Exploring deep learning for multimodal understanding
Lao, M.

Acknowledgements

First and foremost, I'd like to extend my deepest appreciation to my promoters, Prof. Dr. Michael Lew and Prof. Dr. Aske Plaat. Michael, I'm profoundly grateful for offering me a chance to pursue my Ph.D in the field of multimodal retrieval. Your guidance during the pivotal moments of my academic journey and your constant push for my growth as an independent and assertive researcher were invaluable. Your unwavering support and insightful advice have been instrumental in my Ph.D. research. It's truly a privilege to have been part of the MediaLab and to have had the opportunity to work under your guidance throughout my doctoral journey. Aske, your insightful feedback consistently encouraged me to refine my thought processes and elevate the quality of my work. I'm deeply thankful for your mentorship and generous support during my Ph.D. studies.

My dissertation would not have been possible without the guidance of Dr. Erwin Bakker. Erwin, your constant inspiration for my presentations and astute feedback on my papers have been invaluable to my academic journey.

I would like to express my appreciation to my doctorate committee members: Prof. Dr. Katherine Wolstencroft, Prof. Dr. Boudewijn Lelieveldt, Prof. Dr. Cees G. M. Snoek and Prof. Dr. K.Joost Batenburg. They reviewed my thesis carefully and gave me insightful comments and suggestions.

Then, I would like to thank Prof Dr. Yanming Guo. Without his academic guidance, I cannot grow up quickly in my Ph.D. journey. Next, I would like to thank Dr. Yu Liu and Wei Chen. I really appreciate the assistance and advice I obtained from him. My big thank goes to Dr. Nan Pu. I appreciate all the discussions and collaborations with him. In the long PhD journey, I have experienced a lot of frustrated and lost moments in research. No matter when I ware confused or fell in trouble, he always tried his best to help me. I would like to express my gratitude to all my colleagues in Computational Imaging and Deep Learning Research Cluster at LIACS. Thank you for sharing your knowledge and good suggestions about my research. I enjoyed and benefited from the communications with you.

I would like to express my deepest appreciation to my beloved grandparents and parents. They convinced me, "god help those who help themselves. heaven blesses the diligent heaven rewards the diligent." You may not understand what my research work is in this thesis, but it is the best feedback to your unconditional love and support. I hope they are happy and healthy forever. My special appreciation goes to my girlfriend Jinting Bian. Thanks for your ever-present encouragement, company, and always supporting me spiritually. To meet you, for me, is the greatest happiness.

Last but not least, I feel lucky to have made so many Chinese friends in Leiden and in the Netherlands. The days with them made my doctoral studies more enjoyable and less stressful.

<div align="right">

Mingrui Lao
August 2023
Leiden, the Netherlands

</div>

Curriculum Vitae

Mingrui Lao was born in Changsha, China on February 9, 1995. In 2013, he started his bachelor's study at Xi'an Jiaotong University in Xi'an, Shaanxi, China, and received his bachelor's degree in 2017. After that, he started his master's study at National University of Defence Technology in Changsha, China, and obtained his master's degree in 2017 under the supervision of Prof. Dr. Weidong Xiao.

In October 2019, he started his PhD research supported by the China Scholarship Council (CSC No. 201907720082) and worked at the MediaLab in the Leiden Institute of Advanced Computer Science (LIACS), Leiden University, the Netherlands, under the supervision of Prof. Dr. Michael Lew and Prof. Dr. Aske Plaat. Mingrui Lao's research interests include vision-language learning, computer vision and deep learning. He has published papers in international journals and conferences, including IEEE Transactions on Multimedia, Neurocomputing, AAAI, ACM MM, and ICASSP *etc.* He also serves as a reviewer for some conferences and journals, such as IEEE Transactions on Image Processing, Artificial Intelligence Review, IEEE Transactions on Circuits and Systems for Video Technology.

Publication List

Main publications for the thesis

- **Lao, M.**, Pu, N., Zhong, Z., Sebe, N., and Lew, M. S. "FedVQA: Personalized Federated Visual Question Answering over Heterogeneous Scenes" ACM International Conference on Multimedia, 2023.

- **Lao, M.**, Pu, N., Liu, Y., Zhong, Z., Bakker, E. M., Sebe, N., and Lew, M. S. "Multi-Domain Lifelong Visual Question Answering via Self-Critical Distillation" ACM International Conference on Multimedia, 2023.

- **Lao, M.**, Pu, N., and Lew, M. S. "Robust Visual Question Answering: Challenges, Benchmarks and Strategies." Submitted to ACM Transactions on Information Systems, 2023.

- **Lao, M.**, Pu, N., Liu, Y., He, K., Bakker, E. M., and Lew, M. S. "COCA: COllaborative CAsual Regularization for Audio-Visual Question Answering." The AAAI conference on artificial intelligence, 2023.

- **Lao, M.**, Guo, Y., Liu, Y., Chen, W., Pu, N., and Lew, M. S. "From Superficial to Deep: Language Bias driven Curriculum Learning for Visual Question Answering." ACM International Conference on Multimedia, 2021.

- **Lao, M.**, Guo, Y., Liu, Y. and Lew, M. S. "A Language Prior based Focal Loss for Visual Question Answering." IEEE International Conference on Multimedia and Expo, 2021.

- **Lao, M.**, Guo, Y., Pu, N., Chen, W., Liu, Y., and Lew, M. S. "Multi-stage hybrid embedding fusion network for visual question answering." Neurocomputing, 2021, 423, pp. 541-550.

Other publications

- **Lao, M.**, Guo, Y., Chen, W., Pu, N., Lew, M. S., "VQA-BC: Robust Visual Question Answering via Bidirectional Chaining.", International Conference on Acoustics, Speech and Signal Processing, 2022.

- He, K., Pu, N., **Lao, M.**, Bakker, E. M., and Lew, M. S. "Dual Selective Knowledge Transfer for Few-Shot Classification." Applied Intelligence, 2023.

- He, K., Pu, N., **Lao, M.**, and Lew, M. S. "Few-shot and meta-learning methods for image understanding: a survey." International Journal of Multimedia

Information Retrieval, 2023.

- Chen, W., Xu, H., Pu, N., Liu, Y., **Lao, M.**, Wang, W., Liu, L., and Lew, M. S., "Lifelong Fine-grained Image Retrieval." IEEE Transactions on Multimedia, 2022.