## Yuanhang Zhang (张远航)

Contact Information	<ul><li>IIP Lab, Institute of Computing Technology</li><li>Chinese Academy of Sciences</li><li>Building 1, 156 Beiqing Rd</li><li>Haidian, Beijing 100095</li><li>P. R. China</li></ul>		
	<pre>Homepage: www.sailorzhang.com Google Scholar: https://scholar.google.com/citations?user=cTZpqrUAAAAJ E-mail: zhangyuanhang15@mails.ucas.ac.cn (permanent, preferred) or yuanhang.zhang@vipl.ict.ac.cn</pre>		
Education	Institute of Computing Technology, Chinese Academy of Sciences (CAS), Beijing, China		
	Ph.D. in Computer Science and Technology	09/2020 to present	
	<ul> <li>Supervisors: Prof. Shiguang Shan and Prof. Shuang Yang</li> <li>Research focus: Talking face analysis (e.g. lip reading, active speaker detection) and multi-modal self-supervised learning</li> <li>Two-time winner at AVA Active Speaker Detection track of the ActivityNet Challenge 2021&amp; 2022, in collaboration with Susan Liang (now at University of Rochester; Google annoucement &amp; project website)</li> </ul>		
	University of Oxford, Oxford, UK		
	Visiting student in Computer Science	10/2018 to $03/2019$	
	University of Chinese Academy of Sciences (UCAS), Beijing, China		
	<ul> <li>B.Eng. in Computer Science and Technology</li> <li>GPA: 3.75/4.0</li> <li>Supervisor: Prof. Shiguang Shan</li> <li>Bachelor thesis: Key Technologies for Large-scale W (Outstanding Bachelor Thesis award)</li> </ul>	09/2015 to 07/2019 Visual Speech Recognition	
Academic Experience	<ul> <li>Institute of Computing Technology, CAS, Beijing, C Research Assistant in Computer Vision</li> <li>Automated Lip Reading of Chinese in the Wild</li> <li>Supervisor: Prof. Shiguang Shan and Dr. Shuang Yang</li> <li>Investigated methods in shot boundary detection, vis modal learning for speaker identification and speech re</li> </ul>	hina 05/2017 to 06/2019 sual tracking, and multi- cognition	
	• <b>Primary contributor</b> to the construction of the first large-scale Chinese lip- reading database, <i>CAS-VSR-W1k</i> (formerly known as LRW-1000), which covers large variations in pose, age and other speaker attributes, using broadcast news videos spanning over more than a year		
	• Helped prepare the 1st Mandarin Audio-Visual Speech Recognition Challenge (MAVSR) at ACM ICMI 2019		
	• <b>Runner-up</b> at AVA Active Speaker Detection track of the ActivityNet Challenge 2019, in collaboration with Jingyun Xiao (Google annoucement)		

	<ul> <li>Department of Foreign Languages, UCAS, Beijing, China         <ul> <li>Research Student</li> <li>12/2016 to 03/20</li> </ul> </li> <li>Developing Coherence by Machine Readable Items In EAP Speaking Activities</li> <li>Supervisor: Prof. Qun Zheng         <ul> <li>Constructed an intepretable scoring mechanism for the speaking section of Interne based TOEFL (iBT) tests</li> </ul> </li> </ul>		
	• Accomplished by ranking the importance of different guistics with various machine learning techniques, suc	coherence metrics in lin- h as LASSO regression	
Additional Experience	<ul> <li>SeetaTech Technology Limited, Beijing, China</li> <li>R&amp;D Intern</li> <li>Integrated lip reading models into a fully functional de internal Mandarin dataset to recognise speech comma</li> </ul>	echnology Limited, Beijing, China 04/2018 to 07/2018 d lip reading models into a fully functional demo system, trained on an Mandarin dataset to recognise speech commands from drivers	
Awards	CAS Undergraduate Student Scholarship UCAS Student Scholarship Merit Student Award, UCAS ICT Master Student Scholarship (sponsored by E Fund Management Co., Ltd.) National Scholarship	11/18, 12/17 16-18, 21-23 06/22, 12/18, 11/17, 11/17 12/21 12/22	
Publications	<b>Y. Zhang</b> , S. Yang, S. Shan and X. Chen, $ES^3$ : Evolving Self-Supervised Learning of Robust Audio-Visual Speech Representations, IEEE/CVF CVPR 2024. First author. [link]		
	<b>Y. Zhang</b> , S. Liang, S. Yang and S. Shan, UNICON+: <i>ICTCAS-UCAS Submission to the AVA-ActiveSpeaker Task at ActivityNet Challenge 2022</i> , technical report for AVA Challenge 2022. Co-first author. [link]		
	<b>Y. Zhang</b> , S. Liang, S. Yang, X. Liu, Z. Wu, S. Shan and X. Chen, <i>UniCon: Unified Context Network for Robust Active Speaker Detection</i> , ACM MM 2021 (oral). Co-first author. [link]		
	<b>Y. Zhang</b> , S. Liang, S. Yang, X. Liu, Z. Wu and S. Shan, <i>ICTCAS-UCAS-TAL Submission to the AVA-ActiveSpeaker Task at ActivityNet Challenge 2021</i> , technical report for AVA Challenge 2021. Co-first author. [link]		
	<b>Y. Zhang</b> , S. Yang, J. Xiao, S. Shan and X. Chen, <i>Can We Read Speech Beyond Lips?</i> <i>Rethinking RoI Selection for Deep Visual Speech Recognition</i> , IEEE F&G 2020 (oral). First author. [link]		
	J. Xiao, S. Yang, Y. Zhang, S. Shan and X. Chen, <i>Deformation Flow Based Two-stream Network for Lip Reading</i> , IEEE F&G 2020 (poster). Third author. [link]		
	<b>Y. Zhang</b> , R. Huang, J. Zeng and S. Shan, $M^3F$ : Multi-Modal Continuous Valence- Arousal Estimation in the Wild, IEEE F&G 2020 (workshop): ABAW Challenge 2020. First author. [link]		
	<b>Y. Zhang</b> , J. Xiao, S. Yang and S. Shan, <i>Multi-Task Learning for Audio-Visual Active Speaker Detection</i> , technical report for AVA Challenge 2019. Co-first author. [link]		
	S. Yang, Y. Zhang, D. Feng, M. Yang, C. Wang, J. Xiao, K. Long, S. Shan and X. Chen, <i>LRW-1000: A Naturally-Distributed Large-Scale Benchmark for Lip Reading in</i>		

## the Wild, IEEE F&G 2019 (oral). Co-first author. [link]

Skills & Miscellaneous

- Programming languages: Python, C++, C, Bash
- Machine learning frameworks: PyTorch, TensorFlow, JAX
- Proofread two chapters of the Chinese translation of Bengio et al.'s textbook, "Deep Learning" (Posts and Telecom Press, Beijing, 2017; ISBN 9787115461476).
- Translated part of the textbook "Applied Machine Learning" (David Forsyth, China Machine Press, Beijing, 2021; ISBN 9787111668299).