doctor of Philosophy**

Sydney, Australia

2021

# CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Xiaolin Zhang, declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Computer Science at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

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

Signature: Production Note: Signature removed prior to publication.

Date: 31 May 2021

# Acknowledgements

I had a remarkable time at UTS pursuing my Ph.D. degree. My research journey have been truly amazing thanks to all the great people who have helped and supported me. I would like to express my sincere thanks to them all.

First and foremost, I would thank my supervisor, Prof. Yi Yang, for patient guidance and kind encouragement. He has given me a lot of suggestions for my research and future career. It is really the luckiest thing to have had him as my supervisor.

Also, I want to thank my co-supervisor, Dr. Yunchao Wei. He guided me through the realm of deep learning. His patience and kindness give me great support to complete every paper and finally finish my Ph.D. research work.

Then, I would like to thank the late Thomas S. Huang. I was so lucky to have met such a honorable scientist and have him as my advisor when I visited the University of Illinois Urbana-Champaign. His wisdom, kindness, and admirable faith in love really moved me.

And, I would like to thank my colleagues and friends for the help and support. I am delighted to have spent four great years in Sydney with all these lovely people.

I appreciate the financial support from the CSC-UTS Program.

Finally, I would like to express my deepest thanks to my parents, sister, and girlfriend, for their trust and love.

Xiaolin Zhang

November 2020 at UTS.

# List of Publications

### Conference Papers

C-1. **X. Zhang**, Y. Wei, J. Feng, Y. Yang and T. Huang, "Adversarial Complementary Learning for Weakly Supervised Object Localization," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

C-2. **X. Zhang**, Y. Wei, G. Kang, Y. Yang and T. Huang, "Self-produced guidance for weakly-supervised object localization," *Proceedings of the European Conference on Computer Vision*, 2018.

C-3. **X. Zhang**, Y. Wei and Y. Yang, "Inter-Image Communication for Weakly Supervised Localization," *Proceedings of the European Conference on Computer Vision*, 2020.

### Journal Papers

J-1. **X. Zhang**, Y. Wei, Y. Yang and T. Huang, "SG-One: Similarity Guidance Network for One-Shot Semantic Segmentation," *IEEE Transactions on Cybernetics*, vol. 50, 2020.

J-1. **X. Zhang**, Y. Wei, Z. Li, C. Yan and Y. Yang, "Rich Embedding Features for One-Shot Semantic Segmentation," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

### Submitted Papers

J-1. **X. Zhang**, Y. Wei, Y. Yang and F. Wu, "Rethinking Localization Map: Towards Accurate Object Perception with Self-Enhancement Maps," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Major revision, 2021.

# Contents

# List of Figures

# Abbreviation

AP - Average Precision

bbox - Bounding Boxes

CAM - Class Activation Map

CNN - Convolutional Neural Network

DNN - Deep Neural Network

FCN - Fully Convolutional Network

FSL - Few-Shot Learning

FSSS - Few-Shot Semantic Segmentation

GAP - Global Average Pooling

IoU - Intersection-over-Union

PR - Precision Recall

WSL - Weakly Supervised Learning

WSOL - Weakly Supervised Object Localization

WSOD - Weakly Supervised Object Detection

WSSS - Weakly Supervised Semantic Segmentation

# ABSTRACT

**LEARNING FOR OBJECT LOCALIZATION WITH IMPERFECT DATA**

by

Xiaolin Zhang

Deep learning has achieved countless remarkable successes in recent years. Learning deep neural networks usually needs tremendous well-labeled examples, which requires intensive investments. A feasible solution for reducing the budget is to learn from imperfect data, *e.g.*, noisy data, synthetic data, weak labels, and datasets with few annotated examples. This thesis dedicates to the weakly supervised learning and few-shot learning.

The first task is to address the challenging object localization problem using weak annotations as supervision. Objects in images are expected to be precisely located with only image-level labels, *i.e.*, category information. Specifically, convolutional networks can only find the most discriminative object regions leading to the unsatisfied predictions of bounding boxes. This thesis tries to solve this problem in three perspectives: 1) forcing the networks to mine more object areas by erasing the discovered object pixels; 2) learning pixel correlations within images under the supervision of self-produced object masks ; 3) communicating with different images to obtain more consistent features, and therefore, activating target object more accurately.

The second task is to predict the semantic masks of objects in a few-shot approach. Finding every pixel of target objects can also be considered as the most delicate localization problem. In the few-shot regime, only few annotated examples are available for an unseen class, and networks are required to locate the semantic category of each pixel with minimal information. This thesis will present two approaches to improve the quality of predicted object masks. Notably, a similarity-guided network is proposed to endow the segmentation process with rough position cues for locating the object pixels. To enhance the guidance process and improve the robustness, we further enrich the guidance embed-

dings and propose to employ multiple diverse support vectors to generate the similarity maps.

In addition, each of the proposed methods is comprehensively verified and analyzed by conducting various experiments.