

Statement of Purpose

Ziyi Wu

Applicant for CS Ph.D.

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 University, 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 CS department, where I wish to further research the efficiency and robustness of 3D vision algorithms. I believe this is fundamental to enabling machines to intelligently perceive and interact with the world.

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 courses at Tsinghua. For example, I trained a Wide-ResNet to classify articles of clothing in the *PRML* course project. To improve its accuracy, I applied various vision techniques, including Snapshot Ensemble, MixUp and GroupNorm. These challenging courses helped me command the fundamentals of computer vision, and prepared me to accomplish further 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 topic was particularly intriguing to me because it can save lives when incorporated in healthcare systems. My first project was efficient cardiac landmark detection. Previous methods failed to leverage global information in the input and were too time-consuming, preventing their practical applications. To solve this problem, I developed a cascade approach which follows a coarse-to-fine manner, and significantly improved the inference speed. This work was published at **MICCAI'19** workshop. Working on 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 understood the importance of efficient inference in real life applications, I was eager to explore it in other computer vision tasks. So, I joined Prof. Jiwen Lu's group in my junior year to study the design of binary neural networks (BNNs) for lightweight object detection. I introduced the Information Bottleneck (IB) principle for removing redundancy in the network to fully utilize the capacity of BNNs. I also regularized the outputs with sparse priors to reduce false positives (FP). This work named *BiDet* was accepted by **CVPR'20**. Later, I extended *BiDet* to *AutoBiDet*, which leverages GANs to automatically adjust the IB trade-off and learns adaptive priors to alleviate FP 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, through 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, only a 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 employed implicit function networks to restore precise object shapes, and adopted an optimization process to recover natural point distributions. Our defense presented state-of-the-art robustness against all the existing 3D attacks on five typical models. A paper based on this work is now in submission to **CVPR'21**, of which I am the first author. This was the first time I went through the whole process of a research project, from identifying the problem to proposing a solution and finally writing a paper. This experience gave me confidence in my abilities to lead novel research and good collaborations.

Throughout my two-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 smartphones. 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. My experience in efficient inference will enable me to contribute to this subfield. In addition, I believe adversarial robustness in 3D vision tasks deserves further investigation. During my summer internship, I discovered that 3D attacks are very flexible and easy to succeed, which poses a severe threat to applications such as robot navigation and autonomous driving. I am currently leading a project studying black-box attacks, which can fool point cloud models without knowledge about them. This can exacerbate the problem of adversarial robustness in real-world 3D applications.

I am applying to the Ph.D. program in the CS department of UT Austin because of its impactful research groups. There are several professors whose projects are particularly appealing to me. **Prof. Qixing Huang**'s work on 3D vision tasks is fascinating to me, and I really like the hybrid representations he proposed. I think this is a promising way to improve the robustness of 3D algorithms. I admire **Prof. Yuke Zhu**'s projects on the intersection of robotics and computer vision. I would be happy to leverage temporal information for more comprehensive scene understanding. Besides, **Prof. Philipp Krähenbühl** has wonderful works on autonomous driving and video analysis. I am eager to design highly efficient for these tasks to enable their real-world deployment. Overall, I believe UT Austin's abundant resources and collaborative environment can provide the best guidance to my academic career.