SHIZHE CHEN

Email: shizhe.chen@inria.fr
Website: https://cshizhe.github.io

EMPLOYMENT

INRIA, Postdoc *Mar. 2021 - Present*
Robotic navigation and manipulation. Advisor: Ivan Laptev and Cordelia Schmid

Carnegie Mellon University, Researcher *Jun. 2020 - Dec. 2020*
Zero-shot action recognition. Advisor: Dong Huang

University of Adelaide, Visiting Scholar *Jul. 2019 - Oct. 2019*
Controllable image caption generation. Advisor: Qi Wu

Microsoft Research Asia, Research Intern *Dec. 2018 - Jun. 2019*
Neural storyboard creation. Mentors: Jianlong Fu and Ruihua Song

Carnegie Mellon University, Visiting Scholar *Oct. 2017 - Oct. 2018*
Video caption generation. Advisor: Alexander Hauptmann

EDUCATION

Renmin University of China *Sep. 2015 - Jun. 2020*
Ph.D in Computer Science Advisor: Qin Jin

Renmin University of China *Sep. 2011 - Jun. 2015*
B.S in Computer Science Advisor: Qin Jin

HONORS

- ⋆ Top Global Young Chinese Female Scholars in AI released by Baidu Academic. 2023
- ⋆ AI 2000 Most Influential Scholar Honorable Mention in Multimedia released by AI TIME. 2022
- ⋆ Outstanding Graduate Award (Highest honor for graduate set by the government of Beijing). 2020
- ⋆ JingDong Scholarship (10 students at Renmin University). 2019
- ⋆ ACM Multimedia Student Travel Grant. 2019
- ⋆ ICMR Best Paper Runner up. 2018
- ⋆ Baidu Scholarship (10 Ph.D students worldwide). 2017
- ⋆ National Scholarship for Ph.D Students. 2016
- ⋆ ACM Multimedia Student Travel Grant. 2016
- ⋆ National Scholarship for Undergraduate Students. 2013

AWARDS

- ⋆ Ranked 2nd in CSIG REVERIE VLN Challenge. 2022
- ⋆ Ranked 1st in ICCV HIRV Workshop REVERIE & SOON VLN Challenges. 2021
- ⋆ Ranked 1st in CVPR ActivityNet Entities Object Localization Challenge. 2021
- ⋆ Ranked 1st in CVPR ActivityNet Dense Video Captioning Challenge. 2018 - 2020
- ⋆ Ranked 1st in NIST Trecvid Video to Text Challenge. 2017 - 2020
- ⋆ Ranked 2nd in CVPR Video Understanding Pentathlon Challenge. 2020
- ⋆ First Prize in Zhijiang Cup Global AI Competition Video Captioning Challenge. 2019
- ⋆ Ranked 2nd and won Outstanding Method Prize in ICCV VATEX Video Captioning Challenge. 2019
- ⋆ Ranked 1st in ACM Multimedia AVEC Emotion Recognition Challenge. 2017 - 2019
- ⋆ Ranked 1st in ACM Multimedia Video to Language Grand Challenge. 2016 - 2017

⋆ Ranked 2nd in ACM Multimedia AVEC Emotion Recognition Challenge. 2016
⋆ Ranked 1st in MediaEval Emotion Impact of Movies Task. 2016
⋆ Ranked 2nd in CCPR Multimodal Emotion Recognition Challenge. 2016
⋆ Second Prize in Chinese Big Data Contest P2P Task. 2015
⋆ Second Prize in IBM Bleumix Cognitive Computation Development Contest. 2015
⋆ First Prize in National College Student Information Security Contest. 2014
⋆ Second Prize in Chinese Big Data Contest Baidu IErMu Task. 2014
⋆ Meritorious Winner in American Mathematical Contest in Modeling. 2014
⋆ National Second Prize in China Undergraduate Mathematical Contest in Modeling. 2013

PROFESSIONAL ACTIVITIES

Area chair

- International Conference on Computer Vision (ICCV) 2023.
- IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2023.
- ACM Multimedia (ACM MM) 2022.

Workshop co-organizer

- ICMR 2021 Workshop on Multi-Modal Pre-Training for Multimedia Understanding
- CVPR 2020 Workshop on Language & Vision with applications to Video Understanding

Conference & journal reviewer

- Conferences: CVPR, ICCV, ECCV, AAAI, ACM MM, ACL, EMNLP
- Journals: TPAMI, TMM, TOMM, T-RO, RA-L

Supervision

- PhD students: Ricardo Garcia (co-advise), Zerui Chen (co-advise), Pierre-Louis Guhur (co-advised), Anwen Hu (co-advised), Sipeng Zheng (co-advised)
- Master students: Yuqing Song (co-advised), Yida Zhao (co-advised)

Teaching

- Course project advisor for Object Recognition and Computer Vision, 2021-2022
- Guest lecturer at Spoken Language Processing, 2019
- Teaching assistant for Data Structure and Algorithm, 2015
- Teaching assistant for Programming Practice with C Language, 2015

Grants

- Participant, Intelligence Advanced Research Projects Activity (IARPA) No. D17PC00340
  Deep Intermodal Video Analytics (DIVA)
- Participant, Large-Scale Pre-Training Program of Beijing Academy of Artificial Intelligence (BAAI)
  "WenLan" - Chinese multi-modal pre-training
- Participant, National Key Research and Development Plan No. 2016YFB1001202
  Computational Principles of Human-Computer Interaction
- Participant, Alibaba Damo Academy
  Image-guided Machine Translation for Commercial Products
- Participant, National Natural Science Foundation of China No. 61772535
  Multimodal Video Captioning with Deep Neural Networks
- Participant, Beijing Natural Science Foundation No. 4192028
  Language Understanding and Interaction Based on Auditory Information

INVITED PRESENTATIONS

⋆ Talk at IMAGINE research group at Ecole des Ponts ParisTech. 12/2022
⋆ Presentation at WILLOW/SIERRA retreat, Saint-Raphaël. 10/2022
⋆ Presentation at Stanford Vision and Learning Lab iGibson and BEHAVIOR team. 09/2022

⋆ Talk at Renmin University of China: Multimodal Perception and Action. 07/2022
⋆ Talk at AI TIME: Embodied Vision-and-Language Navigation in 3D Environments. 04/2022
⋆ Talk at Microsoft Research Asia: Recent Advances in Vision-and-Language Navigation. 03/2022
⋆ Presentation at WILLOW/SIERRA retreat, Avignon. 10/2021
⋆ Talk at Tencent: Fine-grained Video-Text Retrieval with Hierarchical Graph Reasoning. 06/2020

PUBLICATION

I have published over 30 scientific papers most of which appeared in international journals and major peer-reviewed conferences. The leading conferences in computer vision (ICCV, ECCV, CVPR), machine learning (NeurIPS), multimedia (ACM MM) and robotic learning (CoRL) have a low acceptance rate typically below 25%, and publications in their proceedings are considered as important as journal publications. Overall, my publications have **over 1,600 citations** and my **h-index is 20** (both obtained from Google Scholar).

1. Zerui Chen, **Shizhe Chen**, Cordelia Schmid, and Ivan Laptev. gsdf: Geometry-driven signed distance functions for 3d hand-object reconstruction. In *CVPR*, 2023
2. Pierre-Louis Guhur, **Shizhe Chen**, Ricardo Garcia, Makarand Tapaswi, Ivan Laptev, and Cordelia Schmid. Instruction-driven history-aware policies for robotic manipulations. In *CoRL*, 2022
3. **Shizhe Chen**, Pierre-Louis Guhur, Makarand Tapaswi, Cordelia Schmid, and Ivan Laptev. Language conditioned spatial relation reasoning for 3d object grounding. In *NeurIPS*, 2022
4. **Shizhe Chen**, Pierre-Louis Guhur, Makarand Tapaswi, Cordelia Schmid, and Ivan Laptev. Learning from unlabeled 3d environments for vision-and-language navigation. In *ECCV*, 2022
5. Sipeng Zheng, **Shizhe Chen**, and Qin Jin. Few-shot action recognition with hierarchical matching and contrastive learning. In *ECCV*, pages 297–313. Springer, 2022
6. **Shizhe Chen**, Pierre-Louis Guhur, Makarand Tapaswi, Cordelia Schmid, and Ivan Laptev. Think global, act local: Dual-scale graph transformer for vision-and-language navigation. In *CVPR*, 2022
7. Sipeng Zheng, **Shizhe Chen**, and Qin Jin. Vrdformer: End-to-end video visual relation detection with transformers. In *CVPR*, pages 18836–18846, 2022
8. **Shizhe Chen**, Pierre-Louis Guhur, Cordelia Schmid, and Ivan Laptev. History aware multimodal transformer for vision-and-language navigation. In *NeurIPS*, 2021
9. Pierre-Louis Guhur, Makarand Tapaswi, **Shizhe Chen**, Ivan Laptev, and Cordelia Schmid. Airbert: In-domain pretraining for vision-and-language navigation. In *ICCV*, 2021
10. **Shizhe Chen** and Dong Huang. Elaborative rehearsal for zero-shot action recognition. *ICCV*, 2021
11. Anwen Hu, **Shizhe Chen**, and Qin Jin. Question-controlled text-aware image captioning. *ACM MM*, 2021
12. Yuqing Song, **Shizhe Chen**, Qin Jin, Wei Luo, Jun Xie, and Fei Huang. Product-oriented machine translation with cross-modal cross-lingual pre-training. In *ACM MM*, pages 2843–2852, 2021
13. Yuqing Song, **Shizhe Chen**, Qin Jin, Wei Luo, Jun Xie, and Fei Huang. Enhancing neural machine translation with dual-side multimodal awareness. *IEEE Transactions on Multimedia*, 2021
14. Chaorui Deng, **Shizhe Chen**, Da Chen, Yuan He, and Qi Wu. Sketch, ground, and refine: Top-down dense video captioning. In *CVPR*, pages 234–243, 2021
15. Yuqing Song, **Shizhe Chen**, and Qin Jin. Towards diverse paragraph captioning for untrimmed videos. In *CVPR*, pages 11245–11254, 2021
16. **Shizhe Chen**, Qin Jin, Peng Wang, and Qi Wu. Say as you wish: Fine-grained control of image caption generation with abstract scene graphs. In *CVPR*, pages 9962–9971, 2020
17. **Shizhe Chen**, Yida Zhao, Qin Jin, and Qi Wu. Fine-grained video-text retrieval with hierarchical graph reasoning. In *CVPR*, pages 10638–10647, 2020

18. Anwen Hu, **Shizhe Chen**, and Qin Jin. Icecap: Information concentrated entity-aware image captioning. In *ACM Multimedia*, pages 4217–4225, 2020

19. Sipeng Zheng, **Shizhe Chen**, and Qin Jin. Skeleton-based interactive graph network for human object interaction detection. In *ICME*, pages 1–6, 2020

20. **Shizhe Chen**, Bei Liu, Jianlong Fu, Ruihua Song, Qin Jin, Pingping Lin, Xiaoyu Qi, Chunting Wang, and Jin Zhou. Neural storyboard artist: Visualizing stories with coherent image sequences. In *ACM Multimedia*, pages 2236–2244, 2019

21. Yuqing Song, **Shizhe Chen**, Yida Zhao, and Qin Jin. Unpaired cross-lingual image caption generation with self-supervised rewards. In *ACM Multimedia*, pages 784–792, 2019

22. Sipeng Zheng, **Shizhe Chen**, and Qin Jin. Visual relation detection with multi-level attention. In *ACM Multimedia*, pages 121–129, 2019

23. **Shizhe Chen**, Qin Jin, and Jianlong Fu. From words to sentences: A progressive learning approach for zero-resource machine translation with visual pivots. In *IJCAI*, pages 4932–4938, 2019

24. **Shizhe Chen**, Qin Jin, and Alexander G. Hauptmann. Unsupervised bilingual lexicon induction from mono-lingual multimodal data. In *AAAI*, pages 8207–8214, 2019

25. Weiying Wang, Yongcheng Wang, **Shizhe Chen**, and Qin Jin. Youmakeup: A large-scale domain-specific multimodal dataset for fine-grained semantic comprehension. In *EMNLP*, pages 5136–5146, 2019

26. **Shizhe Chen**, Qin Jin, Jia Chen, and Alexander G Hauptmann. Generating video descriptions with latent topic guidance. *IEEE Trans. Multimedia*, 21(9):2407–2418, 2019

27. **Shizhe Chen**, Jia Chen, Qin Jin, and Alexander Hauptmann. Class-aware self-attention for audio event recognition. In *ICMR*, pages 28–36, 2018

28. **Shizhe Chen**, Jia Chen, Qin Jin, and Alexander Hauptmann. Video captioning with guidance of multimodal latent topics. In *ACM Multimedia*, pages 1838–1846, 2017

29. Qin Jin, Jia Chen, **Shizhe Chen**, Yifan Xiong, and Alexander Hauptmann. Describing videos using multi-modal fusion. In *ACM Multimedia*, pages 1087–1091, 2016

30. **Shizhe Chen** and Qin Jin. Multi-modal conditional attention fusion for dimensional emotion prediction. In *ACM Multimedia*, pages 571–575, 2016

Paris

Mar 1, 2023