Wenrui Li

+86 17602834618 | liwr618@163.com
Wechat:hitlwr | https://liwrui.github.io/

EDUCATION

Harbin Institute of Technology (HIT) Sep 2021 - Present

Ph.D Candidate, Computer Science (Supervisor： Xiaopeng Fan)

Research Topics: Multimedia learning, Computer vision, Joint Source Channel Coding, Spiking Neural Network, AI4S

University of Electronic Science and Technology of China (UESTC) Sep 2017 - Jun 2021

Bachelor of Science, Software Engineering

PROJECT EXPERIENCE

Project Leader:

[1] 2D/3D Image and Graphics Alignment, Fusion, and Perception-Based Collaborative Interaction System supported by National Natural Science Foundation of China, 2025.01-2027.12.

This project addresses major challenges in constructing 2D/3D image-graphic alignment, fusion, and perception collaborative interaction systems, focusing on three key areas: First, **cross-modal feature alignment, matching, and generative techniques,** including a word-domain association reasoning module based on probabilistic cross-modal embedding, a multi-scale text-point cloud matching framework based on Riemannian geometry, and a text-driven complex scene generation framework based on 3DGS. Second, **adaptive multi-modal feature fusion technology based on Tucker decomposition**, including a spatiotemporal alignment-based hierarchical fusion framework for audio-visual features and a content-based deep relationship disentanglement and reasoning framework for audio-visual features. Third, **multi-modal 3D perception and semantic communication technology based on spiking neural networks**, including an efficient feature compression and encoding framework for multi-modal semantic communication based on spiking neural networks, a joint optimization strategy for multi-modal inference models and joint source-channel coding, and a distributed multi-agent collaborative decision-making scheme based on multi-modal information alignment and scene generation.

[2] Trusted multimodal alignment fusion and transmission supported by Fundamental Research Funds for the Central Universities, 2024.08-2025.08.

This project focuses on three key research directions: First, **a trustworthy multi-modal alignment framework based on cross-modal probabilistic embedding**, including the design and optimization of cross-modal probabilistic projection functions and the interpretability of multi-scale image-text alignment based on probabilistic embedding. Second, **an adaptive multi-modal feature fusion framework based on Tucker decomposition**, including hierarchical fusion of audio-visual features using deep networks and the reasoning and optimization of deep relationships between audio-visual features. Third, **a multi-modal spiking transmission framework based on spiking neural networks**, including efficient compression and encoding of multi-modal semantic communication transmission features, and the joint optimization of multi-modal inference models and source-channel joint coding.

Project Participator:

[1] National Key R&D Program: Research and Application Demonstration of Key Technologies for Network Audio-Visual Panoramic Interactive New Business Forms, Dec 2021 - Nov 2024.

I am primarily responsible for **multi-modal content retrieval and 3D asset library generation**. My research focuses on exploring multi-modal alignment techniques, cross-modal retrieval methods, and constructing asset libraries based on efficient 3D scene generation. As part of this effort, I proposed a Riemann-based Multi-scale Attention Reasoning Network (RMARN) for text-3D cross-modal retrieval and developed the large-scale Text-3D Retrieval dataset T3DR-HIT. The related research findings have been published in AAAI 2025.

[2] National Natural Science Foundation Key Project: Research on Key Technologies of Intelligent Generation and Real-time Encoder of 8K Ultra High Definition Video, Jun 2023 - May 2026.

I am primarily responsible for **high-definition video encoding and multi-modal feature fusion**. I proposed a novel method combining Spiking Neural Networks (SNNs) and Transformers for efficient multi-modal feature extraction and encoding in ultra-high-definition video scenarios, significantly improving encoding efficiency and the robustness of multi-modal interactions. The related research results have been published in IEEE Transactions on Image Processing (TIP) and ACM Multimedia 2023 (ACMMM'23).

[3] Pengcheng Laboratory Key Research Project: New Low Power Computing Architecture and System, Jun 2021 - May 2023.

I am primarily responsible for **designing a low-power spiking semantic communication framework.** My research focused on leveraging Spiking Neural Networks (SNNs) to optimize the semantic communication process by efficiently extracting and transmitting multi-modal information through neuron activation mechanisms. I proposed a spiking activation-based sender and a recurrent architecture-based receiver design, which significantly reduced bandwidth requirements and enhanced noise robustness in wireless communication scenarios. These contributions have been applied to image-text retrieval tasks and published in IEEE Transactions on Circuits and Systems for Video Technology (TCSVT).

[4] School-Enterprise Cooperation Project (Huawei): Task-Oriented Source Channel Joint Coding, May 2021 - Apr 2024.

I am primarily responsible for **designing a novel spiking semantic communication framework**. My research focused on a source-channel joint coding method based on Spiking Neural Networks (SNNs), leveraging SNNs' sparsity and temporal modeling capabilities to optimize semantic communication. This framework significantly reduces data volume in low-bandwidth transmission scenarios while enhancing noise robustness and ensuring task performance stability. The related outcomes have been applied to food recognition tasks and published in IEEE Transactions on Multimedia (TMM).

[5] AIGC Joint Laboratory (Peking University Shenzhen Research Institute and TuZhan Intelligence): AIGC Project, Sep 2023 - Jul 2024.

I am primarily responsible for **text-driven 3D scene generation**. My research focused on developing a framework for generating high-quality, 3D-consistent scenes based on textual descriptions. By integrating panoramic image generation with 3D Gaussian Splatting techniques, I proposed the SceneDreamer360 framework, which significantly improves the visual consistency and detail quality of generated scenes. The related outcomes have been applied to complex scene generation tasks and demonstrated superior performance compared to existing methods.

RESEARCH EXPERIENCE

Peking University ShenZhen Graduate School Oct 2023 - Jun 2024

Visiting Student (Supervisor : Yonghong Tian) ShenZhen

- Focused on research in multimodal AI-generated content (AIGC), especially making significant progress in text-to-3D complex scene generation. By introducing innovative neural network architectures and optimization algorithms, we have significantly enhanced the model's ability to understand and generate complex scenes.
- This research emphasizes effectively integrating text, images, and other data sources to generate high-quality 3D scenes.
- We can generate detailed and complex 3D indoor scenes from a single user sentence, with a strong focus on the spatial semantic consistency of the generated 3D scenes. This work has been submitted to IEEE TCSVT.

Peng Cheng Laboratory Mar 2022 - Mar 2023

Visiting Student (Supervisor : Yonghong Tian) ShenZhen

- Designed the spiking neural network for image-text retrieval to apply wireless communication scenarios.
- Utilized the pre-trained architecture to implement state-of-the-art techniques to ensure accurate and reliable results.
- Implemented the first application of spiking neural networks in semantic communication, which reduces bandwidth and is robust to channel noise while ensuring the performance of image-text retrieval, successfully published in IEEE TCSVT.
- Designed the spiking transformer model to extract fine-grained temporal information for audio-visual zero-shot learning.
- proposed spiking transformer gains Nearly 10% on performance for audio-visual generalized zero-shot learning and published in ICME-23.

PUBLICATIONS (*Corresponding Author)

Accepted Papers:

First Author or Corresponding Author (TIP: 1, TCSVT: 2, TMM: 1, ACM MM: 2, AAAI: 2):

[1] **Wenrui Li**, Penghong Wang, Ruiqin xiong and Xiaopeng Fan*. “Spiking Tucker Fusion Transformer for Audio-Visual Zero-Shot Learning” IEEE Transactions on Image Processing. (IEEE TIP)

[2] **Wenrui Li**, Ruiqin Xiong and Xiaopeng Fan*. “ Multi-layer Probabilistic Association Reasoning Network for Image-Text Retrieval” in IEEE Transactions on Circuits and Systems for Video Technology (IEEE TCSVT)

[3] **Wenrui Li**, Jiahui Li, Mengyao Ma, Xiaopeng Hong and Xiaopeng Fan*.” Multi-Scale Spiking Pyramid Wireless Communication Framework for Food Recognition,” in IEEE Transactions on Multimedia (IEEE TMM), 2024.

[4] **Wenrui Li**, Zhengyu Ma, LiangJian Deng, Xiaopeng Fan* and Yonghong Tian.”Neuron-Based Spiking Transmission and Reasoning NetworkFor Robust Image-Text Retrieval,” in IEEE Transactions on Circuits and Systems for Video Technology (IEEE TCSVT), 2022, doi:10.1109/TCSVT.2022.3233042.

[5] **Wenrui Li**, Zhe Yang, Wei Han, Hengyu Man, Xingtao Wang* and Xiaopeng Fan. “Hyperbolic-constraint Point Cloud Reconstruction from Single RGB-D Images”, in AAAI, 2025.

[6] **Wenrui Li**, Wei Han, Yandu Chen, Yeyu Chai, Yidan Lu, Xingtao Wang# and Xiaopeng Fan. “Riemann-based Multi-scale Attention Reasoning Network for Text-3D Retrieval”, in AAAI, 2025.

[7] **Wenrui Li**, Zhengyu Ma*, LiangJian Deng, Penghong Wang, Jinqiao Shi and Xiaopeng Fan*. “Reservoir Computing Transformer for Image Text Retrieval,” in ACM International Conference on Multimedia (ACM MM), Ottawa, Canada, 2023.

[8] **Wenrui Li**, XiLe Zhao, Zhengyu Ma*, Xingtao Wang, Xiaopeng Fan* and Yonghong Tian. “Motion Decoupled Spiking Transformer for Audio Visual Zero Shot Learning,” in ACM International Conference on Multimedia (ACM MM), Ottawa, Canada, 2023.

[9] **Wenrui Li**, Zhengyu Ma, Jinqiao Shi* and Xiaopeng Fan. “The style transformer with common knowledge optimization for image text retrieval,” in IEEE Signal Processing Letter (SPL), 2023.

[10] **Wenrui Li**, Zhengyu Ma, LiangJian Deng and Xiaopeng Fan*. “Modality Fusion Spiking Transformer Network for AudioVisual Zero Shot Learning,” in IEEE International Conference on Multimedia and Expo (ICME), Oral, Brisbane, Australia, 2023.

[11] **Wenrui Li** and Xiaopeng Fan*. “Image-Text Alignment and Retrieval Using Light-Weight Transformer”, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022.

[12] Zhe Yang, **Wenrui Li***, Jingxiu Hou and Guanghui Cheng. Multi-Modal Spiking Tensor Regression Network for Audio-Visual Zero-Shot Learning, Neurocomputing, 2025.

[13] Jinyu Guo, Yuejia Li, Guanghui Cheng and **Wenrui Li***. Based-CLIP early fusion transformer for image caption. Signal, Image and Video Processing, 19, 112 (2025). https://doi.org/10.1007/s11760-024-03721-0.
[14] **Wenrui Li**, Jifei Miao and Guanghui Cheng*. “A Jacobi-Like Algorithm for the General Joint Diagonalization Problem with Its Application to Blind Source Separation,” 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, 2019.

Collaborative Paper (AAAI: 1, ACM MM: 2, IOTJ: 1):

[15] Jisheng Chu, **Wenrui Li**, Xingtao Wang*, Ning Kanglin, Yidan Lu and Xiaopeng Fan. “Digging into Intrinsic Contextual Information for High-fidelity 3D Point Cloud Completion”, in AAAI, 2025.
[16] Haonan Zheng, Xinyang Deng*, Wen Jiang, and **Wenrui Li**, “A Unified Understanding of Adversarial Vulnerability Regarding Unimodal Models and Vision-Language Pre-training Models” in ACM International Conference on Multimedia, 2024.
[17] Haonan Zheng, Wen Jiang*, Xinyang Deng and **Wenrui Li**, “Sample-agnostic Adversarial Perturbation for Vision-Language Pre-training Models”, in ACM International Conference on Multimedia, 2024.
[18] Penghong Wang, Xingtao Wang, **Wenrui Li**, Xiaopeng Fan* and D. Zhao. 2024. “DV-Hop Localization Based On Distance Estimation Using Multi-Node and Hop Loss in IoT” in IEEE Internet of Things Journal (IEEE IoTJ),doi: 10.1109/JIOT.2024.3404492.
[19] Ming Guo, **Wenrui Li**, Chao Wang, Yuxin Ge and Chongjun Wang*. 2024. “SMILE: Spiking Multi-modal Interactive Label-Guided Enhancement Network for Emotion Recognition,” 2024, in IEEE International Conference on Multimedia and Expo.
[20] Yuchuan Feng, Jihang Jiang, Jie Ren, Ruotong Li*, **Wenrui Li** and Xiaopeng Fan. “Text-Guided Editable 3D City Scene Generation,” 2025, in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
[21] Jifei Miao, Guanghui Cheng*, **Wenrui Li** and Gong Zhang. “Non-orthogonal approximate joint diagonalization of non-Hermitian matrices in the least-squares sense,” Neurocomputing, 2019.
[22] Jifei Miao, Guanghui Cheng*, **Wenrui Li** and Eric Moreau. “A unitary joint diagonalization algorithm for nonsymmetric higher-order tensors based on Givens-like rotations,” Numerical Linear Algebra with Applications, 2020.
[23] Guanghui Cheng*, Jifei Miao, **Wenrui Li**, “Two Jacobi-like algorithms for the general joint diagonalization problem with applications to blind source separation,” Chinese Journal of Electronics, doi: 10.23919/cje.2019.00.102, 2022.

Submitted Paper:

[1] **Wenrui Li**, Wei Han, Xingtao Wang, Wangmeng Zuo, Xiaopeng Fan* and Yonghong Tian, “Video Summarization with Mixed Bernoulli distribution of Graph Representation Learning,” in IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI) –Revise and Resubmit as New.

[2] **Wenrui Li**, Wei Han, Liang-jian Deng, Ruiqin Xiong and Xiaopeng Fan*. “Spiking Variational Graph Representation Inference for Video Summarization”, in IEEE Transactions on Image Processing, (IEEE TIP) –Submitted.

[3] **Wenrui Li**, Penghong Wang, Xingtao Wang, Wangmeng Zuo, Xiaopeng Fan* and Yonghong Tian, “Multi-Timescale Motion-Decoupled Spiking Transformer for Audio-Visual Zero-Shot Learning,” in IEEE Transactions on Circuits and Systems for Video Technology (IEEE TCSVT) – Under review.

*** Three papers submitted to ICCV-25, two as first author, one as corresponding author.**

*** Two papers submitted to ACM MM-25 as first author.**

SELECTED AWARDS

- Ph.D. China National Scholarship, Ministry of Education of the People's Republic of China, (1.5%) 2023
- Bydauto Scholarship, Bydauto, (1/69) 2024.
- Mathematical Contest in Modeling, Meritorious Winner. ID:2018520. (winning ratio 7.09%) 2020
- Outstanding Graduate of UESTC (5%) 2021
- Outstanding Student Scholarship of UESTC (12%) 2018,2019,2020

ACADEMIC SERVICES

- **Serving as Journal Reviewer:**

IEEE Transactions on Circuits and Systems for Video Technology (IEEE TCSVT),

IEEE Transactions on Multimedia (IEEE TMM),

IEEE Transactions on Image Processing (IEEE TIP)

- **Serving as Conference Reviewer:**

IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2024-2025
IEEE/CVF International Conference on Computer Vision (ICCV) 2025
Annual Conference on Neural Information Processing Systems(NeurIPS) 2024
AAAI Conference on Artificial Intelligence (AAAI) 2023–2025

ACM Multimedia Conference (ACM MM) 2024-2025

IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2023-2025

IEEE International Conference on Multimedia and Expo (ICME) 2024-2025