

Luwei Zhou

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

EDUCATION

University of Michigan

Ann Arbor, Michigan, USA

Ph.D. program in Robotics (Computer Vision)

Sept. 2015 – May 2020 (expected)

- **Research Interests:** Vision and language, video understanding, deep learning
- **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, et al., “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, Y. Shi, J. Wang, and P. Yang, “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

Facebook AI Research (FAIR)

Menlo Park, CA, USA

Research Intern with Marcus Rohrbach, Yannis Kalantidis, and Xinlei Chen

May 2018 – Aug. 2018

Microsoft Research

Redmond, WA, USA

Research Intern with Hamid Palangi, Lei Zhang, and Jianfeng Gao

May 2019 – Aug. 2019

Salesforce Research (Metamind)

Palo Alto, CA, USA

Deep Learning Research Intern with Caiming Xiong and Richard Socher

May 2017 – Aug. 2017

University of Michigan, EECS

Ann Arbor, MI, USA

Graduate Student Instructor on Justin Johnson’s Deep Vision class

Sept. 2019 – present

INVITED TALKS

NVIDIA AI Lab

Hosted by Dr. Sanja Fidler

Toronto, Ontario, Canada

Dec. 2018

SAMSUNG AI Centre

Hosted by Dr. Afsaneh Fazly and Dr. Allan Jepson

Toronto, Ontario, Canada

Dec. 2018

PROFESSIONAL ACTIVITIES

Co-organizer, CVPR 2018 Workshop on Fine-grained Instructional Video Understanding (FIVER), with Jason Corso, Josef Sivic and Ivan Laptev

Program Committee Member / Reviewer: CVPR, TPAMI, IJCV, AAAI, NIPS, ACM MM, ITS, ICRA

Volunteer, RSS 2016

HONORS AND AWARDS

Outstanding Winner Awards (0.2%), Mathematical Contest in Modeling (MCM) 2013

Sienhua New and Tsu Way Shen Memorial Award (Top 1), of University of Michigan 2015

Best Undergrad Thesis (Top 1), of Jiangsu Province 2015

National Scholarship (1%), of Nanjing University 2012

Red Sun Scholarship, of Nanjing University 2014

RESEARCH EXPERIENCE

Grounded Video Description

Facebook AI Research

Supervisors: Dr. Marcus Rohrbach, Dr. Yannis Kalantidis, and Dr. Xinlei Chen May 2018 – present

- 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 labelled boxes for 52k video clips.
- Proposed a unified framework for video and image description, where a grounding module dynamically detects objects in the scene and provides visual clues to the captioning module.
- Set new state-of-the-art 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 3 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: Action-conditioned object grounding, modeling of object state transitions, multi-task learning of object grounding, action recognition, and captioning on videos.

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, Dr. 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 model evaluation.

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, Associate Professor, University of Michigan, jjcorso@umich.edu

Prof. Chenliang Xu, Assistant Professor, University of Rochester, chenliang.xu@rochester.edu

Dr. Jianfeng Gao, Partner Research Manager, Microsoft Research, jfgao@microsoft.com

Dr. Lei Zhang, Principle Research Manager, Microsoft Research, leizhang@microsoft.com

Dr. Marcus Rohrbach, Research Scientist, Facebook AI, mrf@fb.com

Dr. Yannis Kalantidis, Research Scientist, Facebook AI, yannisk@fb.com