Fanyi Xiao

Contact Information

2078 Academic Surge
University of California, Davis
Davis, CA 95616, USA

WWW: fanyix.cs.ucdavis.edu
Mobile: 949-491-2293
E-mail: fyxiao@ucdavis.edu

Research Interests

I am broadly interested in computer vision and machine learning, with a focus on developing models to learn from videos. At a high level, my research interests can be categorized into three aspects: **(i)** Learning to understand videos. This includes developing algorithms for video recognition/detection/segmentation. **(ii)** Learning video representations with minimal human supervision (weakly and self-supervised learning from videos). **(iii)** Learning across modalities (e.g., image, video, language and audio).

Education

University of California Davis, Computer Science Dept., Davis, CA, USA

Ph.D., Computer Science, 2020

- Advisor: Prof. Yong Jae Lee
- Thesis: *Multimodal Learning with Minimal Human Supervision from Videos and Natural Language*

Carnegie Mellon University, Robotics Institute, Pittsburgh, PA, USA

M.S., Robotics, 2014

- Advisors: Prof. Martial Hebert and Prof. Yaser Sheikh
- Thesis: *Model Recommendation for Large Scale Exemplar-based Object Detection*

Central South University, Computer Science Department, Changsha, China

B.S., Computer Science, 2012

- Thesis: *Facial Expression Analysis with Active Appearance Model*

Publications

[1] Haotian Liu, Rafael A. Rivera-Soto, **Fanyi Xiao**, and Yong Jae Lee. YolactEdge: Real-time Instance Segmentation on the Edge. In *International Conference on Robotics and Automation (ICRA)*, 2021.

[2] Xueyan Zou, **Fanyi Xiao**, Zhiding Yu, and Yong Jae Lee. Delving Deeper into Anti-aliasing in ConvNets. In *British Machine Vision Conference (BMVC)*, 2020. Best Paper Award.

[3] **Fanyi Xiao**, Yong Jae Lee, Kristen Grauman, Jitendra Malik, and Christoph Feichtenhofer. Audiovisual SlowFast Networks for Video Recognition. *arXiv preprint arXiv:2001.08740*, 2020.

[4] Daniel Bolya, Chong Zhou, **Fanyi Xiao**, and Yong Jae Lee. YOLACT++: Better Real-time Instance Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 2020.

[5] **Fanyi Xiao**, Haotian Liu, and Yong Jae Lee. Identity from here, Pose from there: Self-supervised Disentanglement and Generation of Objects using Unlabeled Videos. In *International Conference on Computer Vision (ICCV)*, 2019.

[6] Daniel Bolya, Chong Zhou, **Fanyi Xiao**, and Yong Jae Lee. YOLACT: Real-time Instance Segmentation. In *International Conference on Computer Vision (ICCV)*, 2019. **Oral presentation**.

[7] Xitong Yang, Xiaodong Yang, Ming-Yu Liu, **Fanyi Xiao**, Larry S Davis, and Jan Kautz. STEP: Spatio-Temporal Progressive Learning for Video Action Detection. In *Computer Vision and Pattern Recognition (CVPR)*, 2019.

[8] **Fanyi Xiao** and Yong Jae Lee. Video Object Detection with an Aligned Spatial-Temporal Memory. In *European Conference on Computer Vision (ECCV)*, 2018.

[9] Wenjian Hu, Krishna Kumar Singh*, **Fanyi Xiao***, Jinyoung Han, Chen-Nee Chuah, and Yong Jae Lee (* equal contribution). Who Will Share My Image? Predicting the Content Diffusion Path in Online Social Networks. In *ACM International Conference on Web Search and Data Mining (WSDM)*, 2018.

[10] **Fanyi Xiao**, Leonid Sigal, and Yong Jae Lee. Weakly-supervised Visual Grounding of Phrases with Linguistic Structures. In *Computer Vision and Pattern Recognition (CVPR)*, 2017.

[11] **Fanyi Xiao** and Yong Jae Lee. Track and Segment: An Iterative Unsupervised Approach for Video Object Proposals. In *Computer Vision and Pattern Recognition (CVPR)*, 2016. **Spotlight presentation**.

[12] Krishna Singh, **Fanyi Xiao**, and Yong Jae Lee. Track and Transfer: Watching Videos to Simulate Strong Human Supervision for Weakly-Supervised Object Detection. In *Computer Vision and Pattern Recognition (CVPR)*, 2016.

[13] **Fanyi Xiao** and Yong Jae Lee. Discovering the Spatial Extent of Relative Attributes. In *International Conference on Computer Vision (ICCV)*, 2015. **Oral presentation**.

[14] **Fanyi Xiao** and Martial Hebert. Efficient Model Evaluation with Bilinear Separation Model. In *Winter Conference on Applications of Computer Vision (WACV)*, 2015.

[15] Zhiding Yu, Chunjing Xu, Deyu Meng, **Fanyi Xiao**, Wenbo Liu, and Jianzhuang Liu. Transitive Distance Clustering with K-Means Duality. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[16] Iljoo Baek, Taylor Stine, Denver Dash, **Fanyi Xiao**, Yaser Ajmal Sheikh, Yair Movshovitz-Attias, Mei Chen, Martial Hebert, and Takeo Kanade. Physical Querying with Multi-Modal Sensing. In *Winter Conference on Applications of Computer Vision (WACV)*, 2014.

Awards

- Best Paper Award, British Machine Vision Conference (BMVC), 2020
- Outstanding Reviewer, European Conference on Computer Vision (ECCV), 2020
- Most Innovative Award, COCO Object Detection Challenge, 2019
- Google Cloud Platform Research Grant, 2019
- Best Graduate Researcher Award, CS Dept. of UC Davis, 2018
- Azure Research Award, Microsoft, 2017
- Graduate Fellowship, UC Davis, 2015
- AWS Research Grant, Amazon Web Services, Inc., 2015
- Excellent Undergraduate Thesis, CSU, 2012
- Top Grade Scholarship (University-wide highest honor, 0.8%), CSU, 2010
- Sunward Scholarship (0.4%), Sunward Corporation, 2010
- 1st Grade Scholarship (6%), CSU, 2009
- National Scholarship (1%), Ministry of Education of China, 2009

Experience

Amazon AI, Seattle, WA

Applied Scientist II **Aug 2020 - Present**

- Conduct cutting-edge research on self-supervised video representation learning.
- Build computer vision solution to understand Media & Entertainment content.

Facebook AI Research (FAIR), Menlo Park, CA

Research Intern **June 2019 - Nov 2019**

- An audiovisual video understanding architecture for recognition, detection and multi-modal self-supervised video representation learning.

NVIDIA Research, Santa Clara, CA

Research Intern **July 2017 - Oct 2017**

- An iterative action tube detection method for action detection.

Disney Research, Pittsburgh, PA

Research Intern **June 2016 - Sept 2016**

- Weakly supervised vision-language alignment (i.e., producing segmentation masks for free-form language inputs) by exploiting linguistic structure.

Skills

- Programming: Python, C, Lua, Java, MATLAB
- Misc: PyTorch, Torch7, Caffe, Caffe2, LINUX, LaTeX

Related Graduate Courses

- CMU: Computer Vision / Machine Learning / Convex Optimization
  Math Fundamentals for Robotics / Learning-based Methods in Vision
  Mechanics of Manipulation
- UC Davis: Visual Recognition

Mentoring

- Daniel Bolya, undergraduate student (now PhD at Georgia Tech)
- Haotian Liu, undergraduate student (now PhD at UC Davis)
- Chong Zhou, MS student (now PhD at UNC Chapel Hill)
- Xueyan Zou, junior PhD student at UC Davis

Service

- Reviewer, Computer Vision and Pattern Recognition (CVPR), 2018-21
- Reviewer, International Conference on Computer Vision (ICCV), 2019
- Reviewer, European Conference on Computer Vision (ECCV), 2020
- Reviewer, International Conference on Machine Learning (ICML), 2021
- Reviewer, Neural Information Processing Systems (NeurIPS), 2020
- Program Committee, AAAI Conference on Artificial Intelligence (AAAI), 2019-20
- Program Committee, Intl. Joint Conference on Artificial Intelligence (IJCAI), 2020
- Reviewer, The British Machine Vision Conference (BMVC), 2020
- Reviewer, Asian Conference on Computer Vision (ACCV), 2018
- Reviewer, Winter Conference on Applications of Computer Vision (WACV), 2015-18
- Reviewer, IEEE Transactions on Multimedia (TMM)