Zhongweiyang Xu

https://xzwy.github.io/alanweiyang.github.io/

Email : zx21@illinois.edu

Mobile : +1-217-693-1582

EDUCATION

- **University of Illinois Urbana-Champaign** — Champaign, IL
  MS/PhD Student in Electrical and Computer Engineering; GPA: 4.00 — *Aug. 2021 – May 2026*

- **University of Illinois Urbana-Champaign** — Champaign, IL
  Bachelor of Science in Computer Engineering; GPA: 3.97 — *Aug. 2018 – May. 2021*

INDUSTRY EXPERIENCE

- **Tencent Ethereal Lab** — Shenzhen, China
  Research Scientist Intern — *May 2022 - Aug 2022*
  - **KiNN**: Develop a Knowledge-inspired **computationally efficient** model for **real-time speech enhancement**. This model is designed for deployment in the **Tencent Meeting**'s (most popular video conferencing app in China) speech AI codec.

- **Tencent AI Lab** — Seattle, US
  Research Scientist Intern — *May 2023 - Aug 2023*
  - **SpatialCodec**: Research on multi-channel speech coding. Propose SpatialCodec to code spatial and spectral information independently with Residual VQ-VAE.
  - **uSee**: Research on text-based speech enhancement with score diffusion.
  - **CodecSE**: Research on codec based generative speech enhancement.

- **Meta Reality Lab Research** — Redmond, US
  Research Scientist Intern — *Aug 2023 - Dec 2023*
  - **FoVNet**: Propose a system for configurable Field-of-View speech enhancement on a smart glass with extremely low computation and distortion.

- **Meta Reality Lab Research** — Redmond, US
  Research Scientist Intern — *May 2025 - Aug 2025*
  - **ArrayDPS++**: Work on diffusion-based unsupervised refinement of predictive multi-channel enhancement model, greatly improving on supervised SOTA in all metrics. (ICASSP2026 in submission)

- **Meta Reality Lab Research** — Redmond, US
  Part-time Student Researcher — *Aug 2025 - Current*
  - **Preference-aligned Speech Enhancement**: Work on diffusion and flow based speech enhancement with preference alignment for MOS and WER.

RESEARCH EXPERIENCE

- **UIUC SiNRG Lab** — Champaign, IL
  Graduate Research Assistant Advised by Prof. Romit Roy Choudhury — *May 2021 - current*
  - **Audio/Speech Inverse Problems**:
    1. Combine signal processing and modern machine learning to solve audio inverse problems like source separation, system identification, speech enhancement, DoA estimation, Beamforming...
  - **Generative AI**:
    1. Apply generative models like VQ-VAE, GAN, Flow, Diffusion, EBM, Autoregressive Language Model on signal of interest.

Teaching Experience

- **UIUC CS/ECE374 SP2021**: Introduction to Algorithms & Models of Computation
- **UIUC CS/ECE434 SP2024**: Real-World Algorithms for IoT and Data Science
- **UIUC CS/ECE434 SP2025**: Real-World Algorithms for IoT and Data Science
- **UIUC ECE598 Fall2025**: Deep Generative Model   (Head TA)
    1. Responsible for class material design: EM, VAE, Latent Variable Models, EBM, Normalizing Flow, Diffusion, Flow Matching. (LLM not covered)
    2. Responsible for written and HW design for diffusion and VAE.

Publications

- **WASPAA2025**: Yulun Wu, *Zhongweiyang Xu, Jianchong Chen, Zhong-Qiu Wang, Romit Roy Choudhury*, "Unsupervised Multi-channel Speech Dereverberation via Diffusion"

- **ICML2025**: *Zhongweiyang Xu, Xulin Fan, Zhong-Qiu Wang, Xilin Jiang, Romit Roy Choudhury*, "ArrayDPS: Unsupervised Blind Speech Separation with a Diffusion Prior"

- **NeurIPS2024 Workshop**: *Zhongweiyang Xu, Debottam Dutta, Yu-Lin Wei, Romit Roy Choudhury*, "Multi-Source Music Generation with Latent Diffusion"

- **ISCA INTERSPEECH2024**: *Zhongweiyang Xu, Ali Aroudi, Ke Tan, Ashutosh Pandey, Jung-Suk Lee, Buye Xu, Francesco Nesta*, "FoVNet: Configurable Field-of-View Speech Enhancement with Low Computation and Distortion for Smart Glasses"

- **IEEE ICASSP2024**: *Muqiao Yang, Chunlei Zhang, Yong Xu, Zhongweiyang Xu, Heming Wang, Bhiksha Raj, Dong Yu*, "uSee: Unified Speech Enhancement and Editing with Conditional Diffusion Models"

- **Information Fusion**: *Andong Li, Guochen Yu, Zhongweiyang Xu, Cunhang Fan, Xiaodong Li, Chengshi Zheng*, "TaBE: Decoupling spatial and spectral processing with Taylor's unfolding method in the beamspace domain for multi-channel speech enhancement"

- **IEEE ICASSP2024**: *Zhongweiyang Xu, Yong Xu, Vinay Kothapally, Heming Wang, Muqiao Yang, Dong Yu*, "SpatialCodec: Neural Spatial Speech Coding"

- **SPEAR Workshop**: *Zhongweiyang Xu, Debottam Dutta, Xulin Fan ,Mark Hasegawa-Johnson ,Romit Roy Choudhury*, "Multichannel Speech Enhancement for SPEAR Challenge: A Three Stage Approach"

- **IEEE ICASSP2023**: *Zhongweiyang Xu*, Xulin Fan*, Mark Hasegawa-Johnson*, "Dual-Path Cross-Modal Attention for Better Audio-Visual Speech Extraction"

- **ACM ICML2022**: *Zhongweiyang Xu, Romit Roy Choudhury*, "Learning to Sparate Voices by Spatial Regions"

Preprints

- **arXiv**: *Sattwik Basu, Chaitanya Amballa, Zhongweiyang Xu, Jorge Vančo Sampedro, Srihari Nelakuditi, Romit Roy Choudhury*, "Contrastive Diffusion Guidance for Spatial Inverse Problems"

- **arXiv**: *Debottam Dutta, Chaitanya Amballa, Zhongweiyang Xu, Yu-Lin Wei, Romit Roy Choudhury*, "Learning Energy-based Variational Latent Prior for VAEs"

- **arXiv**: *Heming Wang, Meng Yu, Hao Zhang, Chunlei Zhang, Zhongweiyang Xu, Muqiao Yang, Yixuan Zhang, Dong Yu*, "Unifying Robustness and Fidelity: A Comprehensive Study of Pretrained Generative Methods for Speech Enhancement in Adverse Conditions"

Honors and Awards

- **2018-2021**: UIUC ECE Highest Honor Graduation
- **2024-2025**: A. R. Buck Knight Fellowship

PROJECTS

- **Blind Image Deblurring with Flow Prior**: Use an image-domain flow-matching models as a prior, to solve blind image deblurring. We use the flow posterior sampling technique with 2 different variants: 1) use a Plug-and-play method to estimate the blurring kernel during diffusion 2) train another flow model for the kernel and then apply parallel diffusion sampling for both kernel and image.

- **Conditional VAE for Semi-Supervised Music Separation**: We design a disentangled VAE for music mixtures, where the music mixtures can be encoded into vocal, drums, bass, and other latents. These latents can then be decoded individually to each corresponding stems for separation. This enables training on both labeled data (clean stems) and unlabeled huge datasets, using ELBO traing.

- **AudioMEME Bechmark (ongoing)**: We collect a huge audio-centered meme dataset and question today's multi-modal LLM's ability to understand multi-modal memes.

- **Linux Kernel Project**: Develop an OS Kernel from scatch for ECE391 (about 10k lines of C and X86).

- **BeatDance on FPGA**: Develop a GuitarHero-like game on a Cyclone-V FPGA.

- **RISC-V CPU**: Use SystemVerilog to develop a RISC-V CPU with cache and pipeline.

SKILLS

- **Languages** Python, C, C++, SystemVerilog

- **Coursework and Skills** Machine Learning, Deep Learning, Multimedia Signal Processing, Random Process, Generative AI, Computer Vision, NLP, Optimization, Adaptive Signal Processing, Information Theory, Speech Processing, Advanced DSP, Vector Space Signal Processing, Detection and Estimation Theory, Wireless Sensing, Digital System Design, Operating System.