Improved Fine-Grained Representation Learning with Data Transformation

by Lianbo Zhang

Thesis submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy
under the supervision of
by A/Prof. Wei Liu and Prof. Dacheng Tao

University of Technology Sydney
Faculty of Engineering and Information Technology
November 2021

To my beloved parents
Chuanyong Zhang and *Jianzhen Li*

Certificate of Original Authorship

I, Lianbo 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/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.

Production Note:
Signature: Signature removed prior to publication.

Date: 23/11/2021

Publications During Enrolment

1. **Lianbo Zhang**, Shaoli Huang, Wei Liu, and Dacheng Tao. *"Learning a Mixture of Granularity-Specific Experts for Fine-Grained Categorization."* In Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019.

2. **Lianbo Zhang**, Shaoli Huang, and Wei Liu. *"Intra-Class Part Swapping for Fine-Grained Image Classification."* Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021.

3. **Lianbo Zhang**, Shaoli Huang, and Wei Liu. *"Enhancing Mixture-of-Experts by Leveraging Attention for Fine-Grained Recognition."* IEEE Transactions on Multimedia, 2021.

4. **Lianbo Zhang**, Shaoli Huang, Xinchao Wang, Wei Liu, and Dacheng Tao *"Structure Aware Feature Generation for Zero-Shot Learning."* arXiv preprint arXiv:2108.07032, 2021.

Acknowledgements

Throughout the doctoral study in University of Technology Sydney, I have received a great deal of support and assistance.

I would first like to thank my supervisors, A/Prof. Wei Liu and Prof. Dacheng Tao, for their invaluable advice, continuous support, and patience during my PhD study. Their immense knowledge and plentiful experience have encouraged me in all the time of my academic research and daily life. Their expertise was invaluable in formulating the research questions and methodology. Their insightful feedback pushed me to sharpen my thinking and brought my work to a higher level.

Besides my supervisors, I would like to offer special thanks to my advisor, Dr. Shaoli Huang, for his technical support in my study. I am grateful for his great patience in answering my questions, as well as improving my academic writing. He taught me how to find interesting ideas, develop solid algorithms, and write academic papers.

In addition, I am deeply grateful to my parents, who gave me unreserved love and support. They are always there for me hearing my experience of study and life, sharing my joys and sorrows. Finally, I could not have completed this thesis without the support of my friends, Shanshan Zhao, Zeyu Feng, Yu Cao, Xinyuan Chen, Chaoyue Wang, Liu Liu, Qi Zheng, Youjian Zhang, Yaxin Shi and Ying Li, who provides stimulating discussion as well as distractions to rest my mind outside of my research.

Abstract

Fine-grained recognition is challenging in computer vision and artificial intelligence. It aims to identify under subcategories of given images but suffers from small inter-class variance and large intra-class variance along with multiple object scales and complex background, leading to a more complex problem space. Recently, deep neural networks have extensively promoted the development of fine-grained recognition. However, the existing methods still suffer from several issues, including data limitation, model interpretation, and performance. In this thesis, we propose several data-transformation models to address these challenges.

First, we develop a unified framework (MGN-CNN) based on a mixture of experts to promote diversity among experts by combing a gradually-enhanced learning strategy and a KullbackLeibler divergence based constraint. The strategy learns new experts on the dataset with prior knowledge from former experts and adds them to the model sequentially. At the same time, the introduced constraint forces the experts to produce diverse prediction distributions. These drive the experts to learn the task from different aspects, making them specialized in various subspace problems.

Second, we propose Intra-class Part Swapping (InPS) that produces new data by performing attention-guided content swapping on input pairs from the same class. Compared with previous approaches, InPS avoids introducing noisy labels and ensures a likely holistic structure of objects in generated images. We demonstrate InPS outperforms the most recent augmentation approaches in both fine-grained recognition and weakly object localization.

Finally, we explore fine-grained zero-shot learning and introduce a novel structure-aware feature generation scheme, termed SA-GAN, to explicitly account for the topological structure in learning both the latent space and the generative networks. This topology-preserving mechanism enables our method to significantly enhance the generalization capability on unseen-classes and consequently improve the classification performance.

Contents

List of Figures

List of Tables