**Statement of Purpose**

**Ziyi Wu**

Applicant for Ph.D. in Robotics

My primary research interests are computer vision and its applications to 3D data. As an undergraduate, I was fortunate to participate in several projects at both Tsinghua and Stanford, from which I have realized the importance of 3D vision tasks, especially those related to robotics and self-driving cars. To further pursue my interests, I am excited to apply for your Ph.D. program in the Robotics Institute, where I wish to further research on improving the efficiency and robustness of 3D vision algorithms. I believe this will enable our machines to perceive the world more effectively and intelligently.

I first became interested in computer vision when I enrolled in Stanford’s online courses about deep learning. Deeply impressed by the success of modern AI algorithms such as AlphaGo, I dedicated myself to more related classes. For example, I trained a Wide-ResNet to classify articles of clothing in the *PRML* course project. In order to improve the accuracy, I applied various tricks including MixUp augmentation and Snapshot Ensembles. The challenging courses in Tsinghua helped me obtain mastery of the fundamentals of computer vision, enabling me to succeed in later research endeavors.

With a keen interest in computer vision, I joined the research group of Prof. Jie Zhou and Prof. Jianjiang Feng in my sophomore year to study medical image analysis. This problem was particularly interesting to me because it could save lives if incorporated in healthcare systems. My first project was about efficient cardiac landmark detection. Previous methods for this task failed to leverage all information in the input and were too time-consuming, preventing their deployment. To solve this problem, I developed a novel cascade approach which followed a coarse-to-fine manner, and thus greatly improved inference speed. This work was published at the MICCAI’19 workshop. This project convinced me of the importance of algorithmic efficiency for deployments of computer vision methods, which is a lesson I have strived to apply to my other projects.

Having learned the importance of efficient inference in real-world deployments, I was eager to explore it in other computer vision tasks. So I joined Prof. Jiwen Lu’s group in my senior year to study how to design binary neural networks (BNNs) for lightweight object detection. I introduced the Information Bottleneck (IB) principle for redundancy removal in the network to fully utilize the capacity of BNNs. I also regularized the outputs with sparse object priors to reduce false positives. This work was accepted by CVPR’20. Later, I extended *BiDet* to *AutoBiDet*, which leverages GANs to adjust the IB trade-off automatically and learns adaptive priors to alleviate false positives more effectively. We have submitted this work to T-PAMI. In this project, I tackled many problems in engineering and cross-platform programming. More importantly, I learned how to combine well-founded theories with deep learning to solve new problems.

While delving deeply into 2D computer vision, I became intrigued by recent trends in 3D vision. Last spring term, I took a course about autonomous driving, from which I realized the importance of point clouds in 3D perception tasks. Therefore, I did a summer internship at Stanford University supervised by Prof. Leonidas J. Guibas on point cloud analysis. I discovered that, despite tremendous progress achieved in this field, few people had explored the robustness of point cloud networks, posing a security threat to those safety-critical applications. To tackle this crucial problem, I led a project studying adversarial attack and defense in 3D point clouds. I novelly employed implicit function networks to restore precise object shapes, and adopted an optimization process to recover natural point distributions. Our defense demonstrated state-of-the-art robustness against all the existing 3D attacks on five typical models. Based on this work, I wrote a first-authored paper and submitted it to CVPR’21. This was the first time I went through the whole process of a research project, from identifying the problem through literature review to method proposal and finally paper writing. Now, I am able to cooperate with others in a team and make every discussion productive and fruitful.

Throughout my three-year immersion in computer vision, I have always been determined to pursue an academic career. Fundamentally, I believe 3D vision is a promising topic with a myriad of unexplored research directions. For example, AR/VR applications often involve reconstruction of complicated object geometries, but recent works in 3D reconstruction rarely consider the computational budget, preventing their deployment on mobile devices such as smartphone. In addition, I think combining 3D vision with temporal information is essential to model our 3D world. In self-driving cars, LiDAR produces point cloud sequences as outputs, requiring lightweight models to process them in real-time, which I think my experience in efficient inference can help. On the other hand, I believe adversarial robustness in 3D vision tasks deserves further investigation. From my summer intern project, I discovered that 3D attacks are very flexible and easy to succeed, which can be a severe threat to applications such as autonomous driving and robot navigation. I am currently leading a project studying black-box attacks, which can fool point cloud models without knowledge of them. This can exacerbate the problem of adversarial robustness in real-world 3D applications.

I am applying to the Ph.D. program in the Robotics Institute of CMU because of its impactful research groups. There are several professors whose projects are particularly appealing to me. Prof. Simon Lucey’s work on object geometry modeling and mobile computer vision is fascinating to me, and I think combining them will enable practical algorithm deployment. I also like his idea of injecting priors with model-based vision, which can improve the reliability of model output. Prof. Deva Ramanan’s projects on scene flow and videos spark my interest because we humans perceive the world as temporal sequences, which need efficient and robust methods to process. Meanwhile, I admire Prof. Abhinav Gupta’s work on the intersection of computer vision and robotics. I want to explore the robustness in robotics tasks, which is crucial for robots operated in real-world scenarios. Overall, I am fully convinced that CMU’s abundant resources and collaborative environment can provide the best guidance to my academic career.