Yiming Dou

✉ yiming.dou.1@gmail.com | 🌐 dou-yiming.github.io | ⓞ Dou-Yiming

Education

Stanford University Mar. 2022 – Present
Visiting Research Intern California, U.S.A.

Shanghai Jiao Tong University (SJTU) Sept. 2019 – June 2023 (Expected)
B.Eng in Computer Science and Technology
B.Ec in Economics (Associate Degree) Shanghai, China
- Member of Zhiyuan Honors Program (Selected top 5% students)
- GPA 88.47/100 (Major: 92.26/100)
- Highlights of A+ Courses: Artificial Intelligence, Computer Graphics, Natural Language Processing, Mathematical Foundations of Computer Science, and Project Workshop of Operating System.

Research Interests

Computer Vision: human activity understanding, object-centric learning, video understanding

Multimodal: multisensory (vision, audio, touch) learning, vision-language

Robotics: multisensory robotics

Research Experience

Stanford Vision and Learning Lab, Stanford University Mar. 2022 – Present
Visiting Research Intern, supervised by **Prof. Jiajun Wu** and **Prof. Fei-Fei Li**. California, U.S.A.
- Multisensory Object-Centric Learning
    - Construct the ObjectFolder Real dataset, including multisensory (vision, audio, and touch) measurements for 100 real-world household objects, building upon a newly designed pipeline for collecting the 3D meshes, videos, impact sounds, and tactile readings of real-world objects.
    - Introduce the ObjectFolder Benchmark, a benchmark suite containing 10 multisensory object-centric learning tasks that are centered around object recognition, reconstruction, and manipulation.
    - Standardize the task setting and evaluation metric for each task, and evaluate the performance of ours/existing baseline methods.

Machine Vision and Intelligence Group, Shanghai Jiao Tong University Oct. 2020 – Present
Undergraduate researcher, supervised by **Prof. Cewu Lu**. Shanghai, China
- Unified Semantic Space for Action Recognition
    - Design a Poincaré action semantic space in view of verb taxonomy hierarchy.
    - Gather various (image/video/skeleton/MoCap) datasets with different class granularities into a unified database in a principled label system by aligning the classes of different datasets to our semantic space.
    - Propose a bidirectional mapping model to conduct both action recognition and generation.
- Interactive-Action Driven Visual Grounding
    - Introduce Referring Scene Understanding (ReSU), a novel benchmark that requires the visual grounding model to localize not only objects in indoor scenes, but also explain the reason for choosing the object by localizing interactive human body parts based on referring language.
    - Propose a transformer-based baseline system to address ReSU and show impressive improvements upon state-of-the-art visual grounding methods.
- 2D-3D Spatial-Temporal Human Object Interaction
    - Introduce a new dataset based on AVA: Discovering Interacted Objects (DIO) including 51 interactions and 1,000+ objects.
    - Propose a Spatio-Temporal Human-Object Interaction benchmark, which expects vision systems to track human actions, detect interactions, and discover interacted objects.

Publications

(* indicates equal contribution)

1. Ruohan Gao*, **Yiming Dou***, Hao Li*, Tanmay Agarwal, Jeannette Bohg, Yunzhu Li, Li Fei-Fei and Jiajun Wu, "The ObjectFolder Benchmark: Multisensory Object-Centric Learning with Neural and Real Objects"
Under review, CVPR2023
Website

2. Yong-Lu Li*, Xiaoqian Wu*, Xinpeng Liu, **Yiming Dou**, Yikun Ji, Junyi Zhang, Yixing Li, Xudong Lu, Jingru Tan and Cewu Lu, "Bridging The Isolated Islands in Human Action Understanding"
Under review, CVPR2023

3. **Yiming Dou***, Yong-Lu Li* and Cewu Lu, "ReSU: A Novel Interactive-Action Driven Benchmark for Embodied Visual Grounding"
Under review, CVPR2023

4. Yong-Lu Li*, Hongwei Fan*, Zuoyu Qiu, **Yiming Dou**, Liang Xu, Hao-Shu Fang, Peiyang Guo, Haisheng Su, Dongliang Wang, Wei Wu and Cewu Lu, "Discovering A Variety of Objects in Spatio-Temporal Human-Object Interactions"
Technical report, 2022
Project page / arXiv

Honors and Awards

STJU Overseas Scholarship (two winners in SJTU), SJTU	2022
Zhanjiajun Scholarship (six winners in SJTU), SJTU	2022
Academic Excellence Scholarship (top 10%), SJTU	2022
Meritorious Winner (top 7%), The Mathematical Contest in Modeling	2022
Merit Student Award (top 5%), SJTU	2021
Honorable Mention (top 24%), The Mathematical Contest in Modeling	2021
Zhiyuan Honors Scholarship (top 5%), SJTU	2019, 2020, 2021, 2022

Skills

Programming Languages: C/C++, Python, JavaScript, Matlab, Verilog

Deep Learning Frameworks: PyTorch, TensorFlow

English Proficiency: TOEFL iBT 108