**Statement of Purpose**

**Ziyi Wu**

Applicant for Ph.D. in Robotics

My primary research interests are **computer vision** and its applications to **3D data**. To be specific, I am working on how to design AI algorithms that enable our machines to perceive the world more efficiently and intelligently. During the past three years, I was fortunate to participate in several computer vision projects at both Tsinghua and Stanford, from which I have realized that 3D-related topics are quite promising, especially with respect to robotics in real-world scenarios. To further pursue my interests, I am excited to apply for your renowned Ph.D. program in the Robotics Institute.

I first became familiar with machine learning and computer vision when I enrolled in Stanford’s CS229 and CS231n online courses. Deeply impressed by the revolutionary performance enabled by modern AI algorithms, I dedicated myself to more related classes, where I accumulated abundant experience implementing computer vision approaches to solve problems. 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 Snapshot Ensembles, MixUp and Group Normalization, and finally **ranked 1/180** in the class. The challenging courses in Tsinghua help me obtain basic knowledge and engineering skills in computer vision, enabling me to quickly adapt to new tasks.

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. My first project was about efficient cardiac landmark detection. The input data of this task are 3D CT images, which require high GPU memory to process. Previous methods thus apply 3D U-Net on local patches and stack each individual output as prediction result. However, such approaches fail to utilize the global information of input and is also time-consuming, harming their performance and deployment. To tackle this problem, I proposed a cascade method that first regresses coarse coordinates of landmarks on a low resolution version of input, and then refines the results on local patches at original resolution. I also developed a two-stage training schedule to accelerate convergence and improve its robustness. Experimental results demonstrated the superiority of our method. A paper based on this work was finally published on **MICCAI’19** workshop.

Having learned the importance of algorithm efficiency in real-world deployments, I was eager to study efficient inference in other computer vision tasks. I joined **Prof. Jiwen Lu**’s group to explore how to design binary neural networks (BNNs) for object detection. At first, I encountered numerous false positives because of the limited representational power of BNNs and the redundancy in their intermediate feature maps. Inspired by weekly discussions with my collaborators, I introduced the Information Bottleneck (IB) principle to remove the redundancy by formulating the detection task as a Markov chain and rewrote the problem from a Bayesian perspective. The proposed algorithm, termed *BiDet*, boosted the performance of binary detectors significantly. Based on this work, our paper was accepted by **CVPR’20**. Later, I extended *BiDet* to *AutoBiDet*, which leverages GANs to measure the amount of information contained in the input data and automatically adjusts the IB trade-off accordingly. I also generalized the techniques adopted in *AutoBiDet* to improve other model compression methods such as low-bit quantization and channel pruning in detection, showing the universality of our approach. We submitted this work to **T-PAMI** and it is currently under review.

Though having delved deeply into 2D computer vision, I gradually shifted my interest to the 3D domain. This summer, I did an internship at Stanford University supervised by **Prof. Leonidas J. Guibas** on point cloud analysis. I discovered that, despite the tremendous progress achieved in this field, few people had explored the robustness of point cloud models. To tackle this interesting problem, I led a project studying adversarial attack and defense in 3D point cloud. I first reviewed existing 3D attacks and summarized their effects into two aspects from a geometric perspective. Then, I realized that none of previous defense approaches can simultaneously address both the effects, resulting in their vulnerability against well-designed attacks. To solve this problem, I employed deep implicit functions for precise shape restoration. Concretely, our defense named *IF-Defense* first performs accurate surface recovery via the predicted implicit field to alleviate geometric distortion, and then utilizes an optimization process to retain point clouds with natural point distribution. As a result, *IF-Defense* demonstrates the state-of-the-art robustness against all the existing 3D attacks on five typical point cloud classification networks. Based on this project, I wrote a **first-authored** paper and submitted it to **CVPR’21**.

Throughout my three-year immersion in computer vision, I am now able to extract good questions from literature review, put ideas into practice, and verify them through well-designed experiments. I also learn how to cooperate with others in a research team. Fundamentally, I believe that 3D vision is a promising topic with a myriad of unexplored research directions. For example, in autonomous driving, LiDAR produces point cloud sequences as raw outputs, requiring light-weight models to process them in real-time, which I think my experience in efficient inference can help. Additionally, I think adversarial robustness in 3D vision deserves further investigation. I am currently leading a project studying black-box attacks, which can fool the point cloud networks without prior knowledge of them. This may bring severe security threats to those safety-critical applications.

I am applying to the Ph.D. program in the Robotics Institute of CMU because of its impactful research groups and insightful projects. There are three professors whose projects are particularly appealing to me. It is my dream to join **Prof. Simon Lucey**’s group on applying deep learning algorithms to modeling object geometry. I believe that his approach of injecting priors with model-based vision is very insightful. **Prof. Deva Ramanan**’s projects on videos also fascinate me because we humans perceive the 3D world as temporal sequences of images. Meanwhile, I admire **Prof. Abhinav Gupta**’s inspiring works on the intersection of robotics and computer vision. I am fully convinced that CMU can provide the best guidance to my academic career.

Overall comments:

You write very well, and I had very few grammatical fixes. Your research and publication experience is very impressive, and your essay definitely informs the reader that you are highly qualified.

My main criticism is that this essay is essentially an extended CV. You have described your experience and your publications. But that isn’t the purpose of a Statement of Purpose. The essay needs to be restructured to better respond to their prompt.

After your introduction, the next paragraph of the Statement of Purpose should briefly outline

1. What research topic you hope to pursue in your PhD program
2. The significance of the research (What is the impact on the world or societal impact if your research succeeds?)
3. Why that topic is interesting to you
4. Why you are the best person to tackle that research question

You can then go on to describe your previous research experience. But the description should not be simply a description of what you did. It should always focus on what you learned from the project, how that shaped your PhD interests, and how that experience will help you succeed in your PhD research.

Remember that this document is complementary to your resume. Your publications and experiences are already on there. The focus here is on evaluation of those experiences and the direction of the future of your research. The essay is also unlikely to be read by someone who is knowledgeable about 3D vision. You can skip the technical details of your papers.