

Ze RONG

Homepage: <https://zerong7777-boop.github.io/> | Email: ZeRong7777@gmail.com

EDUCATION

Nantong University

09/2022-07/2026

- B.Sc. in Computer Science

RESEARCH INTERESTS

Reliable Multimodal Learning, Computer Vision, fMRI Semantic Decoding, and Evidence-Grounded AI Agents

PUBLICATIONS

† Equal Contribution. * Corresponding author.

Published / Accepted

- Zhaoxin Wang, Wenwen Yang, Zhengyu Li, **Ze Rong**, Xing Wang, Jincong Han, Lei Ma*. *A 25-Year Retrospective of the Use of AI for Diagnosing Acute Stroke: A Systematic Review*. Journal of Medical Internet Research (JMIR), 2024.
- Zhaoxin Wang, Wenwen Yang, **Ze Rong**, Zhengyu Li, Xing Wang, Lei Ma*. *Multi-Scale Crossed Algorithm for Ultrasound Medical Image Segmentation Based on MSC-LSAM*. Journal of Data Acquisition and Processing, 2025.
- **Ze Rong**, Ziyue Zhao, Zhaoxin Wang, Lei Ma*. *FaRMamba: Frequency-based Learning and Reconstruction Aided Mamba for Medical Segmentation*. In Proceedings of the 32nd International Conference on Neural Information Processing (ICONIP 2025).
- **Ze Rong**, Xiaofeng Shen, Haoyang Qin, Yue Xu, Hongjun Li, Lei Ma*. *FIRM: Fusion-Injected Residual Memory Brings Token-Level Alignment to Unsupervised VI-ReID*. In Proceedings of the 17th Asian Conference on Machine Learning (ACML 2025).
- Xiaofeng Shen†, **Ze Rong**†, Haoyang Qin, Yue Xu, Binhao Zhao, Lei Ma*. *SCORE: Soccer Curriculum for Ordinal Affect Recognition with Evidence*. International Conference on Multimedia Retrieval (ICMR 2026).

Under Review

- Lei Ma†, **Ze Rong**†, Xiaofeng Shen, Ting Li*. *FuzzyStateFormer: A Neuro-Fuzzy State-Aware Framework with Transformer Temporal Encoding for Reliable Non-Invasive Long-Sequence fMRI Semantic Decoding*. IEEE Transactions on Fuzzy Systems. (Major Revision).
- **Ze Rong**, Xiaofeng Shen, Lei Ma*. *BASS: Brain-Aligned Semantic Scaffold for Temporally Structured fMRI-to-Text Decoding*. 34th ACM International Conference on Multimedia (ACMMM2026).
- Xiaofeng Shen†, **Ze Rong**†, Haoyang Qin, Yue Xu, Li Wang, Lei Ma*. *DART: Multimodal Deviation-Aware Reasoning for Grounded Tactical Explanation in Broadcast Soccer*. 34th ACM International Conference on Multimedia (ACMMM2026).

RESEARCH EXPERIENCE

Project: Cochlear Reconstruction and Segmentation for Small-Animal Biomedical Imaging

07/2024-01/2025

Project Leader

- Collaborated with the Nantong University Institute of Special Medicine on institution-provided mouse cochlear imaging data, leading dataset construction, preprocessing, and task formulation for reconstruction and segmentation workflows.
- Developed NAFA-DIKNet for unsupervised blind super-resolution and denoising of degraded medical images by integrating DIP-based reconstruction, DKP-based degradation estimation, and adaptive gated fusion.
- Designed FaRMamba, a frequency-aware medical segmentation framework with multi-scale frequency modeling and reconstruction-aided learning to improve fine-structure segmentation and boundary recovery.
- Evaluated the segmentation framework on mouse-cochlea, CAMUS, and Kvasir-Seg datasets; FaRMamba reported strong segmentation performance, including 89.81 DSC on CAMUS-2CH, 95.98 DSC on CAMUS-4CH, and 88.97 DSC on Kvasir-Seg.

- Representative outputs: FaRMamba, ICONIP 2025; NAFA-DIKNet patent application.

Project: Multimodal Video Understanding and Reasoning for Soccer Analysis

03/2025-09/2025

Core Contributor

- Built a soccer-oriented multimodal research pipeline under an application background linked to the Nantong University Zhiyun Football Academy, covering video segmentation, task-specific annotation, feature extraction, and multimodal evaluation.
- Re-annotated broadcast soccer clips into structured supervision for segment-level affect prediction, player emotion recognition, tactical risk prediction, and grounded tactical explanation.
- Developed MADE for segment-level team affect prediction by combining pose-anchored team representation, face-pose reliability fusion, and event-biased temporal modeling; manuscript reports 0.49/0.35 F1/QWK on SR-1988 and 0.58/0.39 on NSVA.
- Developed RAVEL for reliability-aware player emotion recognition under noisy pose and appearance cues; manuscript reports 64.2% accuracy and 61.5 Macro-F1 on SoccerNet-GSR, improving Macro-F1 by 11.9 points over naive multimodal fusion.
- Contributed to SCORE and DART for evidence-grounded soccer VLM reasoning and tactical explanation, with SCORE accepted by ICMR 2026 and DART under review at ACM Multimedia 2026.

Project: Cross-Modal Representation Learning

02/2025-04/2025

Project Leader

- Designed FIRM, an unsupervised visible-infrared person re-identification framework addressing cross-modal semantic misalignment, noisy pseudo labels, and unstable memory prototypes.
- Implemented Vision-Semantic Prompt Fusion to inject language-derived semantic prompts into multi-depth visual features, improving token-level alignment between visible and infrared modalities.
- Built Evolving Multi-view Cluster Memory with transport-guided clustering and merge-split prototype updates to stabilize pseudo-label learning under cross-modal distribution shift.
- Evaluated FIRM on SYSU-MM01 and RegDB; the method reported 66.5 Rank-1 / 62.1 mAP on SYSU-MM01 and 90.1 Rank-1 / 84.2 mAP on RegDB Visible-to-Thermal under the unsupervised setting.
- Representative output: FIRM, ACML 2025.

Project: Non-Invasive fMRI Semantic Reconstruction

10/2025-Present

Project Leader

- Developed a continuing research line on non-invasive fMRI-to-text decoding, focusing on how to recover temporally structured semantics from noisy, high-dimensional brain signals.
- Designed FuzzyStateFormer, a neuro-fuzzy state-aware framework with Transformer temporal encoding for long-sequence fMRI semantic decoding, aiming to improve state continuity and reduce unstable generation from weak neural evidence.
- Designed BASS, a brain-aligned semantic scaffold framework that separates transcript-side semantic state induction, brain-side posterior scaffold inference, and language-model-conditioned text generation.
- Representative outputs: FuzzyStateFormer, and BASS under review.

OPEN-SOURCE & ENGINEERING HIGHLIGHTS

- GTCRN-Light: Developed a deployment-oriented lightweight speech enhancement system by compressing GTCRN at the operator level while preserving the ERB-CRM pipeline and DPGRNN structure. The released implementation includes training, batch inference, paired-output generation, and model-complexity scripts; the v3 spec-core model reports 0.008M parameters and 0.017G MACs/s for 1-second 16 kHz input.
- Agent Harness: Built a local-first orchestration framework for long-running AI coding and research workflows, with task/phase/decision/experiment tracking, durable task and session memory, JSON index generation, a local control-plane server, and a GUI workspace for task inspection.
- Research Engineering Portfolio: Maintained multiple research-oriented engineering workspaces across open-vocabulary segmentation, frequency-domain image analysis, medical segmentation, graph-temporal forecasting, and

recommender systems, emphasizing reproducible training/evaluation, visualization tools, data processing pipelines, and rapid prototype-to-experiment conversion.

HONORS & AWARDS

- China Robot and Artificial Intelligence Competition (CRAIC) 2025 – **National Second Prize**
- China Robot and Artificial Intelligence Competition (CRAIC) 2024 – **National First Prize**
- Chinese Collegiate Computing Competition 2024 – **National Second Prize**