Faculty of Engineering and Information Technology
University of Technology Sydney

Towards Automatic Construction of Diverse, High-quality Image Dataset

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

by

Yazhou Yao

September 2018

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.

Signature of Candidate

Production Note:
Signature removed prior to publication.

Acknowledgments

Foremost, I would like to express my sincere gratitude to my supervisors Prof. Jian Zhang for the continuous support of my Ph.D study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis.

I also would like to appreciate my co-supervisor Prof. Zhenming Tang and Fumin Shen for providing me with continuous support throughout my Ph.D study and research. Without their professional guidance and persistent help, this thesis would not have been possible.

I am very grateful to Senior Researcher Xian-sheng Hua at Alibaba Research, and I have benefited greatly from every high-quality group meeting. Their cutting-edge research perspectives and solid theoretical foundation make me open-minded and inspire me to work harder.

I thank my fellow labmates in Global Big Data Technologies Center: Yucheng Wang, shangrong Huang for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last three years.

Last but not the least, I would like to thank my family: my wife and my parents, for their unconditional support, both financially and emotionally throughout the whole Ph.D studying.

Yazhou Yao
March 2018 @ UTS

Contents

List of Figures

List of Tables

List of Publications

Papers Published

- **Yazhou Yao**, Fumin Shen, Jian Zhang, Li Liu, Zhenmin Tang, and Ling Shao (2018), Extracting Privileged Information for Enhancing Classifier Learning, *in* 'Proceedings of IEEE Transactions on Image Processing (**TIP**)', doi=10.1109/TIP.2018.2869721.

- **Yazhou Yao**, Fumin Shen, Jian Zhang, Li Liu, Zhenmin Tang, and Ling Shao (2018), Discovering and Distinguishing Multiple Visual Senses for Web Learning, *in* 'Proceedings of IEEE Transactions on Multimedia (**TMM**)', doi=10.1109/TMM.2018.2847248.

- **Yazhou Yao**, Jian Zhang, Fumin Shen, Xian-Sheng Hua, Wankou Yang and Zhenmin Tang (2018), Extracting Privileged Information from Untagged Corpora for Classifier Learning, *in* 'Proceedings of the International Joint Conference on Artificial Intelligence (**IJCAI 18**)', pp, 1085-1091.

- **Yazhou Yao**, Jian Zhang, Fumin Shen, Wankou Yang, Pu Huang and Zhenmin Tang (2018), Discovering and Distinguishing Multiple Visual Senses for Polysemous Words. *in* 'Proceedings of the AAAI Conference on Artificial Intelligence (**AAAI 18**)', pp, 523-530.

- **Yazhou Yao**, Jian Zhang, Fumin Shen, Xian-Sheng Hua, Jingsong Xu and Zhenmin Tang (2017), Exploiting Web Images for Dataset

Construction: A Domain Robust Approach. *in* 'Proceedings of IEEE Transactions on Multimedia (**TMM**)', pp. 1771-1784.

- **Yazhou Yao**, Xiansheng Hua, Fumin Shen, Jian Zhang and Zhenmin Tang (2016), A Domain Robust Approach for Image Dataset Construction. *in* 'Proceedings of the ACM International Conference on Multimedia (**ACM MM 16**)', pp, 212-216.

- **Yazhou Yao**, Jian Zhang, Fumin Shen, Xian-Sheng Hua, Jingsong Xu and Zhenmin Tang (2016), Automatic Image Dataset Construction with Multiple Textual Metadata. *in* 'Proceedings of the IEEE International Conference on Multimedia and Expo (**ICME 16**)', pp, 1-6.

- **Yazhou Yao**, Jian Zhang, Fumin Shen, Xian-Sheng Hua, Jingsong Xu and Zhenmin Tang (2017), A New Web-supervised Method for Image Dataset Construction. *in* 'Proceedings of Neurocomputing', pp. 23-31.

- **Yazhou Yao**, Wankou Yang, Pu Huang, Qiong Wang, Yunfei Cai and Zhenmin Tang (2018), Exploiting Textual and Visual Features for Image Categorization. *in* 'Proceedings of Pattern Recognition Letters', doi=https://doi.org/10.1016/j.patrec.2018.05.028.

- **Yazhou Yao**, Jian Zhang, Xian-Sheng Hua, Fumin Shen and Zhenmin Tang (2016), Extracting Visual Knowledge from the Internet: Making Sense of Image Data. *in* 'Proceedings of the International Conference on Multimedia Modelling', pp, 862-873.

Papers Under Review

- **Yazhou Yao**, Jian Zhang, Fumin Shen, Dongxiang Zhang, Zhenmin Tang, and Heng Tao Shen (2017), Towards Automatic Construction of Diverse, High-quality Image Dataset, IEEE Transactions on Knowledge and Data Engineering (**TKDE**) (**major revision**).

Abstract

The availability of labeled image datasets has been shown critical for high-level image understanding, which continuously drives the progress of feature designing and models developing. However, the process of manual labeling is both time-consuming and labor-intensive. To reduce the cost of manual annotation, there has been increased research interest in automatically constructing image datasets by exploiting web images. Datasets constructed by existing methods tend to suffer from the disadvantage of low accuracy and low diversity. These datasets tend to have a weak domain adaptation ability, which is known as the "dataset bias problem".

This research aims at automatically collect accurate and diverse images for given queries from the Web, and construct a domain robust image dataset. Thus, within this thesis, various methods are developed and presented to address the following research challenges. The first is the retrieved web images are usually noisy, how to remove noise and construct a relatively high accuracy dataset. The second is the collected web images are often associated with low diversity, how to address the dataset bias problem and construct a domain robust dataset.

In Chapter 3, a framework is presented to address the problem of polysemy in the process of constructing a high accuracy dataset. Visual polysemy means that a word has several semantic (text) senses that are visually (image) distinct. Solving polysemy can help to choose appropriate visual senses for sense-specific images collection, thereby improving the accuracy of the collected images. Unlike previous methods which leveraged the human-

developed knowledge such as Wikipedia or dictionaries to handle polysemy, we propose to automate the process of discovering and distinguishing multiple visual senses from untagged corpora to solve the problem of polysemy.

In Chapter 4, a domain robust framework is presented for image dataset construction. To address the dataset bias problem, our framework mainly consists of three stages. Specifically, we first obtain the candidate query expansions by searching in the Google Books Ngram Corpus. Then, by treating word-word (semantic) and visual-visual distance (visual) as features from two different views, we formulate noisy query expansions pruning as a multi-view learning problem. Finally, by treating each selected query expansion as a "bag" and the images therein as "instances", we formulate image selection and noise removal as a multi-instance learning problem. In this way, images from different distributions can be kept while noise is filtered out.

Chapter 5 details a method for noisy images removing and accurate images selecting. The accuracy of selected images is limited by two issues: the noisy query expansions which are not filtered out and the error index of image search engine. To deal with the noisy query expansions, we divide them into two types and propose to remove noise from visual consistency and relevancy respectively. To handle noise induced by error index, we classify the noisy images into three categories and filter out noise by different mechanisms separately.

Chapter 6 proposes an approach for enhancing classifier learning by using the collected web images. Different from previous works, our approach, while improving the accuracy and robustness of the classifier, greatly reduces the time and labor dependence. Specifically, we proposed a new instance-level MIL model to select a subset of training images from each selected privileged information and simultaneously learn the optimal classifiers based on the selected images.

Chapter 7 concludes the thesis and outlines the scope of future work.