

BOWEN ZUO

bowen.zuo@email.ucr.edu

EDUCATION

University of California, Riverside Electrical & Computer Engineering, PhD Student	Sept 2024 – Present
Georgia Institute of Technology, Atlanta M.S. in Computer Science Cumulative GPA: 3.86 / Major GPA: 3.91	Aug 2022 – May 2024
University of Washington, Seattle Electrical & Computer Engineering, Embedded Systems Cumulative GPA: 3.76 / Major GPA: 3.80	Sept 2018 – June 2022

RESEARCH EXPERIENCE

Zhu Lab, UC Riverside, Riverside · Conduct research on large language models, currently focusing on test-time scaling across domains.	Sept 2024 – Present
AVA Lab, Georgia Institute of Technology, Atlanta · Led work on talking-face generation and surveyed recent digital-human research and applications. · Investigated audio-to-gesture modeling, with a focus on dyadic conversation settings.	Jan 2023 – May 2024
ACME Lab, University of Washington, Seattle · Worked on High-Level Synthesis for Machine Learning (hls4ml) and built an ML algorithm accelerator.	Oct 2021 – June 2022

PROJECTS

Test-time Scaling with LLMs

- Scaled LLM inference for math reasoning and code generation by developing a dynamic compute-allocation algorithm that replaced uniform inference budgets, which achieved consistent performance gains across evaluated benchmarks. Extended the base method with bandit-based adaptive allocation, improving robustness on skewed, hard-question-dominant distributions; this work became an ICLR 2026 accepted paper.
- Developed an adaptive allocation and sampling-distribution algorithm to address static-sampling limitations in conventional test-time scaling; this work is under review for ACL 2026.

3D Talking Face Generative Models

- Reproduced training of a 3D talking-face model on Meta's Multiface dataset using PyTorch3D.
- Added Multiface facial textures to models that originally produced texture-free outputs.
- Built an interactive pipeline using Google TTS and the ChatGPT API to generate talking-face video responses.
- Worked with multimodal embeddings (identity, audio, images, and text) in encoder-decoder architectures.

Fairer Diffusion Model

- Used Stable Diffusion 1.5 and controlled latent semantics by updating latent-space weights while freezing the rest of the model.
- Updated weights using global attribute distance and an external woman/man classifier with KL-divergence loss.
- Improved race/gender distribution by 10% compared with the baseline.

Learning on Imbalanced Datasets (Long-tailed Recognition)

- Built a CNN-based image classifier for long-tailed, imbalanced CIFAR-60-LT datasets.
- Implemented the dataset class, PyTorch data loader, and a baseline CNN classifier.

- Applied focal loss and data augmentation (including random crops), improving average classification accuracy by over 5%.

TEACHING EXPERIENCE

Teaching Assistant, CS 4644/7643 Deep Learning, Georgia Institute of Technology.

Head Teaching Assistant, CS 228 Introduction to Deep Learning, UC Riverside.

Teaching Assistant, CS 260 Large Models and Advances in AI, UC Riverside.

PUBLICATIONS

B. Zuo, Y. Zhu. “Strategic Scaling of Test-Time Compute: A Bandit Learning Approach.” ICLR 2026.

B. Zuo, D. Zhou, Y. Zhu. “Test-Time Compute Allocation with Self-Adaptive Demonstration Sampling,” Findings ACL 2026.

Coauthor. “Conversational Gestures: Transforming Text into Full-Body Virtual Interactions,” SouthNLP, 2024.

Coauthor. “Machine Learning Evaluation in the Global Event Processor FPGA for the ATLAS Trigger Upgrade.” Journal of Instrumentation (JINST), 2024.