UNIVERSITY OF TECHNOLOGY SYDNEY
Faculty of Engineering and Information Technology

# Exploring Region-based Deep Learning to Understand Objects in Real-world Scenarios

by

**Ruiheng Zhang**
**Supervisor: A/Prof. Min Xu**

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree

**Doctor of Philosophy**

Sydney, Australia

2021

# Certificate of Authorship/Originality

I, Ruiheng Zhang declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Electrical and Data Engineering, Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced 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.

Also, I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree at any other academic institution except as fully acknowledged within the text. This thesis is the result of a Collaborative Doctoral Research Degree program with Beijing Institute of Technology, China.

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

Production Note:

Signature: Signature removed prior to publication.

© Copyright 2021 Ruiheng Zhang

To my loving parents, my dear grandparents, my beloved wife and daughter.

# ABSTRACT

**Exploring Region-based Deep Learning to Understand Objects in Real-world Scenarios**

by

Ruiheng Zhang

Supervisor: A/Prof. Min Xu

One way to infer about the real scenes is by understanding the object that presents in it, involving object localization, object recognition, object tracking, etc. Despite many advances in computer vision techniques, object understanding in real-world scenarios still remains many challenging tasks. There is no universal algorithm that can solve all of the scenarios with their own practical difficulties. This dissertation focuses on exploring region-based deep learning to understand objects in three typical real-world scenarios.

The first part of the dissertation studies facial landmark detection in the condition of lack of finely labeled training data. We generate weakly labeled training data to replace finely labeled data using generative adversarial networks. Then, we propose a region-based convolutional neural network to detect facial components and landmarks simultaneously. Notably, our approach can handle the situation when large occlusion areas occur, as we localize visible facial components before predicting corresponding landmarks. Extensive evaluations on several datasets indicate the effectiveness of the proposed approach.

In the second part, multi-player identification and tracking tasks in sports video are discussed. We build a robust multi-camera multi-player tracking with identification framework, from player detection, to identification, to tracking. To handle the identity switches, we design a distinguishable deep representation for player identity, considering pose-guided partial features, team class, and jersey number. For data association, a robust multi-player tracker incorporating with player identity is further developed to produce identity-coherent trajectories. Experiment results illustrate that our framework handles the identity switches effectively, and outperforms state-of-the-art trackers on the sports video benchmarks.

Finally, we study vehicle detection in infrared images with poor texture information, low resolution and high noise levels. To deal with these difficulties, we propose a backbone network to exploit discriminative features, composing of a frequency feature extractor, a spatial feature extractor and a dual-domain feature resource allocation model. Hypercomplex Infrared Fourier Transform is developed to calculate the infrared intensity saliency, while a convolutional neural network is used to extract feature maps in the spatial domain. To efficiently integrate and recalibrate the frequency and spatial features, we propose a Resource Allocation model for Features based on the well-designed attention blocks. The experiments substantiate the merits of the proposed method through comparisons with state-of-the-art methods.

Dissertation directed by Associate Professor Min Xu
School of Electrical and Data Engineering

# Acknowledgements

First and foremost, I want to extend my heartfelt gratitude to my supervisor, A/ Prof. Min Xu. Her kind supervision, valuable suggestions and warm encouragement helped me to overcome difficulties and to successfully complete this thesis. She had empowered and inspired me to work on multiple research projects which led to publications in top journals and conferences. She will always be my most beloved supervisor. I have learned so much from her, not only how to be a successful scholar but also how to be a kind person in life. I also would like to appreciate my co-supervisor Dr. Xiaoying Kong for providing me with continuous support throughout my study and research.

Second, I wish to express my sincere appreciation to Prof. Lixin Xu and A/ Prof. Chengpo Mu, my Dual-PhD supervisors at Beijing Institute of Technology. Although we are far away from each other, they gave me many suggestions and much help.

Third, I would like to pay tribute to Prof. Dacheng Tao, Prof. Dayong Jin, Prof. Linlin Ge, Prof. Sean He, Prof. Qiang Wu, Prof. Shenghong Hu, Dr. Yu Peng, Dr. Wenguan Wang, Dr. Yang Yu, and Dr. Zhengyu Yu for their inspiring words, continued support, and elaborate instructions about my presentations and academic reports. They had a significantly positive influence on me during my Ph.D. career.

Forth, I thank my fellow lab mates and my excellent friends: Tianrong Rao, Haimin Zhang, Lingxiang Wu, Wanneng Wu, Yukun Yang, Lei Sang, Zhongqin Wang, Xiaoxu Li, Qiyu Liao, Shuo Yang, Jiahao Xia, Caoyuan Li, Yang He, Yaxin Shi, Shilei Zhou, Huan Li, Wei Wei, Xiaofeng Xu, Qi Zhang, Zhibin Li, Yan Huang, Huidong Xu, Deyang Liu, Yuexi Zhang, Jiayan Qiu, Ye Shi, Yulong Sun and others that I cannot list them all for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the funs we have had.

I thank, School of Electrical and Data Engineering, Faculty of Engineering and IT, University of Technology Sydney for providing the infrastructures, computing power, and financial support.

Finally, I am grateful for my dear family, especially my father Bin Zhang, my mother Juan Li, my grandpa Wenhua Li, my grandma Yachun Liu, my father in law Shengchun Shi, my mother in law Conghua Lin, my uncle Donghai Zhang and Yunhai Ma, my brother Zhe Yuan and Zichao Ma, who had been supportive during my PhD candidature stages, researching and writing my thesis. I am sincerely grateful for everything and especially the opportunities given in this life, from being born to the completion of this Doctor of Philosophy thesis.

Special thanks to my wife Dr. Qiaolin Shi, who is always being with me. She was the one who encouraged me to grow as a real man and gave the pretty birth to my beloved daughter Yuxi Zhang.

Ruiheng Zhang

Sydney, Australia, 2020.

# List of Publications

## Journal Papers

J-1. **R. Zhang**, L. Wu, Y. Yang, W. Wu, Y. Chen, M. Xu, 'Multi-camera Multi-player Tracking with Deep Identification in Sports Video', *Pattern Recognition*, Volume 102, 2020, 107260, ISSN 0031-3203.

J-2. **R. Zhang**, C. Mu, M. Xu, L. Xu, Q. Shi, J. Wang, 'Synthetic IR Image Refinement using Bidirectional Mappings with Adversarial Learning', *IEEE Access*, 2019, 7: 153734-153750.

J-3. **R. Zhang**, C. Mu, M. Xu, L. Xu, X. Xu, 'Facial Component-Landmark Detection with Weakly-supervised LR-CNN', *IEEE Access*, 2019, 7: 10263-10277.

J-4. **R. Zhang**, L. Xu, Z. Yu, Y. Shi, C. Mu, M. Xu, 'Deep-IRTarget: An Automatic Target Detector in Infrared Imagery using Dual-domain Feature Extraction and Allocation', *IEEE Transactions on Multimedia*, 2021.

J-5. Q. Liang, W. Wu, Y. Yang, **R. Zhang**, Y. Peng, M. Xu, 'Multi-Player Tracking for Multi-View Sports Videos with Improved K-Shortest Path Algorithm', *Applied Sciences*, 2020, 10, 864.

J-6. Q. Zhang, **R. Zhang**, L. Ge, M. Xu, 'A High Accuracy Burned Area Detection Framework Based on The Joint Processing of Sentinel-1&2 Data', *Remote Sensing of Environment.* (under review)

J-7. Y. Yang, **R. Zhang**, W. Wu, M. Xu, '3D Localization for Multiple Players in Multiview Sport Videos with Deep Identification Reasoning', *Pattern Recognition.* (under review)

**Conference Papers**

C-1. **R. Zhang**, M. Xu, Y. Shi, J. Fan, C. Mu, L. Xu, 'Infrared Target Detection using Intensity Saliency and Self-Attention', *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020.

C-2. Y. Yang, **R. Zhang**, W. Wu, Y. Peng, M. Xu, 'Multi-camera sports players 3D localization with identification reasoning', *2020 25 th IEEE International Conference on Pattern Recognition(ICPR)*. IEEE, 2020.

C-3. Y. Yang, M. Xu, W. Wu, **R. Zhang**, and Y. Peng, '3D Multiview Basketball Players Detection and Localization Based on Probabilistic Occupancy', *2018 Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 2018.

# Contents

# List of Figures

# List of Tables