

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

Ph.D in Computer Science

Sep. 2015 - Present

Advisor: Qin Jin

Renmin University of China

B.S in Computer Science

Sep. 2011 - Jun. 2015

Advisor: Qin Jin

RESEARCH EXPERIENCE

University of Adelaide, Visiting Scholar

Image caption generation.

Jul. 2019 - Oct. 2019

Advisor: Qi Wu

Microsoft Research Asia, Research Intern

Neural storyboard generation.

Dec. 2018 - Jun. 2019

Mentors: Jianlong Fu and Ruihua Song

Carnegie Mellon University, Visiting Scholar

Video caption generation.

Oct. 2017 - Oct. 2018

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-pivoted Zero-resource Machine Translation

- Bilingual lexicon induction: proposed to employ grounded visual representation and word contextual representation learned from multilingual image caption datasets to translate words.
- Sentence translation: propose a progressive learning framework which firstly grasps word translation from image-pivots and then progresses to more challenging zero-resource sentence translation.

Multimodal Emotion Recognition

- Multimodal fusion: proposed a conditional multimodal fusion model which dynamically attends to more convincing modalities in different situations for emotion recognition.
- Emotion in dyadic dialogs: proposed to dynamically employ both self and interlocutor's contexts to improve emotion understanding in dialogs; winner of ACM Multimedia AVEC 2017-2019.
- Cross-culture emotions: proposed an adversarial framework to transfer the emotion recognition ability from one culture to another culture without sufficient training data.

TECHNICAL SKILLS

**Programming
Framework**

Python, C/C++, Javascript, MATLAB, Bash
Pytorch, Tensorflow

AWARDS

- ★ Ranked 1st in CVPR ActivityNet Dense Video Captioning Challenge. 2018 - 2019
- ★ Ranked 2nd in ICCV VATEX Video Captioning Challenge. 2019
- ★ Ranked 1st in NIST Trecvid Video to Text Challenge. 2017 - 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 (developed an APP for learning English) 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*, 2019
2. Yuqing Song, **Shizhe Chen**, Yida Zhao, and Qin Jin. Unpaired cross-lingual image caption generation with self-supervised rewards. In *ACM Multimedia*, 2019
3. **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
4. **Shizhe Chen**, Qin Jin, and Alexander G. Hauptmann. Unsupervised bilingual lexicon induction from mono-lingual multimodal data. In *AAAI*, pages 8207–8214, 2019
5. **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
6. Jinming Zhao, **Shizhe Chen**, Jingjun Liang, and Qin Jin. Speech Emotion Recognition in Dyadic Dialogues with Attentive Interaction Modeling. In *Proc. Interspeech 2019*, pages 1671–1675, 2019
7. 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
8. **Shizhe Chen**, Jia Chen, Qin Jin, and Alexander G. Hauptmann. Class-aware self-attention for audio event recognition. In *ICMR*, pages 28–36, 2018
9. **Shizhe Chen**, Jia Chen, Qin Jin, and Alexander G. Hauptmann. Video captioning with guidance of multimodal latent topics. In *ACM MM*, pages 1838–1846, 2017

10. **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, 2017
11. **Shizhe Chen** and Qin Jin. Multi-modal conditional attention fusion for dimensional emotion prediction. In *ACM MM*, pages 571–575, 2016
12. Qin Jin, Chengxin Li, **Shizhe Chen**, and Huimin Wu. Speech emotion recognition with acoustic and lexical features. In *ICASSP*, pages 4749–4753, 2015