

# Xiang Li

✉ xiangli8 [at] cs.stonybrook.edu 🏠 <https://xxli.me>

## Education

---

**Stony Brook University**, Stony Brook, NY, USA  
Ph.D. Candidate in Computer Science

2020 - Present  
Advisor: Prof. Michael S. Ryou

**Shanghai Jiao Tong University**, Shanghai, China  
M.S. in Control Engineering  
Thesis: Continuous Visual Object Tracking with View Morphing

2015 - 2018  
Advisor: Prof. Yue Zhou

**Shanghai Jiao Tong University**, Shanghai, China  
B.S. in Automation

2011 - 2015

## Research Interest

---

Self-supervised visual representation learning; Diffusion models & VLMs for robotics.

## Select Publications

---

1. **Li, X.**, Belagali, V., Shang, J. & Ryou, M. S. Crossway Diffusion: Improving Diffusion-based Visuomotor Policy via Self-supervised Learning. *IEEE International Conference on Robotics and Automation (ICRA) (2024)*.
  - We explore multiple self-supervised objectives to improve diffusion-based visuomotor policy for behavior cloning.
  - We propose Crossway Diffusion, introducing a state decoder and a simple objective to reconstruct observations from the intermediate representation of the diffusion network. The reconstruction loss is jointly optimized with the diffusion loss during training.
  - We achieve consistent and significant improvement over multiple simulated and real-world tasks.
2. Ranasinghe, K., **Li, X.**, Kahatapitiya, K., and Ryou, M.S. Understanding Long Videos in One Multimodal Language Model Pass. *arXiv preprint (2024)*.
3. Das, S., Jain, T., Reilly, D., Balaji, P., Karmakar, S., Marjit, S., **Li, X.**, Das, A. & Ryou, M. S. Limited Data, Unlimited Potential: A Study on ViTs Augmented by Masked Autoencoders. *Winter Conference on Applications of Computer Vision (WACV) (2024)*.
4. Burgert, R., **Li, X.**, Leite, A., Ranasinghe, K., and Ryou, M.S. Diffusion Illusions: Hiding Images in Plain Sight. *arXiv preprint (2023)*.
5. Burgert, R., Ranasinghe, K., **Li, X.** and Ryou, M.S. Peekaboo: Text-to-image Diffusion Models are Zero-shot Segmentors. *Workshop on Open-Domain Reasoning Under Multi-Modal Settings @ CVPR (2023)*.
6. **Li, X.**, Shang, J., Das, S., & Ryou, M. Does Self-supervised Learning Really Improve Reinforcement Learning from Pixels? *Advances in Neural Information Processing Systems (NeurIPS) (2022)*.
  - Unfortunately, NO (when the same amount of data and augmentation is used).
  - We conduct an extensive comparison of various self-supervised losses under the existing joint learning framework for pixel-based reinforcement learning in many simulated and real-world environments.
  - We perform evolutionary searches for the optimal combination of multiple self-supervised losses and the magnitudes of image augmentation, and confirm its limitations.
  - We conduct detailed ablations on multiple factors and demonstrate multiple properties of representations learned by different methods.

7. Burgert, R., Shang, J., **Li, X.** & Ryoo, M. S. TRITON: Neural Neural Textures Make Sim2Real Consistent. *Conference on Robot Learning (CoRL) (2022)*.
8. Shang, J., **Li, X.**, Kahatapitiya, K., Lee, Y.-C. & Ryoo, M. S. StARformer: Transformer with State-Action-Reward Representations for Robot Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) (2022)*.
9. Shang, J., Kahatapitiya, K., **Li, X.** & Ryoo, M. S. StARformer: Transformer with State-Action-Reward Representations for Visual Reinforcement Learning. *European Conference on Computer Vision (ECCV) (2022)*.

## Experience

---

- Beijing Falcon Image Technologies Co., Ltd**, Beijing, China 2018 - 2019  
*Software engineer*  
 Deploying LiDAR-based SLAM and navigation system on multiple types of self-driving cars with ROS.
- Hubei Cheng Shi Electronic Business Co., Ltd**, Shanghai, China 2016 - 2018  
*IoT software engineer*  
 Designing an intelligent video surveillance system in three IoT warehouses.  
 Key features: video recording, video streaming, and visual motion detection.

## Professional Activities

---

- Conference Reviewer: ICML'23&24, NeurIPS'23, WACV'24, ICRA'24, ICLR'24, RLC'24, and ECCV'24
- Teaching Assistant:  
 Introduction to Digital Image Processing (undergrad level, 2017, 2018, SJTU)  
 CSE353 Machine Learning (undergrad level, 2020 Spring, SBU)  
 CSE527 Introduction to Computer Vision (graduate level, 2021 Fall, SBU)  
 CSE525 Introduction to Robotics (graduate level, 2023 Spring, SBU)

## Patent

---

- Utility Model: Universal Serial Control Touch-screen Test Device, China 2016  
 An embedded system that interacts with touchscreens like a human finger. No.2016201772460

## Honors and Awards

---

- CVPR Outstanding Demos Award [4] 2023
- Outstanding Graduate of Colleges and Universities in Shanghai, China (Top 5%) 2018
- National Scholarship for Graduate Students, China 2017
- Outstanding Graduate of SJTU, China 2015