SHIZHE CHEN

Email: cszhe1@ruc.edu.cn
Website: https://cshizhe.github.io

RESEARCH INTEREST

Vision-and-Language, Video Understanding, Affective Computing, Multimodal Machine Learning

EDUCATION

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

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

RESEARCH EXPERIENCE

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

RESEARCH PROJECTS

Vision & Language

- **Video Caption Generation**: proposed topic-guided video captioning model and contextual-aware event captioning system; published in ICMR, ACM MM and TMM; winner of MSR-VTT 2016-2017, Trecvid VTT 2016-2019, CVPR ActivityNet Dense Captioning 2018-2019.
- **Video-Text Retrieval**: proposed to disentangle videos and texts into hierarchical representations for fine-grained matching; published in CVPR.
- **Image Caption Generation**: proposed fine-grained controllable image caption generation and self-supervised cross-lingual image captioning; published in CVPR and ACM MM.
- **Visual-pivoted Machine Translation**: proposed to employ images as pivots to translate words and sentences in zero-resource condition; published in AAAI and IJCAI.

Multimodal Emotion Recognition

- **Multimodal Fusion**: proposed a conditional multimodal fusion model to dynamically fuse visual, acoustic and other modalities; published in ACM MM; winner of ACM AVEC 2017-2019.
- **Emotion in Dyadic Dialogs**: proposed to employ self and interlocutor's contexts to improve emotion understanding in dialogs; published in Interspeech.
- **Cross-culture Emotions**: proposed an adversarial framework to transfer emotion recognition models trained in one culture to another culture; published in ICASSP.

TECHNICAL SKILLS

Programming	Python, C/C++, Javascript, HTML, MATLAB, Bash
Framework	Pytorch, Tensorflow

AWARDS

- Ranked 1st in CVPR ActivityNet Dense Video Captioning Challenge. 2018 - 2019
- Ranked 1st in NIST Trecvid Video to Text Challenge. 2017 - 2019
- 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

HONORS

- JingDong Scholarship (10 students in 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

PUBLICATION

1. **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*, 2020

2. **Shizhe Chen**, Yida Zhao, Qin Jin, and Qi Wu. Fine-grained video-text retrieval with hierarchical graph reasoning. In *CVPR*, 2020

3. **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

4. 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

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

6. **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

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

8. 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

9. **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
10. Jinming Zhao, **Shizhe Chen**, Jingjun Liang, and Qin Jin. Speech emotion recognition in dyadic dialogues with attentive interaction modeling. In *Interspeech*, pages 1671–1675, 2019
11. Jingjun Liang, **Shizhe Chen**, Jinming Zhao, Qin Jin, Haibo Liu, and Li Lu. Cross-culture multimodal emotion recognition with adversarial learning. In *ICASSP*, pages 4000–4004, 2019
12. **Shizhe Chen**, Jia Chen, Qin Jin, and Alexander Hauptmann. Class-aware self-attention for audio event recognition. In *ICMR*, pages 28–36, 2018
13. **Shizhe Chen**, Jia Chen, Qin Jin, and Alexander Hauptmann. Video captioning with guidance of multimodal latent topics. In *ACM Multimedia*, pages 1838–1846, 2017
14. Qin Jin, Jia Chen, **Shizhe Chen**, Yifan Xiong, and Alexander Hauptmann. Describing videos using multi-modal fusion. In *ACM Multimedia*, pages 1087–1091, 2016
15. **Shizhe Chen**, Qin Jin, Jinming Zhao, and Shuai Wang. Multimodal multi-task learning for dimensional and continuous emotion recognition. In *ACM Multimedia AVEC Workshop*, pages 19–26, 2017
16. **Shizhe Chen** and Qin Jin. Multi-modal conditional attention fusion for dimensional emotion prediction. In *ACM Multimedia*, pages 571–575, 2016
17. Qin Jin, Chengxin Li, **Shizhe Chen**, and Huimin Wu. Speech emotion recognition with acoustic and lexical features. In *ICASSP*, pages 4749–4753, 2015

TEACHING EXPERIENCE

Guest Lecturer	Spoken Language Processing	2019
Teaching Assistant	Data Structure and Algorithm	2015
Teaching Assistant	Programming Practice with C Language	2015

SERVICE

Conference & Journal Reviewing

- Conferences: ACM MM, ACL, AAAI
- Journals: TMM, TOMM

Conference & Workshop Organizing

- Session Volunteer, ACM Multimedia 2019
- Program Committee, Language & Vision with applications to Video Understanding, CVPR 2020

PARTICIPATED GRANTS

- National Key Research and Development Plan No. 2016YFB1001202
Computational Principles of Human-Computer Interaction

- National Natural Science Foundation of China No. 61772535
Multimodal Video Captioning with Deep Neural Networks

- Beijing Natural Science Foundation No. 4192028
Language Understanding and Interaction Based on Auditory Information