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 generation. Mentors: Jianlong Fu and Ruihua Song

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

RESEARCH PROJECTS

Visual Caption Generation

- Short Video Captioning: proposed topic-guided video captioning models; published in ICMR, ACM MM and TMM; winner of MSR-VTT 2016-2017, Trecvid VTT 2016-2019.
- Dense Video Captioning: proposed event proposal generation and contextual-aware event caption generation system; winner of CVPR ActivityNet Dense Captioning 2018-2019.

Visual and Multi-lingual Languages

- Visual-pivoted Zero-resource Bilingual Lexicon Induction: proposed to employ images as pivots to translate bilingual words without parallel texts; published in AAAI 2018.
- Visual-pivoted Zero-resource Machine Translation: proposed a progressive learning framework for zero-resource sentence translation using images as bridge; published in IJCAI 2019.
- Cross-lingual Caption Generation: proposed to generate cross-lingual image descriptions without pairs in target language using self-supervised learning; published in ACM MM 2019.

Multimodal Emotion Recognition

- Multimodal Fusion: proposed a conditional multimodal fusion model which is able to dynamically focus on different modalities in different situations; published in ACM MM 2016.
- Emotion in Dyadic Dialogs: proposed to employ self and interlocutor's contexts to improve emotion understanding in dialogs; published in Interspeech 2019; winner of ACM MM AVEC 2017-2019.
- Cross-culture Emotions: proposed an adversarial framework to transfer emotion recognition models trained in one culture to another culture; published in ICASSP 2018.

TECHNICAL SKILLS

Programming	Python, C/C++, Javascript, 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. 2018 - 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

⋆ ACM Multimedia Student Travel Grant. 2019
⋆ ICMR Best Paper Runner up. 2018
⋆ Baidu Scholarship **(10 Ph.D student 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**, 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

2. 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

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

4. **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

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

6. 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

7. **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

8. 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

9. 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

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

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

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

13. **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

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

15. **Shizhe Chen**, Xinrui Li, Qin Jin, Shilei Zhang, and Yong Qin. Video emotion recognition in the wild based on fusion of multimodal features. In *ICMI*, pages 494–500, 2016

16. **Shizhe Chen** and Qin Jin. Multi-modal dimensional emotion recognition using recurrent neural networks. In *ACM Multimedia AVEC Workshop*, pages 49–56, 2015

17. Qin Jin, Chengxin Li, **Shizhe Chen**, and Huimin Wu. Speech emotion recognition with acoustic and lexical features. In *ICASSP*, pages 4749–4753, 2015

18. **Shizhe Chen**, Qin Jin, Xirong Li, Gang Yang, and Jieping Xu. Speech emotion classification using acoustic features. In *ISCSLP*, pages 579–583, 2014