

# Statement of Purpose

Chengyong Cui

Contact: hahayunc3c3y25@gmail.com

## Brief Introduction

As a research assistant at Zhejiang University of Technology's City Science and Social Computing(CSSC) Lab, tutored by Professors Guojiang Shen and Xiangjie Kong, I specialize in gesture recognition, computer vision, human-computer interaction(HCI) and recommendation system. My work includes developing a robust gesture recognition pipeline with a 92.68% accuracy rate, a gesture recognition network achieving 94.56% accuracy on the AUTSL dataset, and conducting a user experiment in a driving simulation environment, demonstrating a 65% improvement in driver performance due to gesture interaction.

## Research Experience

At the CSSC lab, I developed an in-air text-writing recognition module, part of an interactive federated active learning project. I integrated fingertip coordinates with a TMBA model using Google's Mediapipe, enhancing trajectory image processing and character prediction with 94.42% accuracy. This work contributed to the publication of "FedAWR: An Interactive Federated Active Learning Framework for Air Writing Recognition" in IEEE Transactions on Mobile Computing.

My interest in gesture recognition, particularly in the nuances of subtle gestures, was sparked by the complex and challenging nature of this field. Motivated by these, I focused my efforts on enhancing gesture recognition methods. This pursuit led to a project with the in-Vehicle Interaction Information System (IVIS). In this role, I engineered a sophisticated gesture recognition system that leveraged RGB cameras, significantly broadening its usability within the in-vehicle environment.

My study on RGB-based gesture recognition methods, in light of the IVIS's computational limitations, led to the adoption of Google's Mediapipe for hand coordinate estimation. This choice was pivotal due to its performance optimization and hardware acceleration capabilities. The need to classify a limited set of gestures for driver-IVIS interaction guided my exploration into back-propagation (BP) and dynamic time warping (DTW) methods, culminating in a lightweight pipeline with an impressive 92.68% accuracy in the demanding in-vehicle environment.

To evaluate the real-world impact of gesture interaction, I set up a driving simulation environment and developed a driver interaction interface supporting both touch and gesture-based interactions. The results were telling; gesture interaction enhanced driver performance by 65%, with a significant reduction in user workload.

The culmination of this research, "Impact of In-Air Gestures on In-Car Task's Driver Distraction," published in April 2023, marked my understanding in gesture recognition technology for automotive environments. However, during comparative experiments, I noticed a decline in recognition accuracy under certain conditions, like rapid gesture sequences or partial finger obstruction. This observation led me to develop an enhanced gesture recognition network based on DenseNet architecture, integrating a DSTformer model for depth estimation and a Tree-Structured Skeleton Image (TSSI) to maintain spatial hierarchy. The implementation of Global Long-Sequence Attention (GLA) mechanisms was a game-changer, allowing the network to focus on critical spatio-temporal details. This innovation resulted in the GLA-DenseNet network achieving top-tier performance, with accuracies of 82.35% on WLASL-100, 94.56% on AUTSL, and 98.87% on LSM,

rivaling state-of-the-art methods.

The paper of “GLA-DenseNet: An Advanced Convolutional Approach for ISLR Leveraging 3D Pose Estimation and Attention Mechanisms,” with me as the lead author, was a testament to my strides made in gesture recognition research. My journey at the CSSC lab has been a blend of challenge and innovation, driving me to push the boundaries of technology in HCI. It has set a solid foundation for my future endeavors in this exciting field.

### **Professional Experience**

Transitioning from theoretical concepts to real-world applications, I had the opportunity to lead a team in the prestigious 7th National University Intelligent Transportation Competition. Our project’s goal was to develop an advanced digital twin system designed to model and analyze traffic scenarios at crossroads. Faced with the inherent challenges of limited visibility and detection constraints, our team employed an innovative approach. We integrated data from surveillance cameras and radar systems, utilizing advanced techniques like longitude and latitude mapping and object pose estimation to enhance our system’s accuracy. This integration of different data sources and technologies was highly effective, as evidenced by our system achieving an impressive 93% accuracy rate in pinpointing automobile locations at crossroads.

The victory in this competition opened the door to an enriching internship at SUPCON, a leader in intelligent systems. Here, I was part of the ambitious “New Mirror City” project, where I applied my expertise in AI and machine learning to improve urban traffic management. This experience was invaluable in understanding how advanced technology can be leveraged to enhance societal well-being.

### **STEM Outreach**

My journey has been complemented by active engagement in various outreach initiatives. As a member of the teacher assistant team at College Shangxue Group, I mentored around 100 freshmen in C++ programming, creating detailed notes on essential concepts and helping students overcome coding challenges. This role also involved acting as a liaison between students and faculty, significantly enhancing the C++ curriculum based on student feedback. These experiences not only honed my mentoring skills but also led to lasting friendships. As President of the ZJUT Yixin CS Club, I organized computer repair events, encouraging members to use their technical skills for community service. I also led the coordination of an e-gaming contest on campus, securing sponsorships from Tencent and Zhejiang Media Group, which enhanced my organizational and communication abilities.

### **Future Work**

In my doctoral research, I aim to significantly advance the fields of machine learning and human-robot interaction. Central to my ambition is the enhancement of robots’ video perception and their interactive capabilities, emphasizing the refinement of goal-conditioned planning mechanisms. By integrating state-of-the-art object-centric representations and leveraging multi-view perception, my objective is to elevate neural networks’ efficiency in interpreting human gestures, thereby enabling robots to perform more intuitive and human-like manipulations. My research will prioritize developing sophisticated models that facilitate natural interaction patterns, aiming for a seamless and intuitive collaboration between humans and robots. Through this, I aspire to push the boundaries of current technology, making significant contributions to the enhancement of machine learning applications in robotics, particularly in areas requiring nuanced understanding of human actions and intentions.