Shizhe Chen

Ph.D. Candidate, Renmin University of China

✉ cszhe1@ruc.edu.cn

Research Interests

Video Captioning, Affective Computing, Deep Learning, Multimodal Machine Learning

Education

Renmin University of China — **Advisor: Prof. Qin Jin**
Ph.D., Computer Science — *Sep 2015 – Present*

Carnegie Mellon University — **Advisor: Prof. Alexander Hauptmann**
Visiting Reserach Scholar — *Oct 2017 – Oct 2018*

Renmin University of China — **Advisor: Prof. Qin Jin**
B.S., Computer Science — *Sep 2011 – Jul 2015*

Research Experiences

○ **Multilingual Captioning**
· unsupervised bilingual lexicon induction with multilingual image caption models.

○ **Video Natural Language Description**
· Dense Video Captioning: winner of the CVPR 2018 ActivityNet dense video captioning challenge.
· Short Video Captioning: winner of the MSR Video to Language Challenge 2016-2017 and the NIST Trecvid Video to Text task 2017-2018.
· Proposed the topic-guided caption models published in ICMR, MM.

○ **Multimodal Emotion Recognition**
· Continuous Emotion Prediction: winner of the ACM Multimedia AVEC 2017-2018, runner up in 2016.
· Discrete Emotion Recognition: top performance in EmotiW 2016-2017, MEC 2016-2017.
· Proposed models to capture temporal and context information and dynamic multimodal fusion, which are published in MM, ICMI etc.

Technical skills

Python, C, C++, Javascript, MATLAB, Bash
Pytorch, Tensorflow, HTML, CSS, Latex, Markdown

Awards and Honors

- Ranked 1st in CVPR 2018 ActivityNet Dense Video Captioning Challenge
- Ranked 1st in NIST Trecvid Video to Text Task 2018
- Ranked 1st in ACM Multimedia 2018 AVEC cross-culture emotion recognition challenge
- Baidu Scholarship in 2018 (10 Ph.D. students worldwide)
- Best paper runner up in ICMR 2018
- Ranked 1st in ACM Multimedia 2017 Video to Language Grand Challenge
- Ranked 1st in NIST Trecvid Video to Text Task 2017
- Ranked 1st in ACM Multimedia 2017 AVEC emotion recognition challenge
- Ranked 5th in ICMI 2017 EmotiW emotion recognition challenge
- National Scholarship for Ph.D. Students 2016
- Ranked 1st in ACM Multimedia 2016 Video to Language Grand Challenge
- Ranked 2nd in AVEC 2016 Continuous Emotion Recognition Sub-challenge
- Ranked 2nd in CCPR 2016 Multimodal Emotion Recognition Challenge
- Ranked 1st in MediaEval 2016 Emotion Impact of Movies Subtask
- ACM Multimedia Student Travel Grant 2016
- Second Prize in the Chinese Big Data Contest 2015
- Second Prize in IBM Bluemix Cognitive Computation Development Contest 2015
- Ranked 4th in AVEC 2015 Continuous Emotion Recognition Challenge
- First Prize in National College Student Information Security Contest 2014
- Second Prize in the Chinese Big Data Contest 2014
- Meritorious Winner in American Mathematical Contest in Modeling 2014
- National Second Prize in China Undergraduate Mathematical Contest in Modeling 2013
- National Scholarship for Undergraduate Students 2013

Publications

Shizhe Chen, Qin Jin, and Alexander Hauptmann. Unsupervised bilingual lexicon induction from mono-lingual multimodal data. In *AAAI*, 2019.

Shizhe Chen, Yuqing Song, Yida Zhao, Jiarong Qiu, Qin Jin, and Alexander Hauptmann. Ruc+cmu: System report for dense captioning events in videos. *CVPR ActivityNet Workshop*, 2018.

Shizhe Chen, Jia Chen, Qin Jin, and Alexander Hauptmann. Class-aware self-attention for audio event recognition. In *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*, pages 28–36. ACM, 2018. **Best Paper Runner-up**.

Shizhe Chen, Qin Jin, Jinming Zhao, and Shuai Wang. Multimodal multi-task learning for dimensional and continuous emotion recognition. In *Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge*, pages 19–26. ACM, 2017.

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

Shizhe Chen, Jia Chen, and Qin Jin. Generating video descriptions with topic guidance. In *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*, pages 5–13. ACM, 2017.

Qin Jin, **Shizhe Chen**, Jia Chen, and Alexander Hauptmann. Knowing yourself: Improving video caption via in-depth recap. In *Proceedings of the 2017 ACM on Multimedia Conference*, pages 1906–1911. ACM, 2017.

Shizhe Chen, Xinrui Li, Qin Jin, Shilei Zhang, and Yong Qin. Video emotion recognition in the wild based on fusion of multimodal features. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, pages 494–500. ACM, 2016.

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

Shizhe Chen and Qin Jin. Multi-modal dimensional emotion recognition using recurrent neural networks. In *Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge*, pages 49–56. ACM, 2015.