Statement of Purpose

I am Haoming Li, a second-year electrical engineering master student at the GRASP Lab of the University of Pennsylvania (UPenn), co-advised by Prof. Nadia Figueroa and Prof. Pratik Chaudhari. My research interests generally lie in building interactive autonomous agents with general problem solving capabilities. My research pursuits primarily involve developing interactive autonomous agents endowed with versatile problem-solving capabilities. Specifically, I am keen on delving into **reactive perception-based safety-critical control**, **deep learning-based robot mapping and navigation**, and other **human-robot interaction (HRI)** applications. My aspiration is to apply for the Ph.D. program in the ESE department at UPenn for the Fall of 2024.

Human-Robot Interaction (HRI) stands as a pivotal factor in the rapidly expanding realm of intelligent personal-service and entertainment robots. Applications span a broad spectrum, ranging from aiding the elderly and individuals with severe disabilities to entertaining visitors at amusement parks. The paramount concerns within HRI encompass efficiency and safety. Contemporary robots increasingly rely on visual feedback to navigate novel and intricate environments, mirroring human capabilities. Leveraging the continuous advancements in computer vision and deep learning techniques holds the potential for significant strides in safe and efficient control. Additionally, the advent of deep learning-based mapping and navigation, propelled by the emergence of the neural radiance field, presents superior reconstruction performance. Harnessing information about the environment, these techniques offer a pathway to further elevate the efficiency and safety of robotic systems.

My enthusiasm for research first took root during my undergraduate studies, specifically in the domain of object detection. As a sophomore, I collaborated with Professor Yan Liu from the University of Chinese Academy of Sciences on automating the detection of ice crevasses. Traditionally, ensuring the safety of vehicles involved operators manually processing extensive data collected by ground-penetrating radar to identify concealed crevasses beneath ice sheets. The inherent limitations of operator judgment prompted my quest for an accurate, real-time, and automated solution for ice crevasse detection. In pursuit of this goal, I conceptualized ice crevasse detection as an object detection problem, with Faster R-CNN emerging as the most suitable method due to its high accuracy and real-time capabilities. I constructed a pipeline capable of preprocessing ice sheet data, learning features, and executing crevasse detection in an end-to-end fashion. The exper results validated my approach, showcasing a remarkable 95% accuracy and real-time performance. This research culminated in an oral presentation at the 15th IEEE International Conference on Signal Processing (ICSP2020). This experience not only honed my skills in data preprocessing, object detection, experiment design, and paper writing but also served as a catalyst for a profound realization: machine perception plays a pivotal role in ensuring the safety of robots and vehicles. Recognizing the potential benefits of fully leveraging visual information, I became inspired to delve deeper into exploring the advantages of computer vision within the robotics community.

Hence, in my senior year, I broadened my footprint in research of computer vision. Under the supervision of Dr. Jinghui Yang and Dr. Keming Chen from the Institute of Electronics, Chinese Academy of Sciences, my bachelor's thesis aimed to solve the Hyperspectral Image Change Detection (HSI-CD) problems. The project targets a challenging computer vision problem aiming to detect areas that have undergone changes within the same region at different times. The difficulty arises from the high spectral resolution and the low spatial resolution of hyperspectral images as well as the complexity of modeling the spatio-temporal dependency. Given the limited availability of hyperspectral remote sensing images, coupled with the ”Curse of Dimensionality” due to the high-dimensional nature of the data, my approach integrated insights from the success of Vision Transformer in computer vision. Viewing each pixel as a token within the context of a Transformer, I successfully reduced data dimensionality while preserving crucial features. To address spatio-temporal dependencies, I employed spatial and temporal self-attention, enhancing the global context and representational qualities of learned tokens. A shallow MLP decoder facilitated bi-classification (assigning ”changed” or ”no change” labels), elevating accuracy from 98.92% to 99.10%. Notably, the absence of skip-connections and a complex decoder ensured computational efficiency. While this project provided a solid foundation for handling high-dimensional vision data, my interest in real robotic applications prompted a shift towards representing 3D scenes. Additionally, I aspire to explore methodologies enhancing neural network generalization to alleviate sample scarcity issues. This shift has directed my research focus toward 3D computer vision and the intricacies of neural network properties upon entering Penn.

During my inaugural semester at Penn, I delved into the realm of exploring the advantages of Lipschitz continuity in neural networks within Prof. George Pappas' group. Recognizing the inherent sensitivity of neural networks to their inputs and the potential repercussions on downstream tasks, I sought to address issues of overfitting to image noise. It is well-established that a small Lipschitz constant is intricately linked to the robustness and generalization of various machine learning problems. Enhancing the Lipschitz continuity of neural networks promises a reduction in excess risk, thereby improving generalization. The project aimed to unravel the benefits of maintaining Lipschitz continuity in Multi-Layer Perceptrons (MLPs) across diverse tasks. Estimating the Lipschitz constant involved computing the product of the norm of the weight matrix for each layer. This estimation was then globally penalized as a regularization term, with additional refinement through the application of weight normalization to restrict the Lipschitz constant further. Experimental validation on tasks such as 2D/3D interpolation and adversarial robustness substantiated the significant improvement in generalization achieved by constraining the Lipschitz constant. Given the critical role of robustness to noisy input in ensuring the safety of robots, particularly when dealing with visual information prone to containing noise, the potential applicability of this approach to vision-based safe control problems has further ignited my interest in the intersection of safe control and computer vision.

Recent studies show that observation-based control barrier functions (CBFs) is more robust than state-based CBFs because the safe region may change. If a robot can directly map the environment and construct CBFs using visual information, both update rate and robustness will be improved. Thus, since May, I worked with Prof. Nadia Figueroa on the reactive perception safe control. I use both parametric and coordinate encodings to implicitly represent the scene. This joint encoding strategy can improve the optimization speed and reconstruction performance. Then I use two simple MLPs to decode the features into signed distance functions (SDFs), color, and density. With SDFs, I construct a CBF-QP to prevent the robot from the obstacles. With density and color, we could better reconstruct the scene, which could be used to refine the estimated SDFs. Also, like a SLAM, this method can give us the estimated poses so that we do not need to provide the poses manually. To prove the advantages of my method, I perform experiments on both a simulation environment and a real-world robot (Fetch). As far as I know, this is the first method that takes RGB-D images as input and decodes the features into SDFs to construct a CBF-QP to avoid collision in real-time. This project afforded me valuable insights into control theory, Robot Operating System (ROS), and SLAM. Furthermore, it underscored the importance of considering the update rate of a system for optimal reactivity in real-world applications. Inspired by this endeavor, my master's thesis aims to delve into the advantages of keyframe selection, encoding, and sampling strategies, with a focus on advancing neural visual SLAM.

I am interested in Prof.xxx, Prof.xxx, and Prof. xxx's work and my experience is closely related to them. Prof.xxx's work, "xxx" inspire to think xxx. Her work 'xxx' uses xxx to realize xxx. I think using xxx would improve the performance. Since xxx is better at xxx, the xxx results will be better. It would be my honor to work with Prof. xxx, Prof. xxx, and Prof. xxx. Their expertise will help me a lot in my academic career.

Aspiring to achieve expertise in robotics, pursuing a doctoral degree stands out as a logical progression in my career trajectory. I am interested in leveraging computer vision and deep learning techniques to enhance the perceptual capabilities of autonomous agents, enabling them to effectively address a myriad of challenges in understanding and navigating their environment. My vision encompasses realizing a world where robots interact intelligently with humans and the physical environment. The prospect of attaining this goal is particularly promising with the abundant research resources and collaborative atmosphere at Penn. It would be an honor for me to be part of the esteemed PhD program at Penn, contributing to the advancement of robotics and the realization of a technologically enhanced future.