Jinghuan Shang

100 Nicolls Road, Stony Brook, NY 11794, USA
jishang [at] cs.stonybrook.edu cs.stonybrook.edu/~jishang

Education

Stony Brook University, NY, USA **2018 – Present**
Ph.D. Candidate in Computer Science, GPA: 3.98/4, Advisor: Prof. Michael S. Ryoo

Shanghai Jiao Tong University, Shanghai, China **2014 – 2018**
B.S. in Computer Science, IEEE Pilot Class

Research Interest

I am interested in creating autonomous system via imitation learning and reinforcement learning. My current research focuses on visual and sequence representation learning for robotics, with a concentration on leaning from multiple viewpoints and multiple visual modalities.

Selected Publications

1. **Shang**, **J.**, Das, S. & Ryoo, M. S. *Learning Viewpoint-Agnostic Visual Representations by Recovering Tokens in 3D Space* In submission to NeurIPS. arXiv: 2206.11895.
2. Li, X., **Shang**, **J.**, Das, S. & Ryoo, M. S. *Does Self-supervised Learning Really Improve Reinforcement Learning from Pixels?* In submission to NeurIPS. arXiv: 2206.05266.
3. Burgert, R., **Shang**, **J.**, Li, X. & Ryoo, M. S. *Neural Neural Textures Make Sim2Real Consistent* In submission to CoRL. arXiv: 2206.05266.
4. **Shang**, **J.** & Ryoo, M. S. *StARformer: Transformer with State-Action-Reward Representations* in *Proceedings of the European Conference on Computer Vision (ECCV)* (2022).
5. **Shang**, **J.** & Ryoo, M. S. *Self-Supervised Disentangled Representation Learning for Third-Person Imitation Learning* in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2021).

Research Experience

Research Assistant, Stony Brook University **May 2020 – Present**
Visual and Sequence Representation Learning for Robotics

- Viewpoint-agnostic representation learning using Transformer [1][Visual][Viewpoint][Transformer]
  - Proposed 3DTRL, a learnable, differentiable layer that learns viewpoint-agnostic representations from monocular 2D image input.
  - It improves tasks including image recognition, multi-view video alignment and action recognition.
  - It is a light-weighted, plug-and-play module that achieves improvements with minimal (2% computation and 4% parameters) overhead than Transformer backbone.
- Transformer for real-robot visual imitation learning (Extension of [4])[Visual][Sequence][Transformer]
  - Collected a $\sim 60,000$-frame, $\sim$ 60-min human following dataset via a mobile robot with spherical camera, depth camera and LiDAR.
  - Trained proposed StARformer to perform human following task using imitation learning.

    - StARformer outperforms the baseline by $10\% - 30\%$ in offline evaluations and $\sim 100\%$ in real-world evaluation.
- Transformer for visual reinforcement learning and imitation learning [4] [Visual][Sequence][Transformer]
    - Adopted Transformers architecture for reinforcement learning tasks under a sequence modeling formulation.
    - Proposed StARformer, which models local representations explicitly from strongly related state, action, and reward tokens, and uses local representations for sequence prediction.
    - Results showed performance improvements over the existed Transformer-based method by over 70%, in both offline-RL and imitation learning.
    - StARformer is also highlighted for better modeling longer trajectories than the existed method.
- Third-person imitation learning for egocentric tasks [5] [Visual][Viewpoint]
    - Developed customized egocentric task environments in Minecraft (navigation) and in PyBullet (manipulation).
    - Introduced dual auto-encoders for learning joint FPV-TPV visual representation.
    - Explicitly split vector representations for disentangling agent state and third-person viewpoint representations.
    - Introduced representation permutation loss to train representations to be disentangled.
    - Results show the learned representations lead to better policies for both TPV and FPV imitation.
- Imitation learning / Navigation for ground mobility robots [Ongoing]
    - Multi-modal fusion of LiDAR and RGB inputs for representation learning.
    - Aiming on flexible scenarios when LiDAR/RGB is not applicable.

Research Assistant, Stony Brook University **Jan 2019 – Jan 2020**
Machine learning and pattern recognition for mobile sensors

- Pattern recognition in gesture typing on touch screen

Professional Activities

Reviewer: CVPR, ECCV
Guest Talk: Google Inc. (2022, Transformer for Robot Learning), CSE527 Introduction to Computer Vision (Fall 2021, graduate level), CSE525 Introduction to Robotics (Spring 2022, 2021, graduate level)
Teaching Assistant: CSE548 Analysis of Algorithms (Spring 2019, graduate level), CSE564 Visualization (Spring 2020, graduate level), CSE101 Computer Science Principles (Fall 2018)

Honors and Awards

- Merit Scholarship, Stony Brook University 2018-2019
- Outstanding Graduate of Colleges and Universities in Shanghai, China (Top 5%) 2018
- 1st Prize in China Undergraduate Mathematical Contest in Modeling 2017
- Academic Excellence Scholarship of SJTU (Top 20%) 2015, 2016, 2017

Technical Skills

Competitive Programming:[My LeetCode] Ranked 9/54 in SBU ACM ICBC Selection Contest, 2020
Programming Languages: [My Github] Python, Java, C/C++, Go, JavaScript
Technologies/Frameworks: PyTorch, Linux, Git, Tensorflow, Unity3D