

# Luowei Zhou

1301 Beal Ave., EECS Building 4338, Ann Arbor, MI 48105, USA  
[luoweizhou.github.io](https://luoweizhou.github.io) | [luozhou@umich.edu](mailto:luozhou@umich.edu) | (734)757-0923

## RESEARCH INTERESTS

---

Computer vision and its relations to natural language and deep learning, with a focus on problems in video understanding such as video captioning, object grounding, question answering, retrieval, activity recognition, and multi-modal unsupervised representation learning.

## EDUCATION

---

### University of Michigan

Ann Arbor, Michigan, USA

*Ph.D. program in Robotics (Computer Vision)*

*Sept. 2015 – March 2020 (expected)*

- **Courses:** Advanced Computer Vision, Natural Language Processing, Machine Learning, Optimization
- **Academics:** Curriculum GPA: **4.00/4.00**

### Nanjing University

Nanjing, Jiangsu, China

*Bachelor of Engineering in Automation*

*Sept. 2011 – Jun. 2015*

- **Courses:** Computer Vision, Artificial Intelligence, Advanced Programming Language, Data Structure
- **Academics:** Overall GPA: **91.8/100**, Major GPA: **93.0/100**

## SELECTED PUBLICATIONS (see all at [Google Scholar](#))

---

**L. Zhou**, Y. Kalantidis, X. Chen, J. J. Corso, and M. Rohrbach, “Grounded Video Description”, CVPR 2019. (oral) AR: 5.6%; h5: 188

**L. Zhou**, Y. Zhou, J. J. Corso, R. Socher, and C. Xiong, “End-to-End Dense Video Captioning with Masked Transformer”, CVPR 2018. (spotlight) AR: 9%; h5: 158

**L. Zhou**, C. Xu, and J. J. Corso, “Towards Automatic Learning of Procedures from Web Instructional Videos”, AAAI 2018. (oral) AR: 11%; h5: 56

H. Huang, **L. Zhou**, W. Zhang, J. J. Corso, and C. Xu, “Dynamic Graph Modules for Modeling Object-Object Interactions in Activity Recognition”, BMVC 2019. AR: 30%; h5: 42

**L. Zhou**, N. Louis, and J. J. Corso, “Weakly-Supervised Video Object Grounding from Text by Loss Weighting and Object Interaction”, BMVC 2018. AR: 30%; h5: 42

**L. Zhou**, P. Yang, C. Chen, and Y. Gao, “Multi-agent Reinforcement Learning with Sparse Interactions by Negotiation and Knowledge Transfer”, IEEE Transactions on Cybernetics 2017, 47 (5): 1238 - 1250. SCI IF: 7.38; h5: 73

**L. Zhou**, C. Xu, P. Koch, and J. J. Corso, “Watch What You Just Said: Image Captioning with Text-Conditional Attention”, ACM Multimedia (Thematic Workshops) 2017: 305-313. (pitch)

**L. Zhou**, P. Yang and C. Chen, “Multi-agent Reinforcement Learning with Sparse Interactions by Negotiation and Knowledge Transfer”, IJCAI (Workshops) 2016. (oral)

**L. Zhou** et al., “A Balanced Heuristic Mechanism for Multi-robot Task Allocation of Intelligent Warehouses”, Mathematical Problems in Engineering 2014: 1–10. SCI IF: 0.80; h5: 39

## WORK EXPERIENCE

---

<b>Microsoft Research (MSR)</b> <i>Research Intern with Hamid Palangi, Lei Zhang, and Jianfeng Gao</i>	Redmond, WA, USA May 2019 – Aug. 2019
<b>Facebook AI Research (FAIR)</b> <i>Research Intern with Yannis Kalantidis, Xinlei Chen, and Marcus Rohrbach</i>	Menlo Park, CA, USA May 2018 – Aug. 2018
<b>Salesforce Research (Metamind)</b> <i>Deep Learning Research Intern with Caiming Xiong and Richard Socher</i>	Palo Alto, CA, USA May 2017 – Aug. 2017
<b>University of Michigan, EECS</b> <i>Graduate Student Research Assistant (GSRA) with Jason Corso</i> <i>Graduate Student Instructor (GSI) with Justin Johnson's Deep Vision Class</i>	Ann Arbor, MI, USA May 2016 – present Sept. 2019 – present

## INVITED TALKS

---

<b>NVIDIA AI Lab</b> <i>Hosted by Dr. Sanja Fidler</i>	Toronto, Ontario, Canada Dec. 2018
<b>SAMSUNG AI Centre</b> <i>Hosted by Dr. Afsaneh Fazly and Dr. Allan Jepson</i>	Toronto, Ontario, Canada Dec. 2018
<b>ICML How2 Workshop</b> <i>Pitch presentation hosted by Dr. Florian Metze</i>	Long Beach, CA, USA June 2019

## PROFESSIONAL ACTIVITIES

---

*Co-organizer*, CVPR 2018 Workshop on Fine-grained Instructional Video Understanding (FIVER), with Jason Corso, Josef Sivic, and Ivan Laptev  
*Co-organizer*, UMich Computer Vision Reading Group  
*Program Committee Member / Reviewer*: CVPR 2019/2018, ICCV 2019, TPAMI 2019/2018, IJCV 2019, AAAI 2019, NIPS 2016, CVIU 2017, ACM MM 2019, ICRA 2018, ITS 2019/2018/2017 etc.  
*Volunteer*, RSS 2016

## HONORS AND AWARDS

---

<i>Outstanding Winner Awards (0.2%)</i> , Mathematical Contest in Modeling (MCM)	2013
<i>Sienhua New and Tsu Way Shen Memorial Award (Top 1)</i> , of University of Michigan	2015
<i>Best Undergrad Thesis (Top 1)</i> , of Jiangsu Province	2015
<i>National Scholarship (1%)</i> , of Nanjing University	2012
<i>Red Sun Scholarship</i> , of Nanjing University	2014

## RESEARCH EXPERIENCE (open-source projects on [Github](#))

---

### Large-Scale Unified Vision-Language Pre-training

Microsoft Research

*Supervisors: Dr. Jianfeng Gao, Dr. Lei Zhang, and Dr. Hamid Palangi*

*May 2019 – present*

- Introduced a generic and unified framework for Vision-Language Pre-training (VLP). VLP is pre-trained on millions of image-text pairs automatically mined from the web and fine-tuned for disparate downstream tasks including image captioning and VQA.
- Proposed to use two unsupervised learning objectives for VLP: bidirectional and sequence-to-sequence (seq2seq) masked vision-language prediction.
- Thanks to our vision-language pre-training, both training speed and overall accuracy have been significantly improved on the downstream tasks compared to other model initialization methods.
- Set new SotA on COCO Captions (CIDEr 129), VQA 2.0 (overall 71) and Flickr30k Captions (CIDEr 67 vs previous SotA 62), all from a single model architecture.
- Current focuses: VLP on videos by leveraging a large amount of instructional video data and the associated ASR scripts. Multi-task learning of captioning, QA, and event proposal.

### Grounded Video Description

Facebook AI Research

*Supervisors: Dr. Marcus Rohrbach, Dr. Yannis Kalantidis, and Dr. Xinlei Chen*

*May 2018 – Dec. 2018*

- Introduced a large-scale video description and grounding dataset, called [ActivityNet-Entities](#), where we annotated noun phrases (& objects) from sentence descriptions in videos as spatial bounding boxes. ActivityNet-Entities contains over 158k labeled boxes for 52k video clips.
- Proposed a unified framework for video and image description, where a supervised grounding module dynamically detects objects in the scene and provides visual clues to the captioning module.
- Set new SotA performance on video description and image description and demonstrated that our generated sentences are more explainable through grounding.

### Fine-grained Instructional Video Understanding

University of Michigan

*Supervisor: Prof. Jason Corso*

*Sept. 2016 – present*

- Introduced [YouCook2](#) dataset, which contains temporally localized recipe sentence annotations and bounding boxes for 2000 YouTube cooking videos.
- Tackled a series of problems related to instructional video understanding: i) event proposal (AAAI 2018), ii) dense video captioning (CVPR 2018), iii) weakly supervised object grounding from language description (BMVC 2018).
- *Event proposal*: Proposed an event proposal and sequential modeling network that can temporally localize procedure steps in web instructional videos and capture the temporal structure of the video.
- *Dense video captioning*: Caption generation for event proposals. See Page 4 for more details.
- *Weakly supervised object grounding*: Given a video and the corresponding description, localize the objects mentioned from the description in the video as bounding boxes. No box is given for training.
- Current focuses: Graph-based procedure structure learning.

### **Dense-Captioning Events in Video and Temporal Action Proposal**

Salesforce Research

*Supervisors: Dr. Caiming Xiong and Dr. Richard Socher*

*May 2017 – Aug. 2017*

- Introduced a self-attention-based video captioning model and improved our previously proposed action/event proposal network with carefully-designed Temporal Convolutional Networks.
- Proposed to bridge event proposal and captioning by a differentiable visual mask and achieved state-of-the-art results on dense video captioning.

### **Text-conditional Visual Captioning with Guiding LSTM**

University of Michigan

*Supervisor: Prof. Jason Corso*

*Mar. 2016 – Nov. 2016*

- Proposed an encoder-decoder image captioner though explicit text-conditional image guidance.
- Extended the work to video captioning by leveraging audio features for the extra guidance.

### **End-to-End Grasping with Deep Reinforcement Learning**

University of Michigan

*Supervisor: Prof. Satinder Singh*

*Sept. 2015 – Apr. 2016*

- Applied state-of-the-art Deep RL algorithm named Deep Q-network (DQN) to robot grasping tasks.
- Built an API between physics engine MuJoCo and the DQN module.

### **Research on Multi-Agent Reinforcement Learning with Sparse Interactions**

Nanjing University

*Supervisors: Prof. Chunlin Chen, Dr. Pei Yang, and Prof. Yang Gao*

*Dec. 2014 – Jul. 2015*

- Introduced the concept of equilibrium into traditional sparse-interaction-based MARL algorithms and proposed a knowledge transfer approach to initialize the joint-state Q table.
- Applied the proposed algorithm in a real-world setting, i.e., our intelligent warehouse simulator.

### **Multi-Robot Task Allocation and Path Planning in Dynamic Environments**

Nanjing University

*Supervisor: Dr. Pei Yang*

*Nov. 2013 – Jul. 2014*

- Proposed a Balanced Heuristic Mechanism to balance task allocation in multi-robot systems.
- Built an intelligent warehouse simulator from scratch using C/OpenGL for the experiments.

## **PROFICIENCY AND SKILLS**

---

*Technical Skills:* PyTorch/Torch, Python, C/C++, Linux, Git, LaTeX, Matlab, Caffe, HTML, CSS, JS etc.

*Languages:* English (proficient) and Mandarin (native)

## **REFERENCES**

---

**Prof. Jason Corso**, Professor, University of Michigan, [jicorso@umich.edu](mailto:jicorso@umich.edu)

**Prof. Chenliang Xu**, Assistant Professor, University of Rochester, [chenliang.xu@rochester.edu](mailto:chenliang.xu@rochester.edu)

**Dr. Jianfeng Gao**, Partner Research Manager, Microsoft Research, [jfgao@microsoft.com](mailto:jfgao@microsoft.com)

**Dr. Lei Zhang**, Principle Research Manager, Microsoft Research, [leizhang@microsoft.com](mailto:leizhang@microsoft.com)

**Dr. Caiming Xiong**, Senior Director, Salesforce Research, [cxiong@salesforce.com](mailto:cxiong@salesforce.com)

**Dr. Marcus Rohrbach**, Research Scientist, Facebook AI, [mrf@fb.com](mailto:mrf@fb.com)

**Dr. Yannis Kalantidis**, Research Scientist, Facebook AI, [yannisk@fb.com](mailto:yannisk@fb.com)