Jinhua Liang

Homepage liangjh0903@gmail.com Github Google scholar

RESEARCH INTEREST

My research expertise lies in the fields of Multimodal Learning for Audio Perception, Generative AI for Audio Content Creation and Editing, (Low-Shot) Acoustic Pattern Recognition, Auditory Scene Analysis, Model Compression and Acceleration, and Audio Emotion Analysis. Specifically, I have hands-on research experience on developing large language models (LLMs) as a universal system for audio understanding or generation.

EDUCATION

Queen Mary University of London 09/2021 - Present
PhD in Electronic Engineering *London, United Kingdom*

- Advised by Dr. Emmanouil Benetos, Dr. Huy Phan, Prof. Mark Sandler
- Funded by QMUL Principle Studentship and EPSRC

Tianjin University 09/2018 – 01/2021
MSc in Information and Communication Engineering *Tianjin, China*

Tianjin University 09/2014 – 06/2018
BEng in Electronic Information Engineering *Tianjin, China*

WORK EXPERIENCE

Microsoft Research 06/2024 – Present
Research Intern *Redmond, WA, United States*

Queen Mary University of London 09/2023 – Present
Teaching Fellow *London, United Kingdom*

- Responsible for curating the MLEnd dataset to let students have a deeper understanding of machine learning in practical applications.
- In charge of the lab sessions and Q&A sessions in the module *Principle of Machine Learning*
- Supervise the final-year projects of post-graduate and undergraduate students at Queen Mary.

ByteDance 02/2024 – 06/2024
Research Intern *Hybrid*

HIGHLIGHTED RESEARCH

Audio-Visual Learning for Understanding and Generation

- Proposed a unified model that learning feature representation and generation.
- Experimented both audio and video trained encoders on classification and retrieval tasks.
- Introduced generative models to model the conditional probability distribution for each modality.

Adapting LLMs for Audio Understanding

- Proposed an audio language model that ingests multiple audio clips and generate text tokens by interleaving the acoustic embeddings with text embeddings in a sequence.
- Proposed a training receipt by combining curriculum learning and multi-task learning.
- Evaluated the proposed audio large language model with various downstream tasks.
- Extended large language models and visual large models to the audio domain.

LLM-Based Agents for Audio Creation and Editing

- Applied LLMs, such as ChatGPT, to create/edit audio content based on user instructions and available recordings.
- Produced audio content in a controllable manner by coordinating various generative models.
- Evaluated the proposed system ability on audio drama where models should manipulate audio content without explicit user commands.

Bootstrapping Audio Language Models in Few-Shot Learning

- Improved Contrastive Language-Audio Pretrained networks (CLAPs) performance with a few examples while preserving its ability to zero-shot classification.
- Proposed a new module to retrieve labels of the test examples by measuring the affinity between test and support embeddings.
- Devised a cosine initialisation strategy such that the proposed methods can benefit from the few-shot settings even without training.

PUBLICATIONS

Conference paper

- **Jinhua Liang**, Yi Yuan, Dongya Jia, Xiaobin Zhuang, Zhengxi Liu, Yuanzhe Chen, Zhuo Chen, Yuping Wang, Yuxuan Wang. (2024) "AudioMorphix: Training-free audio editing with diffusion probabilistic models", in peer review.
- Jiehui Jia, Huan Zhang, **Jinhua Liang**. (2024) "Bridging Discrete and Continuous: A Multimodal Strategy for Complex Emotion Detection", in peer review.
- Tanisha Hisariya, Huan Zhang, **Jinhua Liang**. (2024) "Bridging Paintings and Music – Exploring Emotion based Music Generation through Paintings", in peer review.
- Wen Qing Lim, **Jinhua Liang**, Huan Zhang. (2024) "Hierarchical Symbolic Pop Music Generation with Graph Neural Networks", in peer review.
- Huan Zhang, Shreyan Chowdhury, Carlos Eduardo Cancino-Chacón, **Jinhua Liang**, Simon Dixon, Gerhard Widmer. "DExter: Learning and Controlling Performance Expression with Diffusion Models". Applied Sciences. 2024; 14(15):6543.
- Huan Zhang, **Jinhua Liang**, Simon Dixon. (2024) "From Audio Encoders to Piano Judges: Benchmarking Performance Understanding for Solo Pian", International Society of Music Information Retrieval (ISMIR).
- **Jinhua Liang**, Ines Nolasco, Burooj Ghani, Huy Phan, Emmanouil Benetos, Dan Stowell. "Mind the Domain Gap: a Systematic Analysis on Bioacoustic Sound Event Detection", The 33rd European Signal Processing Conference (EUSIPCO 2025).
- **Jinhua Liang**, Huan Zhang, Haohe Liu, Yin Cao, Qiuqiang Kong, Xubo Liu, Wenwu Wang, Mark D. Plumbley, Huy Phan, Emmanouil Benetos."WavCraft: Audio Editing and Generation with Natural Language Prompts", International Conference on Learning Representations (ICLR) 2024 Workshop on LLM Agents.
- **Jinhua Liang**, Huy Phan, Emmanouil Benetos, "Learning from Taxonomy: Multi-Label Few-Shot Classification for Everyday Sound Recognition," 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Korea, Republic of, 2024, pp. 771-775.
- **Jinhua Liang**, Xubo Liu, Haohe Liu, Huy Phan, Emmanouil Benetos, Mark D. Plumbley, Wenwu Wang. "Adapting Language-Audio Models as Few-Shot Audio Learners", in Proc. INTERSPEECH 2023, 2023, pp. 276-280.
- Yi Yuan, Haohe Liu, **Jinhua Liang**, Xubo Liu, Mark D. Plumbley and Wenwu Wang, "Leveraging Pre-Trained AudioLDM for Sound Generation: A Benchmark Study," 2023 31st European Signal Processing Conference (EUSIPCO), Helsinki, Finland, 2023, pp. 765-769, doi: 10.23919/EUSIPCO58844.2023.10289975.
- Ren Li, **Jinhua Liang**, Huy Phan "Few-Shot Bioacoustic Event Detection: Enhanced Classifiers for Prototypical Networks". Proceedings of the 7th Detection and Classification of Acoustic Scenes and Events 2022 Workshop (DCASE2022). Nancy, France.
- **Jinhua Liang**, Huy Phan, Emmanouil Benetos (2022). "Leveraging Label Hierachies for Few-Shot Everyday Sound Recognition". Proceedings of the 7th Detection and Classification of Acoustic Scenes and Events 2022 Workshop (DCASE2022). Nancy, France.
- **Jinhua Liang**, Huy Phan, and Emmanouil Benetos. "Everyday Sound Recognition with Limited Annotations" Digital Music Research Network (DMRN +16), 2021.

Journal paper

✉ *denotes the corresponding author*

- **Jinhua Liang**, Xubo Liu, Wenwu Wang, Mark D. Plumbley, Huy Phan, Emmanouil Benetos."Acoustic Prmpt Tuning: Empowering Large Language Models with Audition Capabilities" (in Peer Review).
- Xubo Liu, Zhongkai Zhu, Haohe Liu, Yi Yuan, Meng Cui, Qiushi Huang, **Jinhua Liang**, Yin Cao, Quiqiang Kong, Mark D. Plumbley, Wenwu Wang. "Wavjourney: Compositional audio creation with large language models", in Peer Review.
- Biyun Ding, Tao Zhang, Chao Wang, Ganjun Liu, **Jinhua Liang**, Ruimin Hu, Yulin Wu, Difei Guo."Acoustic scene classification: A comprehensive surve", Expert Systems with Applications, vol. 238, p. 121902, 2024.
- Andrew Mitchell, Emmeline Brown, Ratneel Deo, Yuanbo Hou, Jasper Kirton-Wingate, **Jinhua Liang**, Alisa Sheinkman, Christopher Soelistyo, Hari Sood, Arin Wongprommoon, Kaiyue Xing, Wingyan Yip, Francesco Aletta; Deep learning techniques for noise annoyance detection: Results from an intensive workshop at the Alan Turing Institute. J. Acoust. Soc. Am. 1 March 2023; 153 (3_supplement): A262.
- Tao Zhang, **Jinhua Liang**✉, and Guoqing Feng. "Adaptive time-frequency feature resolution network for acoustic scene classification." Applied Acoustics 195 (2022): 108819.
- Tao Zhang, Shuang Li, Guoqing Feng, **Jinhua Liang**✉, Lun He, Xin Zhao. "Local channel transformation for efficient convolutional neural network." Signal, Image and Video Processing (2022): 1-9.
- Tao Zhang, Guoqing Feng, **Jinhua Liang**, Tong An. "Acoustic scene classification based on Mel spectrogram decomposition and model merging," in Applied Acoustics, 182, 108258, 2021.

- **Jinhua Liang**, Tao Zhang✉ and Guoqing Feng, "Channel Compression: Rethinking Information Redundancy Among Channels in CNN Architecture," in IEEE Access, vol. 8, pp. 147265-147274, 2020.
- Tao Zhang, **Jinhua Liang**✉, and Biyun Ding, "Acoustic scene classification using deep cnn with fine-resolution feature", Expert Systems with Applications, vol. 143, pp. 113067, 2020.

Technical Report

- Andrew Mitchell, **Jinhua Liang**, et al. "Deep Learning Techniques for Noise Annoyance Detection", Data Study Group, the Alan Turing Institute, Dec. 20, 2022.
- Ren Li, **Jinhua Liang**, Huy Phan. "FEW-SHOT BIOACOUSTIC EVENT DETECTION USING PROTOTYPICAL NETWORKS WITH RESNET CLASSIFIER", DCASE2022 Challenge, June 2022.
- Guoqing Feng, **Jinhua Liang**, and Biyun Ding, "Acoustic Scene Classification Based on Lightweight CNN With Efficient Convolutions," Tech. Rep., DCASE2020 Challenge, June 2020.
- Biyun Ding, Ganjun Liu, and **Jinhua Liang**, "Acoustic Scene Classification Based on Ensemble System," Tech. Rep., DCASE2019 Challenge, June 2019.

TEACHING EXPERIENCE

ECS423U Teaching Assistant in *Engineering Maths* (Employer: Dr. Emmanouil Benetos) 2023
ECS766P Teaching Assistant in *Data Mining* (Employer: Dr. Emmanouil Benetos) 2022
ECS7013P Teaching Assistant in *Deep Learning for Audio & Music* (Employer: Dr. Huy Phan) 2022
Teaching Assistant in *Digital Logic Circuit* (Employer: Dr. Tao Zhang) 2020

AWARDS & HONORS

Scholarships

UK EPSRC Studentship 2021 - Present
QMUL Principle Studentship 2021 - Present

Prizes

The Third Prize of the North China Division, the 14th China Graduate Electronic Design Competition 2019

Other funds

Postgraduate Research Fund 2023
DCASE 2022 Workshop Travel funds 2022

Honors

UK Global Talent Endorsement (from *The Royal Academy of Engineering*) 2024

COMMUNITY SERVICES

Chef student reps in School of Electronic Engineering and Computer Science, QMUL 2021 - 2024
Organiser for DCASE challenge Task 5 Few-shot Bioacoustic Event Detection
Reviewer for IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)
Reviewer for IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)
Reviewer for Detection and Classification of Acoustic Scenes and Events (DCASE)
Reviewer for Expert Systems with Applications
Reviewer for ACM Transactions on the Web
Reviewer for IEEE Access
Reviewer for IEEE Signal Processing Letters
Reviewer for Applied Acoustics
Reviewer for Transactions on Audio, Speech and Language Processing (TASLP)

SKILLS

Concepts: Machine Listening, Deep Learning, Audio Signal Processing, Natural Language Processing, Computer Vision
Languages: Python, LINUX, MATLAB, C++
Deep learning: Pytorch, TensorFlow 1.x/2.x, Keras