Di Hu

dihu@ruc.edu.cn | dtaoo.github.io | gewu-lab.github.io

Education

Northwestern Polytechnical University 2014 – 2019
Ph.D in Computer Science and Technologys
Advisor: Feiping Nie, Xuelong Li
Thesis: Research on Machine Multimodal Perception

Honor College, Northwestern Polytechnical University 2010 – 2014
Bachelor in Computer Science and Technology

Experience

Tenure-track Associate Professor Jul. 2023 – Present
Gaoling School of Artificial Intelligence, Renmin University of China

Tenure-track Assistant Professor Aug. 2020 – Jul. 2023
Gaoling School of Artificial Intelligence, Renmin University of China

Research Scientist Jul. 2019 – Aug. 2020
Baidu Inc.

Research Interest

Interested in how to understand and interact with the environment via the natural multimodal messages, *e.g., sound, vision and touch*. I'm strongly convinced that the pervasive multimodal messages can provide sufficient information for perceiving, interacting, learning and understanding environment, even the agent itself, which promisingly makes multimodal learning become one of the key to achieve machine general intelligence.

Distinction

WuWenJun AI Excellent Young Scientist Award 2023
The Young Elite Scientists Sponsorship Program 2022
SHAANXI Outstanding Doctoral Dissertation Award 2021
CAAI Outstanding Doctoral Dissertation Award 2020
ACM Xi'an Doctoral Dissertation Award 2019
Baidu AIDU Recruitment Program 2019
CVPR Doctoral Consortium 2019
CSC Scholarship to CMU as a Visiting Scholar 2018
National Scholarship 2017,2018
RoboCup China Open 2014
The First Prize, Service Robot@Home

Publications

Conference Paper (†: Equal Contribution, *: Corresponding Author)

26. Guangyao Li, Wenxuan Hou, and **Di Hu***. Progressive Spatio-temporal Perception for Audio-Visual Question Answering. *In Proceedings of the ACM Conference on Multimedia* (**ACMMM**), 2023.
25. Hongpeng Lin†, Ludan Ruan†, Wenke Xia†, Peiyu Liu, Jingyuan Wen, Yixin Xu, **Di Hu**, Ruihua Song, Wayne Xin Zhao, Qin Jin, and Zhiwu Lu. TikTalk: A Video-Based Dialogue Dataset for Multi-Modal Chitchat in Real World. *In Proceedings of the ACM Conference on Multimedia* (**ACMMM**), 2023.
24. Andong Deng, Xingjian Li, **Di Hu***, Tianyang Wang, Haoyi Xiong, Chengzhong Xu. Towards Inadequately Pre-trained Models in Transfer Learning. *In Proceedings of the IEEE Conference on Computer Vision* (**ICCV**), 2023.
23. Guangyao Li, Yixin Xu, and **Di Hu***. Multi-Scale Attention for Audio Question Answering. *Interspeech*, 2023.
22. Wenke Xia, Xingjian Li, Andong Deng, Haoyi Xiong, Dejing Dou, and **Di Hu***. Robust Cross-Modal Knowledge Distillation for Unconstrained Videos. *IEEE International Conference on Multimedia and Expo* (**ICME**), 2023.

21. Ruize Xu, Ruoxuan Feng, Shi-xiong Zhang, and **Di Hu***. MMCosine: Multi-Modal Cosine Loss Towards Balanced Audio-Visual Fine-Grained Learning. *The International Conference on Acoustics, Speech, & Signal Processing* (**ICASSP**), 2023.
20. Xinchi Zhou, Dongzhan Zhou, **Di Hu**, Hang Zhou, and Wanli Ouyang. Exploiting Visual Context Semantics for Sound Source Localization. *IEEE/CVF Winter Conference on Applications of Computer Vision* (**WACV**), 2022.
19. Xinchi Zhou, Dongzhan Zhou, Wanli Ouyang, Hang Zhou, and **Di Hu**. SeCo: Separating Unknown Musical Visual Sounds with Consistency Guidance. *IEEE/CVF Winter Conference on Applications of Computer Vision* (**WACV**), 2022.
18. Xiaokang Peng†, Yake Wei†, Andong Deng, Dong Wang, and **Di Hu***. Balanced Multimodal Learning via On-the-fly Gradient Modulation. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (**CVPR**), 2022. **Oral Presentation**
17. Guangyao Li†, Yake Wei†, Yapeng Tian†, Chenliang Xu, Ji-Rong Wen, and **Di Hu***. Learning to Answer Questions in Dynamic Audio-Visual Scenarios. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (**CVPR**), 2022. **Oral Presentation**
16. Xian Liu, Rui Qian, Hang Zhou, **Di Hu**, Weiyao Lin, Ziwei Liu, Bolei Zhou, and Xiaowei Zhou. Visual Sound Localization in-the-Wild by Cross-Modal Interference Erasing. *In Proceedings of the AAAI Conference on Artificial Intelligence* (**AAAI**), 2022.
15. Dongzhan Zhou, Xinchi Zhou, **Di Hu***, Hang Zhou, Lei Bai, Ziwei Liu, and Wanli Ouyang. SepFusion: Finding Optimal Fusion Structures for Visual Sound Separation. *In Proceedings of the AAAI Conference on Artificial Intelligence* (**AAAI**), 2022.
14. Zechen Bai, Zhigang Wang, Jian Wang, **Di Hu***, and Errui Ding*. Unsupervised Multi-Source Domain Adaptation for Person Re-Identification. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (**CVPR**), 2021. **Oral Presentation**
13. Yapeng Tian, **Di Hu***, and Chenliang Xu*. Cyclic Co-Learning of Sounding Object Visual Grounding and Sound Separation. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (**CVPR**), 2021.
12. Dong Wang, **Di Hu***, Xingjian Li, and Dejing Dou. Temporal Relational Modeling with Self-Supervision for Action Segmentation. *In Proceedings of the AAAI Conference on Artificial Intelligence* (**AAAI**), 2021.
11. **Di Hu**, Rui Qian, Minyue Jiang, Xiao Tan, Shilei Wen, Errui Ding, Weiyao Lin, and Dejing Dou. Discriminative Sounding Objects Localization via Self-supervised Audiovisual Matching. *In Advances in Neural Information Processing Systems* (**NeurIPS**), 2020.
10. **Di Hu**, Xuhong Li, Lichao Mou, Pu Jin, Dong Chen, Liping Jing, Xiaoxiang Zhu, and Dejing Dou. Cross-Task Transfer for Geotagged Audiovisual Aerial Scene Recognition. *In Proceedings of the European Conference on Computer Vision* (**ECCV**), 2020.
9. Rui Qian, **Di Hu**, Heinrich Dinkel, Mengyue Wu, Ning Xu, and Weiyao Lin. Multiple Sound Sources Localization from Coarse to Fine. *In Proceedings of the European Conference on Computer Vision* (**ECCV**), 2020.
8. **Di Hu**, Dong Wang, Xuelong Li, Feiping Nie, and Qi Wang. Listen to the Image. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (**CVPR**), 2019.
7. **Di Hu**, Feiping Nie, and Xuelong Li. Deep Multimodal Clustering for Unsupervised Audiovisual Learning. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (**CVPR**), 2019.
6. **Di Hu**, Chengze Wang, Feiping Nie, and Xuelong Li. Dense Multimodal Fusion for Hierarchically Joint Representation. *In Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing* (**ICASSP**), 2019.
5. Xuelong Li, **Di Hu**, and Feiping Nie. Large Graph Hashing with Spectral Rotation. *In Proceedings of the AAAI Conference on Artificial Intelligence* (**AAAI**), 2017.
4. Xuelong Li, **Di Hu**, and Feiping Nie. Deep Binary Reconstruction for Cross-modal Hashing. *In Proceedings of the ACM Conference on Multimedia* (**ACMMM**), 2017.
3. Xuelong Li, **Di Hu**, and Xiaoqiang Lu. Image2song: Song Retrieval via Bridging Image Content and Lyric Words. *In Proceedings of the IEEE Conference on Computer Vision* (**ICCV**), 2017.
2. **Di Hu**, Xiaoqiang Lu, and Xuelong Li. Multimodal Learning via Exploring Deep Semantic Similarity. *In Proceedings of the ACM Conference on Multimedia* (**ACMMM**), 2016.
1. **Di Hu**, Xuelong Li, and Xiaoqiang Lu. Temporal Multimodal Learning in Audiovisual Speech Recognition. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (**CVPR**), 2016.

Journal Paper

8. Ziyun Li, Jona Otholt, Ben Dai, **Di Hu**, Christoph Meinel, and Haojin Yang. Supervised Knowledge May Hurt Novel Class Discovery Performance. *Transactions on Machine Learning Research* (**TMLR**), 2023.
7. Konrad Heidler, Lichao Mou, **Di Hu**, Pu Jin, Guangyao Li, Chuang Gan, Ji-Rong Wen, Xiao-Xiang Zhu. Self-supervised Audiovisual Representation Learning for Remote Sensing Data. *In International Journal of Applied Earth Observation and Geoinformation*, 2023.
6. **Di Hu**, Zheng Wang, Feiping Nie, Rong Wang, Xuelong Li. Self-supervised Learning for Heterogeneous Audiovisual Scene Analysis. *In IEEE Trans. Multimedia* (**TMM**), 2022.
5. **Di Hu**, Yake Wei, Rui Qian, Weiyao Lin, Ruihua Song, Ji-Rong Wen. Class-aware Sounding Objects Localization via Audiovisual Correspondence. *In IEEE Trans. Pattern Analysis and Machine Intelligence* (**TPAMI**), 2021.
4. Sijia Yang, Haoyi Xiong, **Di Hu**, Kaibo Xu, Licheng Wang, Peizhen Zhu, Zeyi Sun. Generalising Combinatorial Discriminant Analysis through Conditioning Truncated Rayleigh Flow. *Knowledge and Information Systems* (**KAIS**), 2021.
3. **Di Hu**, Feiping Nie, and Xuelong Li. Deep Linear Discriminant Analysis Hashing. *In SCIENTIA SINICA Informationis*, 2019.
2. **Di Hu**, Feiping Nie, and Xuelong Li. Discrete Spectral Hashing for Efficient Similarity Retrieval. *In IEEE Trans. Image Processing* (**TIP**), 2018.
1. **Di Hu**, Feiping Nie, and Xuelong Li. Deep Binary Reconstruction for Cross-modal Hashing. *In IEEE Trans. Multimedia* (**TMM**), 2018.

Workshop Paper

5. Wenke Xia† , Xu Zhao† , Xincheng Pang , Changqing Zhang, and **Di Hu***. Balanced Audiovisual Dataset for Imbalance Analysis. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop* (**CVPRW**), 2023.
4. **Di Hu**, Zheng Wang, Haoyi Xiong, Dong Wang, Feiping Nie, and Dejing Dou. Heterogeneous Scene Analysis via Self-supervised Audiovisual Learning. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop* (**CVPRW**), 2020.
3. **Di Hu***, Lichao Mou*, Qingzhong Wang*, Junyu Gao, Yuansheng Hua, Dejing Dou, and Xiaoxiang Zhu. Does Ambient Sound Help? - Audiovisual Crowd Counting. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop* (**CVPRW**), 2020.
2. Yapeng Tian*, **Di Hu***, and Chenliang Xu. Co-Learn Sounding Object Visual Grounding and Visually Indicated Sound Separation in A Cycle Video. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop* (**CVPRW**), 2020.
1. Rui Qian, **Di Hu**, Heinrich Dinkel, Mengyue Wu, Ning Xu, and Weiyao Lin. A Two-Stage Framework for Multiple Sound-Source Localization. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop* (**CVPRW**), 2020.

Professional Services

Organizing Committee

CVPR Tutorial on Audio-visual Scene Understanding	2021
WACV Tutorial on Audio-visual Scene Understanding	2021
ICDM Tutorial on Automated Deep Learning: Theory, Algorithms, Platforms, and Applications	2019

Senior Program Committee

The AAAI Conference on Artificial Intelligence (AAAI)	2023,2024
The International Joint Conference on Artificial Intelligence (IJCAI)	2023

Program Committee

IEEE Conference on Computer Vision and Pattern Recognition (CVPR)	2018, 2020-2023
IEEE International Conference on Computer Vision (ICCV)	2019, 2021, 2023
European Conference on Computer Vision (ECCV)	2020, 2022
The AAAI Conference on Artificial Intelligence (AAAI)	2018, 2020-2022
International Conference in Learning Representations (ICLR)	2021-2023
Neural Information Processing Systems (NeurIPS)	2020-2023
The International Conference on Machine Learning (ICML)	2021-2023

Asian Conference on Computer Vision (ACCV) 2018, 2020

IEEE Winter Conference on Applications of Computer Vision (WACV) 2021

Journal Reviewer

IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)

IEEE Transactions on Neural Networks and Learning Systems (TNNLS)

IEEE Transactions on Image Processing (TIP)

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

EEE Transactions on Multimedia (TMM)

IEEE Transactions on Knowledge and Data Engineering (TKDE)

ACM Transactions on Intelligent Systems and Technology (TIST)

Invited Talks

Effective Multimodal Learning Mechanism and Scene Understanding Mar. 2023
Beijing Institute of Technology

Effective Multimodal Learning Mechanism and Scene Understanding Sep. 2022
ByteDance AI Lab

Hear the Object, See the Sound Nov. 2020
VALSE Webinar

Audiovisual Machine Perception and Learning Dec. 2019
Beijing Jiaotong Univeristy

Machine Multimodal Perception Aug. 2019
Xidian Univeristy

Machine Audio-visual Perception Dec. 2018
Big Data Lab, Baidu Inc.