UNIVERSITY OF TECHNOLOGY SYDNEY
Faculty of Engineering and Information Technology

Hierarchical Convolutional Neural Networks for Vision-Based Feature Detection

by

Qiuchen Zhu

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

Doctor of Philosophy

Sydney, Australia

2022

Certificate of Authorship/Originality

I, Qiuchen Zhu 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.

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

Signature: Production Note: Signature removed prior to publication.

Date: 04/05/2022

ABSTRACT

Hierarchical Convolutional Neural Networks for Vision-Based Feature Detection

by

Qiuchen Zhu

This thesis is devoted to the problem of feature detection, an essential prerequisite to machine vision applications. The key to feature detection rests with the development of effective algorithms, which could incorporate machine intelligence to achieve such attributes as accuracy in pixel-wise terms and robustness against structural and stochastic uncertainty. To this end, a hierarchical convolutional neural network (HCNN) with feature preservation, is proposed to present the probability map of feature candidates. Specifically, the abstraction of features in consideration is enhanced by bidirectional branch nets. The outputs of previous convolutional blocks are unified and concatenated to the current ones to reduce the visual impairment in the up/down-sampling stage and the overall information loss. Besides, an Intercontrast-based Iterative Thresholding (IIT) approach is developed for the proposed network hierarchy at the post-processing step, whereby patterns of interest are clustered within the probability map of identified features and generate a solid feature map. To effectively overcome uncertainty, network prediction is conducted by a customised variational inference. Here, deterministic weights are converted into a probability distribution with learnable hyperparameters to adapt the interference of outliers and alleviate overfitting. Furthermore, to incorporate Bayesian modelling into high-level tasks such as resource allocation, an additional module for Gaussian heatmap is developed to meticulously present the location of the geometrical target. Then, a physics-driven training scenario is designed to gradually shrink the benchmark kernel for continuous calibration to avoid local minima. In summary, the contributions of this thesis include 1. a new hierarchical network proposed for feature detection, whereby the abstractions of the image can be bidirectionally extracted to improve prediction performance, 2. unsupervised and gradient-sharing approaches incorporating Bayesian inference in the proposed network for enhancing its uncertainty handling capability, 3. a new training strategy for representation learning via spatial indexing to link the primary geometrical features with quantitative allocation, and 4. an average F-measure proposed for evaluation of robustness along with other metrics for performance evaluation. An extensive comparison with existing techniques is conducted using various datasets and evaluation criteria for evaluation. Experimental results demonstrate the crack detection merits of the proposed architecture over existing techniques applied to numerous images. To illustrate generality of the developed network architecture, additional tests are also conducted for various applications, including salient object detection, anthropometric and facial landmark detection, and measurement retrieval. The results obtained show the scalability and robustness of the proposed model to medium and high-level image processing tasks.

Dissertation directed by Professor Quang Ha
School of Electrical and Data Engineering, University of Technology Sydney

Dedication

To my parents Jialin Zhu and Bin Jia, and my supervisor Professor Quang Ha.

Acknowledgements

First and foremost, I would like to thank my supervisor Quang Ha for his patient guidance and encouragement for this work. I respectfully acknowledge Dinh Tran Hiep, Manh Duong Phung and Ricardo P. Aguilera for the technical support, and thank my fellow group-mates for the stimulating discussions and nice team work. I would also like to pay my respect to Ngaiming Kwok for his invaluable suggestions on this thesis. I would like acknowledge University of Technology Sydney, Chinese Scholarship Council, SmartFit and ApparelTech for the financial assistance to my research.

Finally, My special thanks to my family for the spiritual support throughout my PhD candidature especially during the hard time of the global epidemic.

Qiuchen Zhu
Sydney, Australia, 2022.

List of Publications

Journal Papers

J-1. **Q. Zhu**, T. H. Dinh, M. D. Phung, and Q. Ha, "Hierarchical Convolutional Neural Network with Feature Preservation and Autotuned Thresholding for Crack Detection," *IEEE Access*, vol. 9, pp. 60201-60214, 2021, DOI: 10.1109/ACCESS.2021.3073921.

J-2. **Q. Zhu** and Q. Ha, "A Bidirectional Self-Rectifying Network with Bayesian Modelling for Vision-Based Crack Detection," *IEEE Transactions on Industrial Informatics*, accepted on 30 APR 2022.

J-3. T. H. Dinh, **Q. Zhu**, M. D. Phung and Q. Ha, "Summit Navigator Automatic Thresholding for Image Binarization with Application to Crack Detection," *IEEE Transactions on Systems, Man and Cybernetics: Systems*, revised and resubmitted.

Conference Papers

C-1. **Q. Zhu**, and Q. Ha, "A Bidirectional Self-Rectifying Network with Bayesian Modelling for Feature Detection and Keypoint Detection," *The 21th International Conference on Machine Learning and Cybernetics*, Adelaide, SA, Australia, December 4-5, 2021, Doi: 10.1109/ICMLC54886.2021.9737243.

C-2. T. X. Tran, T. H. Dinh, H. V. Le, **Q. Zhu**, and Q. Ha, "Defect Detection Based on Singular Value Decomposition and Histogram Thresholding," in *Proc. 2020 IEEE/ASME International Conf. on Advanced Intelligent Mechatronics (AIM)*, Boston, MA, USA, July 6-9, 2020, pp. 149-1154.

C-3. **Q. Zhu**, M. D. Phung, and Q. Ha, "Crack Detection Using Enhanced Hierarchical Convolutional Neural Networks," *Proc. Australasian. Conf. on*

Robotics and Automation, Adelaide, SA, Australia, December 11-13, 2019, pp. 1-8. **Best Paper Award Winner**.

C-4. V. T. Hoang, M. D. Phung, T. H. Dinh, **Q. Zhu** and Q. Ha, "Reconfigurable Multi-UAV Formation using Angle-Encoded PSO," in *Proc. 2019 IEEE 15th International Conf. on Automation Science and Engineering (CASE)*, Vancouver, Canada, August 22, 2019, pp. 1670-1675.

C-5. **Q. Zhu**, T. H. Dinh, V. T. Hoang, M. D. Phung, and Q. Ha, "Crack Detection Using Enhanced Thresholding on UAV based Collected Images," in *Proc. Australasian. Conf. on Robotics and Automation*, Lincoln, Canterbury, New Zealand, Dec. 9-11, 2018, pp. 1-7.

Contents

List of Figures

Abbreviation

Adam: Adaptive moment estimation

BSNBM: Bidirectional self-rectifying network with Bayesian modelling

CBAT: Contrast-based autotuned thresholding

CNN: Convolutional neural network

CRF: Conditional random field

DCNN: Deep convolutional neural networks

DCB: Dilated convolutional block

ELBO: Evidence lower bound

FCN: Fully convolutional network

F/REB: Forward/reverse enhancement branch

HCNNFP: Hierarchical convolutional neural networks with feature preservation

JI: Jaccard Index

KL: Kullback-Leibler

MAPE: Mean absolute error

MAPE: Mean absolute percentage error

MLP: Multilayer perceptron

MSE: Mean square error

RNN: Recurrent neural network

RCNN: Regional convolutional neural network

ReLU: Rectified linear units

RF: Receptive field

ROI: Region of interest

Nomenclature and Notation

Lower-case italic alphabets denote scalar values or vectors.

Upper-case italic letters denote matrices denotes matrices.

In a sequence, lower-case and capital italic letters respectively indicate the index and the maximum value.

Roman or calligraphic letters indicate a function.

$\hat{x}$ represents the estimated/predicted value.

$*$ denotes the convolutional operation.

$\odot$ is the component-wise multiplication.

$\|\cdot\|_F$ denotes the Frobenius norm.

w represents the scalar weight.

c_o denotes the coordinates of landmarks.

I is the identity diagonal matrix.

R represents the pixel region of interest.

$\mathit{\Omega}$ represents a sample space.

AF_β is the average F-measure.

F_β is the F-measure.

$\mathrm{h}_{\mathrm{smx}}$ represents the spatial softmax function.

P denotes the probability function.

$\mathbb{E}$ denotes the mathematical expectation.

$\mathcal{F}$ is the Fourier transformation.

$\mathcal{L}$ represents the loss function.

$\mathcal{N}$ represents the Normal/Gaussian distribution.