

YE ZHU (SHE/HER)

[Homepage](#)

[Google Scholar](#)

yezhu@princeton.com

RESEARCH INTERESTS

Main Research: Multimodal Generation (Vision, Audio and Language) , Computer Vision

Interdisciplinary: Machine Learning for Astrophysics

PROFESSIONAL APPOINTMENT

Princeton University, USA

September 2023 - Now

Postdoctoral Research Associate in Computer Science.

Advisor: Prof. Olga Russakovsky.

EDUCATION

Illinois Institute of Technology, USA

September 2023

Ph.D. in Computer Science.

Thesis: Multimodal Learning and Generation - Toward a Multisensory and Creative AI System.

Advisor: Prof. Yan Yan.

Princeton University, USA

September 2022 - June 2023

Visiting Ph.D. in Computer Science.

Advisor: Prof. Olga Russakovsky.

Shanghai Jiao Tong University (SJTU), China

March 2019

M.S. in Mechanical Engineering.

French Engineering Diploma.

Ecole Polytechnique (X), France

September 2016 - March 2017

Exchange Master Student in Engineering.

Shanghai Jiao Tong University (SJTU), China

August 2016

B.S. in Mechanical and Automation.

Pre-enrollment before Chinese College Entrance Examination.

FEATURED PUBLICATIONS

(* for equal contributions, representative works marked in shades.)

Preprints

[1] **Ye Zhu**, Yu Wu, Zhiwei Deng, Olga Russakovsky, and Yan Yan. Unseen Image Synthesis with Diffusion Models. (arXiv preprint, arXiv:2310.09213), 2023. [[Paper](#)]

[2] **Ye Zhu**, 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. [[Survey Paper](#)]

[3] Ruoyu Wang*, Yongqi Yang*, Zhihao Qian, **Ye Zhu**, Yu Wu. Diffusion in Diffusion: Cyclic One-Way Diffusion for Text-Vision-Conditioned Generation. (arxiv preprint, arXiv:2306.08247), 2023. [[Paper](#)] [[Project](#)]

Computer Science Conference and Journal Publications, 2020 - Now

- [1] Bin Duan, Hao Tang, Changchang Sun, **Ye Zhu**, Yan Yan. Mining and Unifying Heterogeneous Contrastive Relations for Weakly-Supervised Actor-Action Segmentation, in *Winter Conference on Applications of Computer Vision (WACV)*, 2024.
- [2] **Ye Zhu**, Yu Wu, Zhiwei Deng, Olga Russakovsky, and Yan Yan. Boundary Guided Learning-Free Semantic Control with Diffusion Models, in *Conference on Neural Information Processing Systems (NeurIPS)*, 2023. [\[Paper\]](#) [\[Code\]](#) [\[Project\]](#) [\[Hugging face\]](#)
- [3] **Ye Zhu**, 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. [\[Paper\]](#) [\[Code\]](#) [\[Project\]](#)
- [4] Matthew Coleman, Olga Russakovsky, Christine Allen-Blanchette, and **Ye Zhu**. Discrete Diffusion Reward Guidance Methods for Offline Reinforcement Learning, in *International Conference on Machine Learning, Sampling and Optimization in Discrete Space (SODS) Workshop (ICML Workshop)*, 2023. [\[Paper\]](#)
- [5] Duo Xu, Jonathan Ta, Chia-Jung Hsu, and **Ye Zhu**. Denoising Diffusion Probabilistic Models to Predict the Number Density of Molecular Clouds in Astronomy, in *International Conference on Learning Representations Physics4ML Workshop (ICLR Workshop)*, 2023. [\[Paper\]](#)
- [6] **Ye Zhu**, 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. [\[Paper\]](#) [\[Code\]](#) [\[Project\]](#)
- [7] **Ye Zhu**, 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)*, 2021. [\[Paper\]](#) [\[Code\]](#)
- [8] **Ye Zhu**, 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. [\[Paper\]](#) [\[Code\]](#)
- [9] Xiaoguang Zhu, **Ye Zhu**, Haoyu Wang, Honglin Wen, Yan Yan, Peilin Liu. Skeleton Sequence and RGB Frame Based Multi-Modality Feature Fusion Network for Action Recognition, in *ACM Transactions on Multimedia Computing Communications and Applications (TOMM)*, 2021. [\[Paper\]](#)
- [10] **Ye Zhu**, Yu Wu, Yi Yang, and Yan Yan. Describing Unseen Videos via Multi-Modal Cooperative Dialog Agents, in *European Conference on Computer Vision (ECCV)*, 2020. [\[Paper\]](#) [\[Code\]](#)
- [11] **Ye Zhu**, Yan Yan, and Oleg Komogortsev. Hierarchical HMM for Eye Movement Classification, in *European Conference on Computer Vision Workshop (ECCV Workshop)*, 2020. [\[Paper\]](#)

Astrophysics Journal Publications, 2023 - Now

- [1] Duo Xu, Jonathan Tan, Chia-Jung Hsu, and **Ye Zhu**. Denoising Diffusion Probabilistic Models to Predict the Density of Molecular Clouds, in *The Astrophysics Journal (APJ)*, 2023. [\[Paper\]](#)

INDUSTRIAL INTERNSHIP

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.

TALKS

Invited speaker at AI forum, Bang & Olufsen, Copenhagen, Denmark - Topics on Multimodal Learning and Generation.	October, 2023
Joint talk with Prof. Olga Russakovsky at ICML Workshop, Hawaii, USA - Topics on Art, Science and Challenges of Generative AI.	July 2023
Invited speaker at Wuhan University, Wuhan, China - Topics on diffusion generative models, ML4Astrophysics.	May 2023
Invited speaker at Shanghai Jiao Tong University (SJTU), Shanghai, China - Topic on diffusion generative models, ML4Astrophysics.	May 2023
Guest course lecture, Princeton University, USA - Guest lecture for the COS429 Computer Vision, topic on diffusion generative models.	April 2023
Invited speaker for ZHIDX Tech, China (<i>Remote</i>) - Topic on multimodal generation for music and images, live talk.	April 2023
Guest speaker at PIXL talk, Princeton University, USA - Topic on diffusion generative models, ML4Astrophysics.	April 2023
Guest speaker for the AI club, Bang & Olufsen, Denmark (<i>Remote</i>) - Topic on multimodal learning and generation for audio and music data.	November 2022

OUTREACH AND TEACHING

AI4ALL Program, Princeton University, USA - Instructor for the NLP project. - Mentor for high school students to create an emotionally supportive Chatbot. - Media Coverage on AI4ALL Princeton from NBC	July 2023
--	-----------

FEATURED HONORS AND AWARDS

NeurIPS 2023 Scholar Award, New Orleans, USA	2023
ICCV 2023 DEI Grant, Paris, France	2023
ICLR 2023 Financial Assistance Award, Kigali, Rwanda	2023
ACM-Women Scholarship [Coverage]	2023
Award for Excellence in Dissertation Research for the College of Computing, IIT, USA	2022
CVPR 2022 Travel Grant Award, New Orleans, USA	2022
Merrick Merit Fellowship, Texas State University, USA	2019
First Class Academic Excellence Scholarship for Graduate Students of SJTU, China	2018
First Class Academic Excellence Scholarship for Graduate Students of SJTU, China	2017
Meritorious Winne in Mathematical Contest in Modeling (MCM)	2015
Second Class Academic Excellence Scholarship for Undergraduate Students of SJTU, China	2015

PROFESSIONAL SERVICE

Conference Reviewer <i>In Computer Vision:</i> CVPR 2022-2024, ECCV 2022, ICCV 2023, ACMMM 2021-2022, WACV 2023-2024 <i>In Machine Learning:</i> NeurIPS 2023, ICLR 2024, ICML 2023-2024, AAAI 2023-2024 <i>In Signal Processing:</i> ICASSP 2022
Journal Reviewer IEEE Transactions on Multimedia, Neurocomputing, Knowledge-Based Systems

STUDENT MENTORING

Matthew Coleman, Discrete Diffusion Reward Guidance for Offline RL	Undergrad, Princeton University
Ruoyu Wang, Cyclic Diffusion for Text-Visual Conditioned Generation	Undergrad, Wuhan University
Yongqi Yang, Cyclic Diffusion for Text-Visual Conditioned Generation	MS student, Wuhan University
Sai Wang, DETER dataset for Detecting Generative Manipulations	PhD student, Wuhan University

LINGUISTIC SKILLS AND OTHERS

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