YE ZHU

Homepage
Google Scholar
yzhu96@hawk.iit.edu

RESEARCH INTERESTS

Computer Vision; Multimodal Learning and Generation; Vision, Audio and Language

ACADEMIC EXPERIENCE

Illinois Institute of Technology, USA January 2021 - Present
Ph.D. candidate in Computer Science, advised by Prof. Yan Yan.
GPA: 4.0/4.0

Princeton University, USA September 2022 - Present
Visiting Researcher at VisualAI lab, advised by Prof. Olga Russakovsky.

Texas State University, USA *(Transferred)* September 2019 - December 2020
Ph.D. in Computer Science, advised by Prof. Yan Yan.
GPA: 4.0/4.0

Shanghai Jiao Tong University(SJTU), China September 2016 - March 2019
M.S. in Mechanical Engineering.
GPA: 3.72/4.0

Ecole Polytechnique(X), France September 2016 - March 2017
Exchange Master Student in Engineering.

Shanghai Jiao Tong University(SJTU), China September 2012 - August 2016
B.S. in Mechanical and Automation.
GPA: 3.63/4.3

SELECTED PUBLICATIONS AND PREPRINTS

Preprints

[1] <u>Ye Zhu</u>, Yu Wu, Nicu Sebe, and Yan Yan. Vision+X: A Survey on Multimodal Learning in the Light of Data, (arXiv preprint: arXiv:2210.02884), 2022. [*Paper*]

[2] <u>Ye Zhu</u>, Xiaoguang Zhu, Yuzhang Shang, Zhenhao Zhao, and Yan Yan. Supplementing Missing Visions via Dialog for Scene Graph Generations. (arXiv preprint: arXiv:2204.11143), 2022. [*Paper*] [*Code*]

Conference and Journal Publications, 2020 - Now

[1] <u>Ye Zhu</u>, Yu Wu, Kyle Olszewski, Jian Ren, Sergey Tulyakov, and Yan Yan. Discrete Contrastive Diffusion for Cross-Modal Music and Image Generation, in International Conference on Learning Representations (**ICLR**), 2023. [*Paper*] [*Code*]

[2] <u>Ye Zhu</u>, Kyle Olszewski, Yu Wu, Panos Achlioptas, Menglei Chai, Yan Yan, and Sergey Tulyakov. Quantized GAN for Complex Music Generation from Dance Videos, in European Conference on Computer Vision (**ECCV**), 2022. [*Paper*] [*Code*] [*Project*]

[3] <u>Ye Zhu</u>, Yu Wu, Yi Yang, and Yan Yan. Saying the Unseen: Video Descriptions via Dialog Agents, in IEEE Transactions on Pattern Analysis and Machine Intelligence (**TPAMI**), 2022. [*Paper*] [*Code*]

[4] <u>Ye Zhu</u>, Yu Wu, Hugo Latapie, Yi Yang, Yan Yan. Learning Audio-Visual Correlations From Variational Cross-Modal Generations, in IEEE International Conference on Acoustics, Speech and Signal Processing (**ICASSP**), 2021. [*Paper*] [*Code*]

[5] Xiaoguang Zhu, <u>Ye Zhu</u>, Haoyu Wang, Honglin Wen, Yan Yan, Peilin Liu. Skeleton Sequence and RGB Frame Based Multi-Modality Feature Fusion Network for Action Recognition, in Transactions on Multimedia Computing Communications and Applications (**TOMM**), 2021. [*Paper*]

[6] <u>Ye Zhu</u>, Yu Wu, Yi Yang, and Yan Yan. Describing Unseen Videos via Multi-Modal Cooperative Dialog Agents, in European Conference on Computer Vision (**ECCV**), 2020. [*Paper*] [*Code*]

[7] <u>Ye Zhu</u>, Yan Yan, and Oleg Komogortsev. Hierarchical HMM for Eye Movement Classification, in European Conference on Computer Vision Workshop (**ECCVW**), 2020. [*Paper*]

INDUSTRIAL EXPERIENCE

Snap Inc., Remote, USA   May 2021 - August 2021
Research intern in Computer Vision, advised by Dr. Kyle Olszewski

· Project: Music generation conditioned on dance videos.

Bang & Olufsen, Struer, Denmark   July 2018 - December 2018
Research intern in Computer Vision, advised by Dr. Sven Ewan Shepstone and Dr. Pablo Martinez-Nuevo

· Project: 3D indoor scene understanding via point clouds.

OUTREACH AND INVITED TALKS

AI on the Big Stage, Bang & Olufsen, Denmark *(Remote)*   November 2022
- Guest speaker for Bang & Olufsen AI club
- Multimodal Learning and Generation for Audio and Music

SELECTED HONORS AND AWARDS

Excellence in Dissertation Research for the College of Computing, IIT	2022
CVPR Travel Grant Award	2022
Merrick Merit Fellowship, Texas State University	2019
First Class Academic Excellence Scholarship for Graduate Students of SJTU	2017&2018
Meritorious Winne in Mathematical Contest in Modeling (MCM)	2015
Second Class Academic Excellence Scholarship for Undergraduate Students of SJTU	2015

PROFESSIONAL ACTIVITIES

Conference Reviewer: CVPR, ECCV, AAAI, ACMM, WACV, ICASSP.	
Journal Reviewer: IEEE Transactions on Multimedia, Neurocomputimg, Knowledge-Based Systems.	
Research Assistant at the Computer Vision and Multimedia Lab, Illinois Institute of Technology	2021-Now
Teaching Assistant for CS4310 Computer Networks, Texas State University	2019&2020
Teaching Assistant for CS4328 Operating Systems, Texas State University	2020

LINGUISTIC SKILLS AND OTHERS

Chinese (Native Proficiency)
English (Professional Proficiency)
French (Professional Proficiency, DALF & TCF C1 Diploma)
Part-time French translator for the European Science Magazine *Science&Vie*