# Hanzhang Cui

## Personal Information

---

**Status:** MS Student

**Program:** Computer Science and Engineering

**School:** Tandon School of Engineering, New York University

**RA Period:** From 2017-09 to 2018-11

## Biography

---

I'm a software development engineer at Amazon. Before that, I was a research assistant in NYU Multimedia and Visual Computing Lab, advised by Professor Yi Fang. I am broadly interested in 3D Computer Vision, Pattern Recognition and Deep Learning.

**Research Project:** Cross-Safe: A Computer Vision-Based Approach to Make All Intersection-Related Pedestrian Signals Accessible for the Visually Impaired

# 1 Description

Intersections pose great challenges to blind or visually impaired travelers who aim to cross roads safely and efficiently given unpredictable traffic control. Due to decreases in vision and increasingly difficult odds when planning and negotiating dynamic environments, visually impaired travelers require devices and/or assistance (i.e. cane, talking signals) to successfully execute intersection navigation. The proposed research project is to develop a novel computer vision-based approach, named Cross-Safe, that provides accurate and accessible guidance to the visually impaired as one crosses intersections, as part of a larger unified smart wearable device. Subsequently, experimental results show robust preliminary findings of our detection and recognition algorithm.

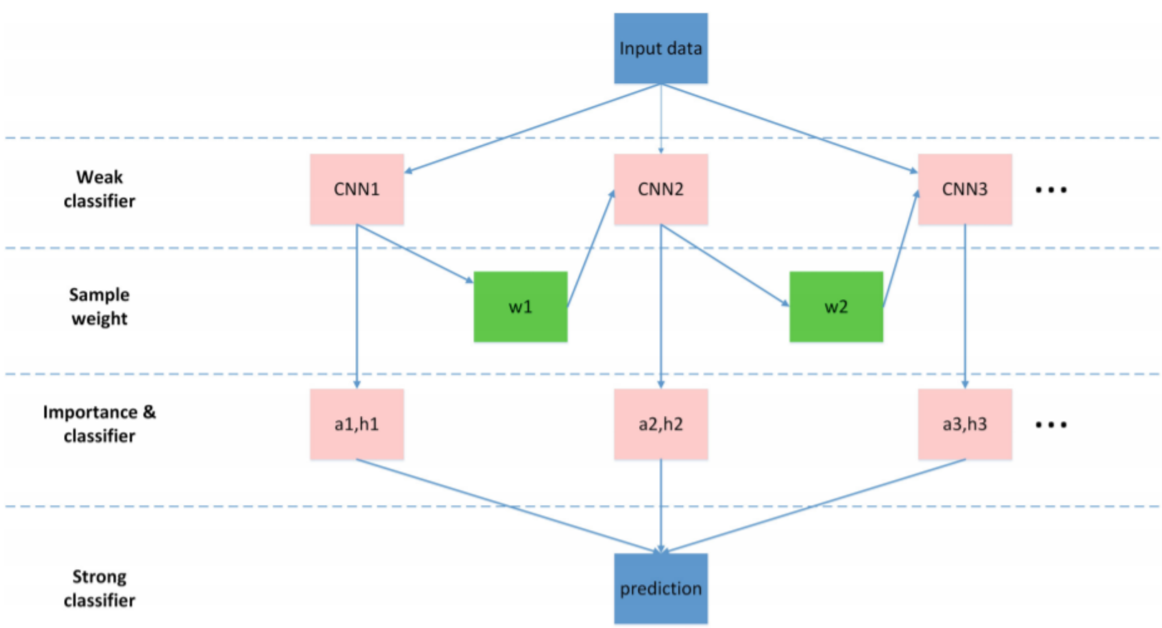


Figure 1: Final End-to-End Hand Pose Estimation Pipeline.

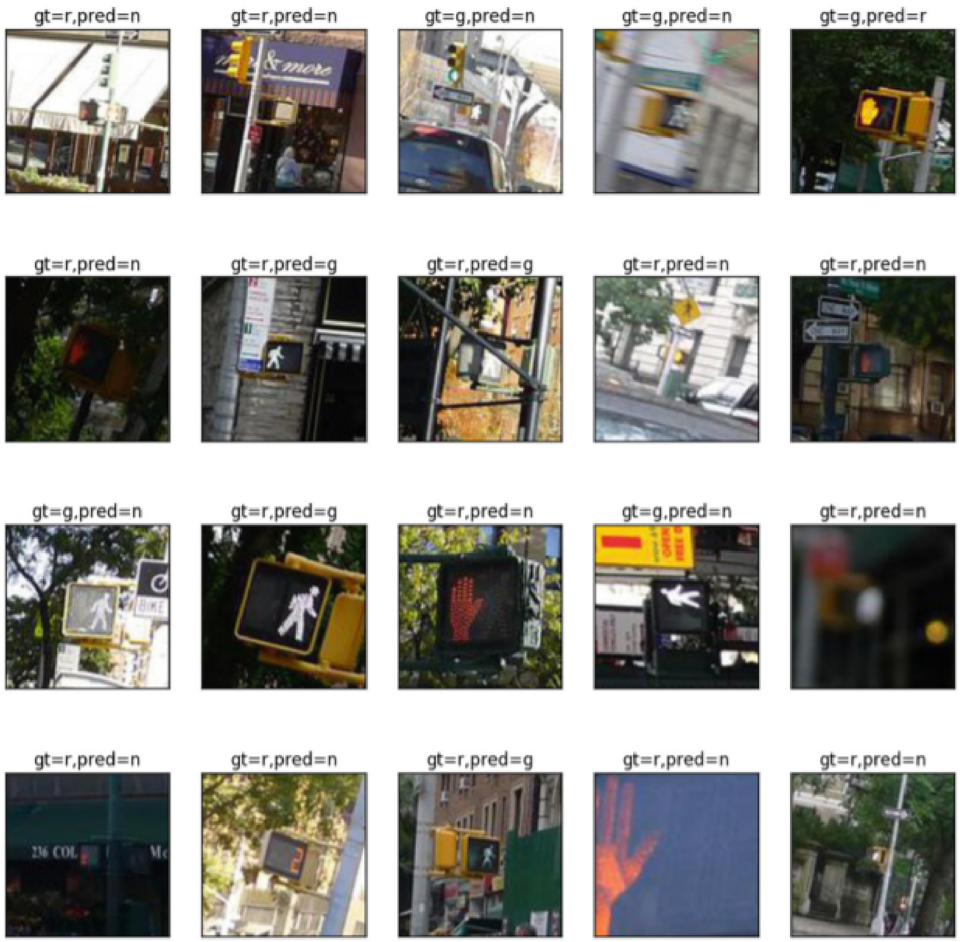


Figure 2: NYU Hand Dataset.

## 2 Method

In this project, we develop a novel computer vision-based approach, named Cross-Safe, that provides accurate and accessible guidance to the visually impaired as one crosses intersections, as part of a larger unified smart wearable device. As a first step, we focused on the red-light-green-light, go-no-go problem, as accessible pedestrian signals are drastically missing from urban infrastructure in New York City. Cross-Safe leverages state-of-the-art deep learning techniques for realtime pedestrian signal detection and recognition. A portable GPU unit, the Nvidia Jetson TX2, provides mobile visual computing and a cognitive assistant provides accurate voice-based guidance. More specifically, a lighter recognition algorithm was developed and equipped for CrossSafe, enabling robust walking signal sign detection and signal recognition. Recognized signals are conveyed to visually impaired end user by vocal guidance, providing critical information for real-time intersection navigation. Cross-Safe is also able to balance portability, recognition accuracy, computing efficiency and power consumption. A custom image library was built and developed to train, validate, and test our methodology on real traffic intersections, demonstrating the feasibility of Cross-Safe in providing safe guidance to the visually impaired at urban intersections. As displayed in Figure.1, The overall pipeline contains a successive of CNN networks as weak classifier, each classifier CNN is trained independently, and takes as input both input data and the sample weight calculated from its preceding classifier CNN. The importance of each classifier and output probability hi are summarized to get the final prediction.

## 3 Results

In this section, we compare the result of this research a large-scale wild scene dataset. Our dataset contains 3,693 images collected by volunteers from New York City with Zed camera. Each image has a size of 900x1200 pixels, and most images contain one traffic light located near the center. Figure.2 shows selected examples of prediction errors. 'gt' indicates ground truth, 'pred' indicates predicted label. 'g' for green light, 'r' for red light, 'n' for no light. One should note that among all prediction errors, (1, 1), (1, 2), (1, 3), (2, 1), (2, 4), (2, 5), (3, 5), (4, 1) are caused by blurry images, and (1, 5), (2, 2), (2, 3), (3, 2), (4, 3), (4, 5) may be caused by labeling errors. Table 1 shows the confusion matrix for the patches in the validation set (2154 patches in total). The rows stand for the ground truth labels, and the columns indicate the predicted labels.

| | Red | Green | No light |
|---|---|---|---|
| Red | 479 | 4 | 16 |
| Green | 2 | 300 | 5 |
| No light | 3 | 3 | 1342 |

Table 1. Comparing on NYU Dataset.