

BOWEN ZHANG

E-mail: zbwglory@gmail.com, Homepage: <http://zbwglory.github.io>
MCB 407, 1002 West Childs Way
University of Southern California

EDUCATION

University of Southern California

Sept. 2017 - Present

Ph.D candidate in Computer Science, Advisor: Prof. Fei Sha.

Tongji University

Sept. 2014 - Mar. 2017

M.Eng. in Computer Technology, Advisor: Prof. Hanli Wang.

Tongji University

Sept. 2010 - Jul. 2014

B.Eng. in Computer Science and Technology, Advisor: Prof. Hanli Wang.
Minor in Mathematics.

Best Undergraduate Thesis of Tongji University.

RESEARCH EXPERIENCE

Google Brain

May. 2021 -

Host: Jiahui Yu and Andrew M. Dai

Student Researcher & Research Intern

- Action recognition with co-training on videos and images. [Arxiv]

Google Research Language

Student Researcher & Research Intern

- *Host: Pete Shaw*

Feb. 2021 - May 2021

Visually grounded compositional generalization. [Findings of EMNLP'21]

- *Hosts: Vihan Jain and Eugene Ie*

Feb. 2020 - Feb. 2021

Language guided clip localization from a large video corpus. [ArXiv]

Learning from structured image and language data. [EMNLP'20 Oral]

Amazon AWS Rekognition

May 2019 - Dec. 2019

Hosts: Yuanjun Xiong, Hao Chen, and Meng Wang

Intern

- Online video understanding system for streaming videos. [ArXiv]

Tencent AI Lab

Jun. 2018 - Aug. 2018

Host: Boqing Gong

Intern

- Model-based reinforcement learning.

University of Southern California

Sept. 2017 -

Theoretical & Empirical Data Science Laboratory, Advisor: Prof. Fei Sha

Research Assistant

- Learning from vision and language.

Learning visual and linguistic representation from video and language. [ECCV'18]

Learning word embedding from image and text. [EMNLP'18]

Visual Storytelling generation. [ICCV'19 Workshop]

Topic based text summarization. [BayLearn'19]

· Large Scale Action Recognition

Real-time action recognition. [CVPR'16, TIP]

ActivityNet Challenge [CVPR'16 Workshop, CVPR'17 Workshop]

PREPRINT

- 1 **Bowen Zhang**, Jiahui Yu, Christopher Fifty, Wei Han, Andrew M. Dai, Ruoming Pang, and Fei Sha, Co-training Transformer with Videos and Images Improves Action Recognition, 2021
- 2 **Bowen Zhang**^{*}, Hexiang Hu^{*}, Joonseok Lee, Ming Zhao, Sheide Chammas, Vihan Jain, Eugene Ie, and Fei Sha, A Hierarchical Multi-Modal Encoder for Moment Localization in Video Corpus, 2020
- 3 **Bowen Zhang**, Hao Chen, Meng Wang, and Yuanjun Xiong, Online Action Detection in Streaming Videos with Time Buffers, 2020
- 4 Melissa Ailem, **Bowen Zhang**, and Fei Sha, Topic Augmented Generator for Abstractive Summarization, 2019 (Short version is on BayLearn 2019)

PUBLICATIONS

- 1 **Bowen Zhang**, Hexiang Hu, Linlu Qiu, Pete Shaw, and Fei Sha, Visually Grounded Concept Composition, *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2021, Findings, Long Paper
- 2 Linlu Qiu, Hexiang Hu, **Bowen Zhang**, Pete Shaw, and Fei Sha, Systematic Generalization on gSCAN: What is Nearly Solved and What is Next?, *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2021. Short paper, *Oral Presentation*
- 3 **Bowen Zhang**^{*}, Hexiang Hu^{*}, Vihan Jain, Eugene Ie, and Fei Sha, Learning to Represent Images and Texts with Denotation Graphs, *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020, Long paper, *Oral Presentation*
- 4 Melissa Ailem, **Bowen Zhang**, Aurélien Bellet, Pascal Denis, and Fei Sha, A Probabilistic Model for Joint Learning of Word Embeddings from Texts and Images, *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2018, Long paper
- 5 **Bowen Zhang**^{*}, Hexiang Hu^{*}, and Fei Sha, Cross-Modal and Hierarchical Modeling of Video and Text, *European Conference on Computer Vision (ECCV)*, 2018
- 6 **Bowen Zhang**, Limin Wang, Zhe Wang, Yu Qiao, and Hanli Wang, Real-time Action Recognition with Enhanced Motion Vector CNNs, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, 2016.
- 7 **Bowen Zhang**, Limin Wang, Zhe Wang, Yu Qiao, and Hanli Wang, Real-Time Action Recognition with Deeply-Transferred Motion Vector CNNs, *IEEE Transactions on Image Processing (TIP)*, 2018
- 8 Zhe Wang, Limin Wang, Yali Wang, **Bowen Zhang**, and Yu Qiao, Weakly supervised patchnets: Describing and aggregating local patches for scene recognition, *IEEE Transactions on Image Processing (TIP)*, 2017
- 9 **Bowen Zhang**, and Hanli Wang, Encoding Scale into Fisher Vector for Human Action Recognition, *IEEE Conference on Visual Communications and Image Processing (VCIP)*, Singapore, 2015, *Oral Presentation*

WORKSHOPS

- 1 **Bowen Zhang**, Hexiang Hu, and Fei Sha, Visual Storytelling via Predicting Anchor Word Embeddings in the Stories, *ICCV 2019 Workshop on Closing the Loop Between Vision and Language*, 2019

- 2 Yue Zhao, **Bowen Zhang**, Zhirong Wu, Shuo Yang, Lei Zhou, Sijie Yan, Limin Wang, Yuanjun Xiong, Dahua Lin, Yu Qiao, and Xiaoou Tang, CUHK & ETHZ & SIAT Submission to ActivityNet Challenge 2017, *ActivityNet Large Scale Activity Recognition Challenge, CVPR'17 Workshop*, Hawaii, USA, 2017.
- 3 Yuanjun Xiong, Limin Wang, Zhe Wang, **Bowen Zhang**, Hang Song, Wei Li, Dahua Lin, Yu Qiao, Luc Van Gool, and Xiaoou Tang, CUHK & ETHZ & SIAT Submission to ActivityNet Challenge 2016, *ActivityNet Large Scale Activity Recognition Challenge, CVPR'16 Workshop*, Las Vegas, USA, 2016.
- 4 Yun Yi, Hanli Wang, **Bowen Zhang**, and Jian Yu, MIC-TJU in MediaEval 2015 Affective Impact of Movies Task, *MediaEval 2015 Workshop*, Wurzen, Germany, 2015.
- 5 **Bowen Zhang**, Yun Yi, Hanli Wang, and Jian Yu, MIC-TJU at MediaEval Violent Scenes Detection (VSD) 2014, *MediaEval 2014 Workshop*, Barcelona, Spain, 2014.

CONTESTS

| | |
|--|-------------------|
| The ActivityNet Challenge 2017 on CVPR'17: Untrimmed Classification. <i>Rank</i> 2/37. | <i>Jul. 2017</i> |
| The ActivityNet Challenge 2017 on CVPR'17: Trimmed Classification. <i>Rank</i> 2/27. | <i>Jul. 2017</i> |
| The ActivityNet Challenge 2017 on CVPR'17: Action Localization. <i>Rank</i> 2/15. | <i>Jul. 2017</i> |
| The ActivityNet Challenge 2016 on CVPR'16: Untrimmed Video Classification. <i>Rank</i> 1/25. | <i>Jul. 2016</i> |
| LSUN Challenge 2016 on CVPR'16: Scene Classification. <i>Rank</i> 1 st . | <i>Jul. 2016</i> |
| MediaEval'15 Affective Impact. Violence: <i>Rank</i> : 2 nd ; Valence: <i>Rank</i> : 2 nd ; Arousal: <i>Rank</i> : 1 st . | <i>Sept. 2015</i> |
| MediaEval'14 Affect in Multimedia. Main: <i>Rank</i> : 4 th ; Generalization: <i>Rank</i> : 4 th . | <i>Sept. 2014</i> |

SELECTED HONORS AND AWARDS

| | |
|--|-------------|
| Shanghai Outstanding Graduates | <i>2017</i> |
| National Graduate Scholarship of China | <i>2016</i> |
| Best Undergraduate Thesis of Tongji University (2 recipients in CS department) | <i>2014</i> |

ACADEMIC SERVICE

Reviewer for CVPR, ECCV, ICCV, NAACL, ACL, EMNLP, NeurIPS, ICML, ICLR, AAAI, IJCV, T-PAMI