Statement of Purpose

I aspire to pursue doctoral research in Computer Science, and my career objective is to become a professor. My major research interests include machine learning and cognitive sciences.

Personal Motivation

The motivation behind my pursuit of doctoral research stems from the realization of challenges encountered in the use of computational models for solving day-to-day tasks. In the final project in the course "Autonomous Driving" under Prof. Longbo Huang, I designed an autonomous model for routing and driving through heavy traffic. The greatest difficulty we encountered was the handling of inevitable traffic jams at crossroads. This is owing to lack of a mechanism for vehicles to beep and/or signal to avoid collisions. I later delved into this problem during the course on "Artificial Intelligence: Principles and Techniques" instructed by Prof. Chongjie Zhang. In the subsequent course project, my team sought to make vehicles communicate with other vehicles operating in regular traffic scenarios. Albeit not a significant research, results of this study demonstrated realization of better control over total traffic flow.

Although I personally deem these experiences educational, the gap between computers and human intelligence in what seems to be the learning of basic skills or common-sense knowledge sparked my interest in relevant research.

Research at MIT

For the purpose of understanding the this problem better, I worked with Prof. Joshua Tenenbaum in 2019 for six months in the Computational Cognitive Science Group at MIT. During this period, I had the opportunity to collaborate with Jiayuan Mao, Chuang Gan and Jiajun Wu, and we chose to investigate the problem on the learning of metaconcepts (abstract relational concepts), such as *synonym* and *hypernym*, in natural language. Learning of these abstract concepts requires a higher categorization level than concrete visual concepts (e.g., *red* and *metal*). For instance, a clear understanding of the statement "*red and green describe the same property of objects"* necessitates (1) understanding of the semantic meaning of *red* and *green*; and (2) comprehension of the *same property* notion in the sense that both *red* and *green* categorize the color of an object.

The paper proposed a visual concept-metaconcept learner (VCML)—a symbolic-reasoning framework for answering questions. I designed the metaconcept neural operator for VCML, which operated on two visual-concept representations and output their metaconcept relations. The VQA datasets used in this project comprise images paired with only visual reasoning questions. Thus, I augmented the datasets with text-only metaconcept questions, such as "*Do red and green describe the same property of objects?*" to support metaconcept learning. Evaluation on both synthetic and real-world datasets demonstrated how metaconcepts learning and visual concepts assist each other. This work will be presented in the poster form at NeurIPS 2019 [1].

Research at Tsinghua

Another problem I have always been interested in is the recognition of the physical world. It appears that humans excel in interpreting visual scenes (often as a collection of objects), and learn the capability quickly even in novel environments, such as a new video game.

Driven by this problem, I worked on a project targeted at incorporating prior knowledge of objects in video games under the supervision by Prof. Chongjie Zhang at Tsinghua University. In the project, we were enlightened by the observation that deconstructing video images into a collection of mutually interacting objects does not compromise the information required for prediction and planning. To leverage this feature, I attempted to employ a modified CNN structure to which an original video image was provided as input. Its output is an "object map" with size much smaller compared to the input image. The "map" contained information to facilitate reconstruction of the original image and predict subsequent frames.

Results of preliminary experiments performed on Sokoban revealed the extracted "object map" as interpretable. The proposed model was capable of modeling inter-object interactions such as collisions. Upon implementation on more complex environments, such as Flappy Bird and Pong, however, the proposed model struggled to find a balance between comprehensiveness (inclusion of all objects in the "object map" for image reconstruction) and succinctness (creation of a simple "object map" for prediction). This project is yet in progress. Nevertheless, what intrigues me most is that it seems natural for humans to deal with such situations easily, as we exclusively attend to important objects in a given scene. I view this direction exciting and worth continue investigating.

Closure

Though pertaining to different topics, the two above-mentioned research projects I have been involved in both draw inspiration from and attempt to model human behavior. In view of these experiences, I am increasingly fascinated by problems on the development of computational models that match human performance in terms of interpretation of the physical world and the ability to reason logically in a unified way. It is interesting to note that even infants outperform computer models with regard to gaining common-sense knowledge and acquiring basic life skills. I see this as an exciting opportunity to narrow this gap between nature and technology as well as gain a deeper understanding of the field over the course of my prospective doctoral research.

(Paragraph to mention faculties I am interested in and their researches topics.)

References

Chi Han*, Jiayuan Mao*, Chuang Gan, Joshua B. Tenenbaum, and Jiajun Wu. Visual Concept-Metaconcept Learning. In *NeurIPS*, 2019