

# Licai Sun

Postdoctoral researcher at University of Oulu  
Address: Pentti Kaiteran katu 1, 90570, Oulu, Finland  
Email: Licai.Sun@oulu.fi or sfoohun@gmail.com  
Homepage: <https://sunlicai.github.io>



## RESEARCH INTERESTS

Affective Computing, Multimodal Emotion Recognition (Sentiment Analysis), Facial Expression Recognition, Speech Emotion Recognition, Self-Supervised Learning, Multimodal Large Language Model

## RESEARCH EXPERIENCE

**Center for Machine Vision and Signal Analysis, University of Oulu**, Oulu, Finland      September 2024 | Present  
Postdoctoral Researcher  
Project: Mutual Gaze and Facial Expression Estimation from Videos  
Collaborator: Prof. Guoying Zhao and Prof. Jukka Leppänen

## EDUCATION

**Institute of Automation, Chinese Academy of Sciences**, Beijing, China      September 2019 | July 2024  
Ph.D. in Computer Applied Technology  
Thesis Title: Research on Emotion Recognition Based on Multimodal Representation Learning and Fusion  
Advisor: Prof. Jianhua Tao

**University of Chinese Academy of Sciences**, Beijing, China      September 2016 | July 2019  
M.Eng. in Computer Technology  
Thesis Title: Study on the Modulation Characteristics of Muscle Synergies during Human Upper-limb Reaching Movements  
Advisor: Prof. Shiwei Ye

**Beijing Forestry University**, Beijing, China      September 2012 | July 2016  
B.Eng. in Electronic and Information Technology

## PUBLICATIONS

Google Scholar: [https://scholar.google.com/citations?user=7qo\\_cTcAAAAJ&hl=en&oi=ao](https://scholar.google.com/citations?user=7qo_cTcAAAAJ&hl=en&oi=ao)  
Journal Paper

- **Licai Sun**, Zheng Lian, Kexin Wang, Yu He, Mingyu Xu, Haiyang Sun, Bin Liu, Jianhua Tao. SVFAP: Self-supervised Video Facial Affect Perceiver. IEEE Transactions on Affective Computing (TAC), 2024.
- **Licai Sun**, Zheng Lian, Bin Liu, Jianhua Tao. HiCMAE: Hierarchical Contrastive Masked Autoencoder for Self-Supervised Audio-Visual Emotion Recognition. Information Fusion (INFFUS), 2024.
- **Licai Sun**, Zheng Lian, Bin Liu, Jianhua Tao. Efficient Multimodal Transformer with Dual-Level Feature Restoration for Robust Multimodal Sentiment Analysis. IEEE Transactions on Affective Computing (TAC), 2023.
- Zheng Lian, **Licai Sun**, Haiyang Sun, Kang Chen, Zhuofan Wen, Hao Gu, Bin Liu, Jianhua Tao. GPT-4V with Emotion: A Zero-shot Benchmark for Generalized Emotion Recognition. Information Fusion (INFFUS), 2024.
- Zheng Lian, Lan Chen, **Licai Sun**, Bin Liu, Jianhua Tao. GCNet: graph completion network for incomplete multimodal learning in conversation. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2023.

### Conference Paper

- **Licai Sun**, Zheng Lian, Bin Liu, Jianhua Tao. MAE-DFER: Efficient Masked Autoencoder for Self-supervised Dynamic Facial Expression Recognition. ACM International Conference on Multimedia (ACM MM), 2023.
- **Licai Sun**, Bin Liu, Jianhua Tao, Zheng Lian. Multimodal Cross- and Self-Attention Network for Speech Emotion Recognition. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2021.
- **Licai Sun**, Mingyu Xu, Zheng Lian, Bin Liu, Jianhua Tao, Meng Wang, Yuan Cheng. Multimodal emotion recognition and sentiment analysis via attention enhanced recurrent model. Proceedings of the 2nd on Multimodal Sentiment Analysis Challenge (MuSe@ACM MM), 2021.
- **Licai Sun**, Zheng Lian, Bin Liu, Jianhua Tao, Mingyue Niu. Multi-modal continuous dimensional emotion recognition using recurrent neural network and self-attention mechanism. Proceedings of the 1st International on Multimodal Sentiment Analysis in Real-life Media Challenge and Workshop (MuSe@ACM MM), 2020.

- Zheng Lian, Haiyang Sun, **Licai Sun**, Kang Chen, Mingyu Xu, Kexin Wang, Ke Xu, Yu He, Ying Li, Jinming Zhao, Ye Liu, Bin Liu, Jiangyan Yi, Meng Wang, Erik Cambria, Guoying Zhao, Björn W Schuller, Jianhua Tao. MER 2023: Multi-label Learning, Modality Robustness, and Semi-supervised Learning. ACM International Conference on Multimedia (ACM MM), 2023.
- Ke Xu, Kang Chen, **Licai Sun**, Zheng Lian, Bin Liu, Gong Chen, Haiyang Sun, Mingyu Xu, Jianhua Tao. Integrating VideoMAE based model and Optical Flow for Micro-and Macro-expression Spotting. ACM International Conference on Multimedia (ACM MM), 2023.
- Haiyang Sun, Zheng Lian, Bin Liu, Ying Li, **Licai Sun**, Cong Cai, Jianhua Tao, Meng Wang, Yuan Cheng. EmotionNAS: Two-stream Architecture Search for Speech Emotion Recognition. Interspeech, 2023.
- Yu He, **Licai Sun**, Zheng Lian, Bin Liu, Jianhua Tao, Meng Wang, Yuan Cheng. Multimodal Temporal Attention in Sentiment Analysis. Proceedings of the 3rd International on Multimodal Sentiment Analysis Workshop and Challenge (MuSe@ACM MM), 2022.
- Kexin Wang, Zheng Lian, **Licai Sun**, Bin Liu, Jianhua Tao, Yin Fan. Emotional reaction analysis based on multi-label graph convolutional networks and dynamic facial expression recognition transformer. Proceedings of the 3rd International on Multimodal Sentiment Analysis Workshop and Challenge (MuSe@ACM MM), 2022.

## Preprint

- Zheng Lian, **Licai Sun**, Yong Ren, Hao Gu, Haiyang Sun, Lan Chen, Bin Liu, Jianhua Tao. MERBench: A Unified Evaluation Benchmark for Multimodal Emotion Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), *Submitted*.
- Zheng Lian, **Licai Sun**, Mingyu Xu, Haiyang Sun, Ke Xu, Zhuofan Wen, Shun Chen, Bin Liu, Jianhua Tao. Explainable Multimodal Emotion Reasoning. ArXiv, 2023.

## PROJECTS

---

### Self-Supervised Emotion Representation Learning

March 2022 | Present

- Previous efforts in video emotion recognition (VER) are dominated by the supervised learning paradigm. Despite significant progress, supervised learning is meeting its bottleneck due to the longstanding data scarcity issue. Motivated by recent advances of supervised learning, we pioneeringly explored masked autoencoder (MAE) for VER and achieved significant improvements over previous SOTA supervised methods on six VER datasets.
- Since audio-visual emotion recognition (AVER) faces the same challenges as VER, we extended the above work to the audio-visual domain and proposed a self-supervised audio-visual emotion recognition framework. It integrates masked data modeling and contrastive learning and introduces a three-pronged strategy to promote hierarchical audio-visual representation learning. The proposed method achieved great success on nine AVER datasets covering both categorical and dimensional tasks.

### Multimodal Fusion for Emotion Recognition

September 2019 | Present

- We utilized self-attention and cross-attention mechanisms to capture fine-grained intra-modal and cross-modal interactions for improved audio-text emotion recognition.
- Previous multimodal fusion methods for unaligned multimodal sequences suffers from quadratic scaling costs and are also vulnerable to random modality feature missing which usually occurs in real-world scenarios. Therefore, we developed an efficient multimodal Transformer which enjoys linear computational complexity over the involved modalities and introduced dual-level feature restoration to improve the model robustness in realistic settings.

### Emotion Recognition in the Era of Multimodal Large Language Models

April 2023 | Present

- We introduced a novel multimodal emotion reasoning task (i.e., not only give the final emotion prediction but also provide reasons or clues in each modality to justify its prediction) and benchmarked several advanced multimodal large language models on this task.
- We evaluated the zero-shot emotion understanding ability of GPT-4V on several unimodal and multimodal benchmarks.

## HONORS and AWARDS

---

- Micro-Expression and Macro-Expression Spotting Task (Winner) in MEGC 2023@ACM MM 2023
- The MuSe-Stress 2022 Multimodal Sentiment Analysis Challenge Prize (Winner) in MuSe 2022@ACM MM 2022
- The MuSe-Wilder 2021 Multimodal Sentiment Analysis Challenge Prize (Winner) in MuSe 2021@ACM MM 2021
- The MuSe-Sent 2021 Multimodal Sentiment Analysis Challenge Prize (Winner) in MuSe 2021@ACM MM 2021
- The MuSe-Physio 2021 Multimodal Sentiment Analysis Challenge Prize (Winner) in MuSe 2021@ACM MM 2021
- The 2020 Multimodal Sentiment in-the-Wild Challenge Prize (Winner) in MuSe 2020@ACM MM 2020

## PATENTS

---

- Jianhua Tao, **Licai Sun**, Bin Liu, and Zheng Lian. Multimodal dimensional emotion recognition method. U.S. Patent 11,281,945, issued March 22, 2022.

- Jianhua Tao, **Licai Sun**, Bin Liu, and Zheng Lian. Multi-modal lie detection method and apparatus, and device. U.S. Patent 11,244,119, issued February 8, 2022.

## PERSONAL SKILLS

---

- Programming Languages: Python, Matlab, LaTeX.
- Deep Learning: PyTorch, TensorFlow.
- Platforms: Windows, Linux.
- Languages: Mandarin (native), English.

## ACADEMIC SERVICES

---

- Journal Reviewer: IEEE Transactions on Affective Computing, Speech Communication, Engineering Applications of Artificial Intelligence, Neurocomputing, Pattern Analysis and Application.
- Conference Reviewer: ACM MM (2024, 2023), PRCV (2024), ICASSP (2022), InterSpeech (2020).
- Program Committee: MER Workshop (2023, 2024)@ACM MM.