

# Shuaiyi Huang

huangshy2019@gmail.com ◇ huangshy@umd.edu ◇ shuaiyihuang.github.io ◇ google scholar ◇ linkedin

## RESEARCH INTEREST

---

My research lies in the intersection of **Computer Vision** and **Autonomous Agents**, with a focus on *video understanding*, *object recognition*, *policy learning* and *generative world modeling*. I've worked on **video-action recognition**, **vision-language-action models**, **video diffusion for robot manipulation**, **video instance segmentation**, and **visual matching**. I aim to build **foundation models** that seamlessly integrate visual, linguistic, and action understanding for real-world applications.

## EDUCATION

---

**University of Maryland, College Park**, Maryland, USA

Ph.D. in Computer Science, Sep 2020 - June 2025

- GPA: 3.93/4.0
- Supervisor: [Prof. Abhinav Shrivastava](#)

**ShanghaiTech University**, Shanghai, China

M.Sc. in Computer Science, Sep 2017 - June 2020

- GPA: 3.71/4.0
- Supervisor: [Prof. Xuming He](#)

**Tongji University**, Shanghai, China

B.E. in Software Engineering, Sep 2013 - July 2017

- GPA: 4.62/5.0
- Rank: 12/180

## EXPERIENCES

---

**Research Engineer, *World Modeling*** at **Meta FAIR Robotics**

July 28 - now, 2025

Work on **vision-language-latent-action models** for general manipulation.

**Research Intern, *Generative World Modeling*** at **NVIDIA**

May - Aug, 2023

Work on **video diffusion for robot manipulation**.

Mentor: [De-An Huang](#), [Linxi “Jim” Fan](#), [Yuke Zhu](#)

**Research Intern, *Video Understanding and Object Recognition*** at **NVIDIA**

May - Nov, 2022

Work on **video instance segmentation with point supervision**.

Mentor: [Zhiding Yu](#), [De-An Huang](#), [Shiyi Lan](#)

## PREPRINTS & PUBLICATIONS

---

- **Tokens: Semantic-Aware Relational Trajectory Tokens for Few-Shot Action Recognition**  
[Pulkit Kumar\\*](#), [Shuaiyi Huang\\*](#), [Matthew Walmer](#), [Sai Saketh Rambhatla](#), [Abhinav Shrivastava](#)  
*ICCV* 2025. [Project](#) [PDF](#) [Code](#)
- **TREND: Tri-teaching for Robust Preference-based Reinforcement Learning with Demonstrations**  
[Shuaiyi Huang](#), [Mara Levy](#), [Anubhav Gupta](#), [Daniel Ekpo](#), [Ruijie Zheng](#), [Abhinav Shrivastava](#).  
*ICRA*, 2025. [Project](#)

- **TraceVLA: Visual Trace Prompting Enhances Spatial-Temporal Awareness for Generalist Robotic Policies**  
Ruijie Zheng\*, Yongyuan Liang\*, Shuaiyi Huang, Jianfeng Gao, Hal Daumé III, Andrey Kolobov, Furong Huang, Jianwei Yang.  
*ICLR*, 2025. [Project](#) [PDF](#) [Code](#)
- **AUTOHALLUSION: Automatic Generation of Hallucination Benchmarks for Vision-Language Models**  
Xiyang Wu\*, Tianrui Guan\*, Dianqi Li, Shuaiyi Huang, Xiaoyu Liu, Xijun Wang, Ruiqi Xian, Abhinav Shrivastava, Furong Huang, Jordan Lee Boyd-Graber, Tianyi Zhou, Dinesh Manocha.  
*EMNLP*, 2024. [PDF](#) [Code](#)
- **ARDuP: Active Region Video Diffusion for Universal Policies**  
Shuaiyi Huang, Mara Levy, Zhenyu Jiang, Anima Anandkumar, Yuke Zhu, Linxi Fan, De-An Huang, Abhinav Shrivastava.  
*IROS*, **Oral**, 2024. [PDF](#) [Code](#)
- **What is Point Supervision Worth in Video Instance Segmentation?**  
Shuaiyi Huang, De-An Huang, Zhiding Yu, Shiyi Lan, Subhashree Radhakrishnan, Jose M. Alvarez, Abhinav Shrivastava, Anima Anandkumar.  
*CVPRW*, 2024. [PDF](#)
- **UVIS: Unsupervised Video Instance Segmentation**  
Shuaiyi Huang, Saksham Suri, Kamal Gupta, Sai Saketh Rambhatla, Ser-nam Lim, Abhinav Shrivastava.  
*CVPRW*, 2024. [PDF](#)
- **Towards Scalable Neural Representation for Diverse Videos**  
Bo He, Xitong Yang, Hanyu Wang, Zuxuan Wu, Hao Chen, Shuaiyi Huang, Yixuan Ren, Ser-Nam Lim, Abhinav Shrivastava.  
*CVPR*, 2023. [PDF](#) [Code](#)
- **Learning Semantic Correspondence with Sparse Annotations**  
Shuaiyi Huang, Luyu Yang, Bo He, Songyang Zhang, Xuming He, Abhinav Shrivastava.  
*ECCV*, 2022. [PDF](#) [Code](#)
- **Confidence-aware Adversarial Learning for Self-supervised Semantic Matching**  
Shuaiyi Huang, Qiuyue Wang, Xuming He.  
*PRCV*, 2020. [PDF](#) [Code](#)
- **Dehazing Evaluation: Real-World Benchmark Datasets, Criteria, and Baselines**  
Shiyu Zhao, Lin Zhang, Shuaiyi Huang, Ying Shen, Shengjie Zhao.  
*TIP*, 2020. [PDF](#) [Code](#)
- **Dynamic Context Correspondence Network for Semantic Alignment**  
Shuaiyi Huang, Qiuyue Wang, Songyang Zhang, Shipeng Yan, Xuming He.  
*ICCV*, 2019. [PDF](#) [Code](#)
- **Evaluation of Defogging: A Real-world Benchmark Dataset, A New Criterion and Baselines**  
Shiyu Zhao, Lin Zhang, Shuaiyi Huang, Ying Shen, Shengjie Zhao, Yukai Yang.  
*ICME*, 2019. [PDF](#) [Code](#)
- **Structured Attentions for Visual Question Answering**  
Chen Zhu, Yanpeng Zhao, Shuaiyi Huang, Kewei Tu, Yi Ma.  
*ICCV*, 2017. [PDF](#) [Code](#)