Yixuan Li

CONTACT INFORMATION
163 Xianlin Road, Qixia District
Nanjing, Jiangsu, P.R. China, 210023
(+86) 18120190595
yixuanli@smail.nju.edu.cn

HOMEPAGE
https://yixuanli98.github.io/

RESEARCH INTERESTS
Computer Vision: video understanding, action recognition and detection
Machine Learning: representation learning, deep learning

EDUCATION

Department of Computer Science and Technology, Nanjing University — Nanjing, China
M.Sc. Candidate in MCG Lab — August 2019 – Present
Supervisor: Prof. Limin Wang

Kuang Yaming Honors School, Nanjing University — Nanjing, China
B.Sc., Major in Computer Science (GPA: 86.2/100) — August 2015 – June 2019
Supervisor: Prof. Gangshan Wu

PUBLICATION
Yixuan Li, Lei Chen, Runyu He, Zhenzhi Wang, Gangshan Wu, Limin Wang. MultiSports: A Multi-Person Video Dataset of Spatio-Temporally Localized Sports Actions. International Conference on Computer Vision (ICCV'21), 2021.
Yixuan Li*, Zixu Wang*, Limin Wang, Gangshan Wu. Actions as Moving Points. European Conference on Computer Vision (ECCV'20), Glasgow, United Kingdom, 2020.

HONORS AND AWARDS

- **National Scholarship** (9/231) — Ministry of Education — 2021
- **1st Place, HC-STVG Challenge (CVPR2021 PIC Workshop)** — Workshop Committee — 2021
- **National Scholarship** (2/231) — Ministry of Education — 2020
- **Outstanding Graduate Students, awardee** (2/231) — Nanjing University — 2020
- **Final Round, Google Girl Hackathon 2020** (18/94) — Google — 2020
- **1st Award, Scholarship for Graduate Students** (20%) — Nanjing University — 2019&2020
- **2rd Award, People's Scholarship** — Nanjing University — 2018
- **3rd Place, ROBOMASTER 2017 (Eastern Division)** — Da-Jiang Innovations — 2017
- **3rd Award, Academic Excellent Scholarship** — Nanjing University — 2016
- **3rd Award, People's Scholarship** — Nanjing University — 2016
- **Excellent League Member of Kuang Yaming Honors School** — Nanjing University — 2015

RESEARCH EXPERIENCE

MultiSports: A Multi-Person Video Dataset of Spatio-Temporally Localized Sports Actions.
Advisor: Prof. Limin Wang — Aug. 2020 - Mar. 2021
HomePage: MultiSports Dataset

- As the first author, presented a large-scale, fine-grained, multi-person, and untrimmed spatio-temporal action detection dataset with well-defined temporal boundaries, *MultiSports*. Besides, adapted several representative methods to it and gave in-depth analysis to inspire new advances in this field.
- *MultiSports* contained 66 fine-grained action categories from 4 different sports, where we collected 3200 video clips and annotated around 37701 action instances with 902k bounding boxes. Our datasets had more fine-grained action categories (66 vs. 21 or 24), more instances per video clip (11.8 vs. 1 or 1.4), and much more instances (37701 vs. 928 or 4458) than the existing datasets JHMDB and UCF101-24.
- Existing methods achieved satisfactory performance on JHMDB and UCF101-24 but obtained low performance on *MultiSports* (video-mAP@0.2 of 77.3 or 82.8 vs. 12.88 for MOC).

Actions as Moving Points
Advisor: Prof. Limin Wang — Jul. 2019 - Mar. 2020

- As the first author, presented an conceptually simple, computationally efficient, and more precise spatio-temporal action detection framework, MOC-detector, which would recognize all the action instances present in a video and localize them in both space and time.
- MOC outperformed the existing state-of-the-art methods under the same setting on the JHMDB and UCF101-24 datasets. The code is available at https://github.com/MCG-NJU/MOC-Detector.
- MOC could handle online real-time video stream and reach 53 fps with only RGB as input.

MR2Flow: Efficient Motion Representations for Real-time Video Recognition

Advisor: Prof. Limin Wang   Sep. 2018 - Apr. 2019

- As the first author, presented an efficient motion representation by enhancing the discriminative power of motion vector for real-time video recognition, termed as MR2Flow.
- The whole pipeline achieved 94.0% with 100 fps on UCF101 dataset, where the accuracy rate of previous method was 95.8% with 12 fps.

CONTEST EXPERIENCE

Human-centric Spatio-Temporal Video Grounding Challenge.   May. 2021 - Jun. 2021

In CVPR2021 Workshop Person in Context.

- We got the **1st place**. First, we extracted tube-level features by SlowFast and CSN on linked tubes based on person boxes predicted by Faster R-CNN. Then we used a 2d-map proposal representation like 2D-TAN and enhanced the feature representation to be more discriminative by multi-modal contrastive learning.
- Contribution: I generated the person boxes for a single frame, then linked the boxes into the tubes, and finally extracted the visual features of the tubes.

INTERNSHIP

Fine-grained Spatio-Temporal Action Detection

Tencent Data Platform, Shenzhen. Advisor: Dr. Zhifeng Li   Jun. 2021 - Present

- As the first author, presented an end-to-end sparse action detection framework for new datasets MultiSports. The network took the video as input and directly output the video-level detection results, which removed the heuristic linking algorithm and explicitly estimated temporal boundaries. This paper will be submitted to a top conference of computer vision area.
- I participated in the construction of a computer vision codebase. I was responsible for organizing the estimation codes of six existing fields (e.g. image/video classification, object detection, spatio-temporal action detection) into a unified framework, and writing README and test samples.

ACADEMIC SERIVCE

- Track organizer of ICCV2021 Workshop DeeperAction on localized-and-detailed understanding of human actions in videos. Our track, MultiSports, focused on localizing all action instances with spatio-temporal tubes and recognizing their labels from untrimmed and multi-person videos.
- Reviewer of IEEE Transactions on Circuits and Systems for Video Technology (T-CSVT) and Pattern Recognition (PR).

SKILLS

- Programming: Python, PyTorch, Matlab, C, Latex,
- Languages: Mandarin, English