HUANG XIE

Machine Learning PhD (6+ years research) | Software Engineer (7+ years)
Audio & Multimodal AI | Low-Resource Learning | Representation Learning

Personal Info

- Huang Xie
- huang.xie@outlook.com
- +358 40 3219383
- Tampere, Finland
- github.com/xieh97
- linkedin.com/in/huang-xie-28b7872bb
- google scholar

Languages

Chinese	*Native*
English	*Professional*
Finnish	*Beginner*

Referees

Prof. Tuomas Virtanen
@ Tampere University
tuomas.virtanen@tuni.fi

Assoc. Prof. Okko Räsänen
@ Tampere University
okko.rasanen@tuni.fi

About Me

Machine Learning researcher and engineer with 6+ years of PhD research in audio intelligence and multimodal representation learning, complemented by 7+ years of software engineering experience. My work focuses on bridging cutting-edge ML techniques with real-world applications—from developing self-supervised audio-language models to optimizing low-resource learning systems. Passionate about turning theoretical advances into deployable solutions, especially in audio understanding and multimodal retrieval.

Employment

Doctoral Researcher | Research Assistant 03/2019 - Present

Audio Research Group, Tampere University, Finland

- Designed and implemented large-scale ML models (CNNs, Transformers) for audio-language representation learning, audio classification, and multimodal retrieval, leveraging self-supervised/contrastive/transfer learning techniques.
- Built and curated large-scale audio and NLP datasets via Amazon MTurk; performed data processing, clustering, feature extraction, semantic and statistical analysis.
- Coordinated the Language-Based Audio Retrieval task for DCASE Challenge (2022-2025), contributing to standardized evaluation protocols.
- Published 9+ first-author papers at top-tier venues (IEEE/ACM TASLP, SPL, ICASSP), advancing ML methods in low-resource learning, representation learning, and multimodal information retrieval.

Software Engineer 10/2015 - 05/2018

Bohai Commodity Exchange, Tianjin, China

- Designed and implemented core features for *CloudBoce*, an e-commerce platform, including shopping cart workflows, order processing systems, and product management tools.
- Collaborated closely with UI/UX designers and product managers to define, design, and deliver new features aligned with business and user needs.
- Diagnosed and resolved performance bottlenecks, crashes, and stability issues, significantly improving platform reliability and user experience.

Skills

Programming: Python, Java, Scala, JavaScript, SQL, C/C++, LaTeX
Machine Learning: PyTorch, TensorFlow, scikit-learn, Ray Tune, Spark
Audio & NLP: librosa, torchaudio, NLTK
Data Analysis: NumPy, SciPy, Pandas, Jupyter, Matplotlib
Web & Backend: Java EE, Spring, Hibernate, Django, Microservices, RESTful APIs
Databases & DevOps: MySQL, PostgreSQL, Linux, Docker, Git
General Skills: problem solving, critical thinking, continuous learning, attention to detail, teamwork

Education

PhD in Signal Processing and Machine Learning 01/2021 - Present
Tampere University, Tampere, Finland

M.Sc. in Data Engineering and Machine Learning 08/2018 - 11/2020
Tampere University, Tampere, Finland

M.Eng. in Software Engineering 09/2010 - 06/2014
University of Science and Technology of China, Hefei, China

Activities

- Active reviewer for IEEE/ACM TASLP, SPL, ICASSP, IJCNN, WASPAA, etc.
- Task coordinator for Language-based Audio Retrieval and Automated Audio Captioning in DCASE Challenge 2022 (), 2023 (), and 2024 ().

Publications (full list)

[1] Text-based Audio Retrieval by Learning from Similarities between Audio Captions
H. Xie, K. Khorrami, O. Räsänen, and T. Virtanen
2024 IEEE Signal Processing Letters arXiv

[2] Integrating Continuous and Binary Relevances in Audio-Text Relevance Learning
H. Xie, K. Khorrami, O. Räsänen, and T. Virtanen
2024 in Proc. Detect. Classif. Acoust. Scenes Events Work. (DCASE) arXiv

[3] Multi-label Zero-Shot Audio Classification with Temporal Attention
D. Dogan, **H. Xie**, T. Heittola, and T. Virtanen
2024 in Proc. Int. Workshop Acoust. Signal Enhanc. (IWAENC) arXiv

[4] Crowdsourcing and Evaluating Text-Based Audio Retrieval Relevances
H. Xie, K. Khorrami, O. Räsänen, and T. Virtanen
2023 in Proc. Detect. Classif. Acoust. Scenes Events Work. (DCASE) arXiv

[5] On Negative Sampling for Contrastive Audio-Text Retrieval
H. Xie, O. Räsänen, and T. Virtanen
2023 in Proc. Int. Conf. Acoustic., Speech and Signal Process. (ICASSP) arXiv

[6] Language-based Audio Retrieval Task in DCASE 2022 Challenge
H. Xie, S. Lipping, and T. Virtanen
2022 in Proc. Detect. Classif. Acoust. Scenes Events Work. (DCASE) arXiv

[7] Unsupervised Audio-Caption Aligning Learns Correspondences Between Individual Sound Events and Textual Phrases
H. Xie, O. Räsänen, K. Drossos, and T. Virtanen
2022 in Proc. Int. Conf. Acoustic., Speech and Signal Process. (ICASSP) arXiv

[8] Zero-Shot Audio Classification using Image Embeddings
D. Dogan, **H. Xie**, T. Heittola, and T. Virtanen
2022 in Proc. Eur. Signal Process. Conf. (EUSIPCO) arXiv

[9] Zero-Shot Audio Classification with Factored Linear and Nonlinear Acoustic-Semantic Projections
H. Xie, O. Räsänen, and T. Virtanen
2021 in Proc. Int. Conf. Acoustic., Speech and Signal Process. (ICASSP) arXiv

[10] Zero-Shot Audio Classification via Semantic Embeddings
H. Xie, and T. Virtanen
2021 IEEE/ACM Trans. Audio Speech Lang. Process. (TASLP) arXiv

[11] Zero-Shot Audio Classification Based on Class Label Embeddings
H. Xie, and T. Virtanen
2019 in Proc. Work. Appl. Signal Process. Audio and Acoustic. (WASPAA) arXiv