

Di Hu

dihu@ruc.edu.cn | [dtaoo.github.io](https://github.com/dtaoo) | github.com/DTaoo

EDUCATION

| | |
|--|-------------|
| Northwestern Polytechnical University <i>Ph.D in Computer Science and Technologys</i> <i>Advisor: Feiping Nie, Xuelong Li</i> <i>Thesis: Research on Machine Multimodal Perception</i> | 2014 – 2019 |
| Honor College, Northwestern Polytechnical University <i>Bachelor in Computer Science and Technology</i> | 2010 – 2014 |

EXPERIENCE

| | |
|--|-----------------------|
| Assistant Professor <i>Gaojing School of Artificial Intelligence, Renmin University of China</i> | Aug. 2020 – Present |
| Research Scientist <i>Baidu Inc.</i> | Jul. 2019 – Aug. 2020 |

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 intelligence.

DISTINCTION

| | |
|---|-----------|
| 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 <i>The First Prize, Service Robot@Home</i> | 2014 |

PUBLICATIONS

Conference Paper

16. Xian Liu, Rui Qian, Hang Zhou, **Di Hu**, Weiyao Lin, Ziwei Liu, Bolei Zhou, 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, Xinchu Zhou, **Di Hu***, Hang Zhou, Lei Bai, Ziwei Liu, 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***, 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***, 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

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

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 |

Program Committee

| | |
|---|-----------------|
| IEEE Conference on Computer Vision and Pattern Recognition (CVPR) | 2018, 2020-2022 |
| IEEE International Conference on Computer Vision (ICCV) | 2019, 2021 |
| European Conference on Computer Vision (ECCV) | 2020 |
| The AAAI Conference on Artificial Intelligence (AAAI) | 2018, 2020-2022 |
| International Conference in Learning Representations (ICLR) | 2021 |
| Neural Information Processing Systems (NeurIPS) | 2020-2021 |
| The International Conference on Machine Learning (ICML) | 2021 |
| 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) |
| IEEE Transactions on Multimedia (TMM) |
| IEEE Transactions on Knowledge and Data Engineering (TKDE) |
| ACM Transactions on Intelligent Systems and Technology (TIST) |

INVITED TALKS

| | |
|--|-----------|
| Hear the Object, See the Sound | Nov. 2020 |
| <i>VALSE Webinar</i> | |
| Audiovisual Machine Perception and Learning | Dec. 2019 |
| <i>Beijing Jiaotong Univeristy</i> | |
| Machine Multimodal Perception | Aug. 2019 |
| <i>Xidian Univeristy</i> | |
| Machine Audio-visual Perception | Dec. 2018 |
| <i>Big Data Lab, Baidu Inc.</i> | |

STUDENT ADVISING

Ph.D students

| | |
|---|-----------------|
| Guangyao Li (Homepage: ayameyao.github.io) | Since Sep. 2020 |
| Yake Wei (Homepage: echo0409.github.io) | Since Sep. 2021 |

Master students

| | |
|---------------|-----------------|
| Xiaokang Peng | Since Sep. 2020 |
| Yixin Xu | Since Sep. 2020 |

Visiting Scholar

| | |
|--|-----------------|
| Andong Deng | Since Aug. 2021 |
| Rui Qian (Homepage: shvdiwnkozbw.github.io) | Since Jan. 2020 |