

# Statement of Purpose

My objective for graduate studies is to prepare myself for becoming a pioneer who can bring real innovations to the field of artificial intelligence. Specifically, I want to develop intelligent agents that can learn from environments by autonomous interactions.

My past research experience focused on the algorithms and applications of reinforcement learning, imitation learning, and computer vision. With my determination to contribute to these fields, I have undertaken many exciting researches during my academic life in University of California San Diego and Peking University.

My internship in SenseTime AI led me to the world of computer vision, where I worked on face recognition and detection. I got the chance to realize techniques of computer vision that could truly contribute to real life. However, although such algorithms can exceed human's performance on some datasets, it requires manually picked models and a large number of computational resources. To me, essential pieces are still missing towards building agents of generic intelligence.

At the same time, my enthusiasm for machine learning was driving me to seek for possible frameworks that would enable the learning from the environments without too many human-designed priors. Fascinated by the success of Alpha-Go, I started to explore multi-agent reinforcement learning under the guidance of Prof. Yizhou Wang at Peking University and Dr. Bo Xin at Microsoft Research Asia. The designed hierarchical multi-agent framework achieved state-of-the-art performance on some near-symbolic environments for micromanagement tasks (e.g. StarCraft I, Traffic Junction), which was finally accepted as an oral presentation in the NIPS 2017 Hierarchical RL Workshop. However, when we tried to extend this framework to video games such as StarCraft II, it did not work as well as expected (though outperformed the baseline proposed by DeepMind). We observed that it was indeed hard for current deep learning algorithms to deal with object detection and decision-making simultaneously only driven by extrinsic rewards, even if we adopt the state-of-the-art backbone architecture for object detection.

These experiences made me rethink the relationship between perception and decision making. As a key component in current intelligence systems, perception should not be limited to pure pattern recognition problems. It should serve as an important mechanism to understand environments, which will not only facilitate the decision-making process but even get evolved driven by feedbacks from interactions.

Therefore, advised by Prof. Hao Su at UC San Diego, I started the journey of bridging environment understanding and interaction. In a collaboration with Mr. Zhiao Huang, I first considered how to model the environment using a low-dimensional representation for planning and reasoning. When solving the goal-reaching problems in large-scale MDPs, we observed that network-based universal function approximators (UVFA) could not learn long-horizon planning well. We felt that reasoning grounded on the structured representation of the environment might help to address the long-horizon planning problem. To this end, we built a graph-based map to encode the topology of the environment and used this map to assist the local-level policy network learning. Note that this map is not pre-defined but emerged during the interaction process. This approach achieved state-of-the-art on goal-oriented robot control, manipulation, and navigation tasks. It was submitted and accepted by NeurIPS 2019, where I shared the co-first authorship.

When addressing the concerns from a reviewer of this paper, I realized that the idea of this work might also benefit imitation learning. In the NeurIPS work, we abstracted the whole state space by a map built upon landmark states. After discussions with Prof. Su, I got interested in further exploring the role of states in policy learning. Particularly, while there are many possible ways to achieve a

given goal, they share some common states in successful trajectories. In other words, we think that a state-based perspective may allow us to better characterize the equivalence of policies. I explored this idea in the context of cross-morphology imitation learning, when the expert and the imitator have different agent dynamics. This is the setting that few papers have explored systematically. I finally established a framework by aligning the state sequence from demonstration policy and imitator policy. Implementation-wise, this design is achieved by using Variational Auto-Encoder to match local state transitions and reward induced from Wasserstein Distance to match the global state visitation distribution. The local and global constraints are combined in the proximal policy optimization (PPO) framework. Experimentally, it achieved sample efficiency and stable performance. This work is accepted by ICLR 2020, where I serve as the first author.

These experiences made me realize that current RL algorithms still solve problems in a brute-force manner and behave quite differently from how humans learn, as their behaviors are adjusted by trial-and-error without effectively distilling reusable knowledge from their past experiences. Furthermore, although imitation learning can partially mitigate this problem, current approaches still lack robustness and flexibility when the agent or environment dynamics are not the same.

On another front, I actively participated in research activities in Prof. Su's lab to build simulators of the physical environment. When targeting at developing real-world intelligence, current simulated environments usually lack either realistic physics or diverse objects, which cause a large domain gap to the real world. To further explore this direction, I am fortunate to participate in a project to build a physics-based simulator of the real environment using part-based 3D data from the PartNet dataset. In this project, I combined 3D perception with reinforcement learning to solve some robot manipulation tasks. To generate state representations, I used PointNet++ to encode object mobility features with an auxiliary task to predict object motions, which showed some potential of generalizability for unseen object manipulation. This work is submitted to CVPR 2020.

I believe that a powerful intelligent system will contain a robust perception module to extract information from observations, model environments with causal structure based on reasoning, and improve policies through active exploration and interactions. In the future, I will leverage knowledge from graphics, physics, visual perception and decision-making to build such intelligence that can truly extrapolate its skills and knowledge for general tasks and purposes.