# Yubo Pan

## Personal Information

---

**Status:** MS Student

**Program:** Computer Science and Engineering

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

**Website:** https://www.linkedin.com/in/yubopan/

**RA Period:** From 2016-07 to 2016-12

## Biography

---

**I'm a software engineer at Bloomberg LP. 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:** Sensor fusion for Ecologically Valid Obstacle Detection and Identification for Visually Impaired

# 1 Description

Sensor fusion represents a robust approach to ecologically valid obstacle identification in building a comprehensive electronic travel aid (ETA) for the blind and visually impaired. A stereoscopic camera system and an infrared sensor with 16 independent elements is proposed to be combined with a multiscale convolutional neural network for this fusion framework. While object detection and identification can be combined with depth information from a stereo camera system, our experiments demonstrate that depth information may be inconsistent given material surfaces of specific potential collision hazards. This inconsistency can be easily remedied by supplementation with a more reliable depth signal from an alternate sensing modality. The sensing redundancy in this multi-modal strategy, as deployed in this platform, may enhance the situational awareness of a visually impaired end user, permitting more efficient and safer obstacle negotiation.

# 2 Method

In this project, we presented a sensor fusion framework for an electronic travel aid (ETA) or a smart service system for spatial intelligence and on-board navigation that has the potential to provide real-time situational and obstacle awareness in one's immediate environment, allowing blind and visually impaired individuals to travel more safely in three-dimensional (3D) space. As displayed in Figure.1, our neural network architecture is a single unified network consisting of a main trunk and four different proposal sub-networks connected to it at different scales in addition to the detection network. The network detects objects at four different scales. The region proposal network outputs a set of rectangular object proposals with their objectness score (class probability). The idea behind using multiple proposal networks at different scales is that object proposals at multiple receptive fields are obtained by the network. The first proposal subnetwork consists of a convolution layer (since the sub-network is close to

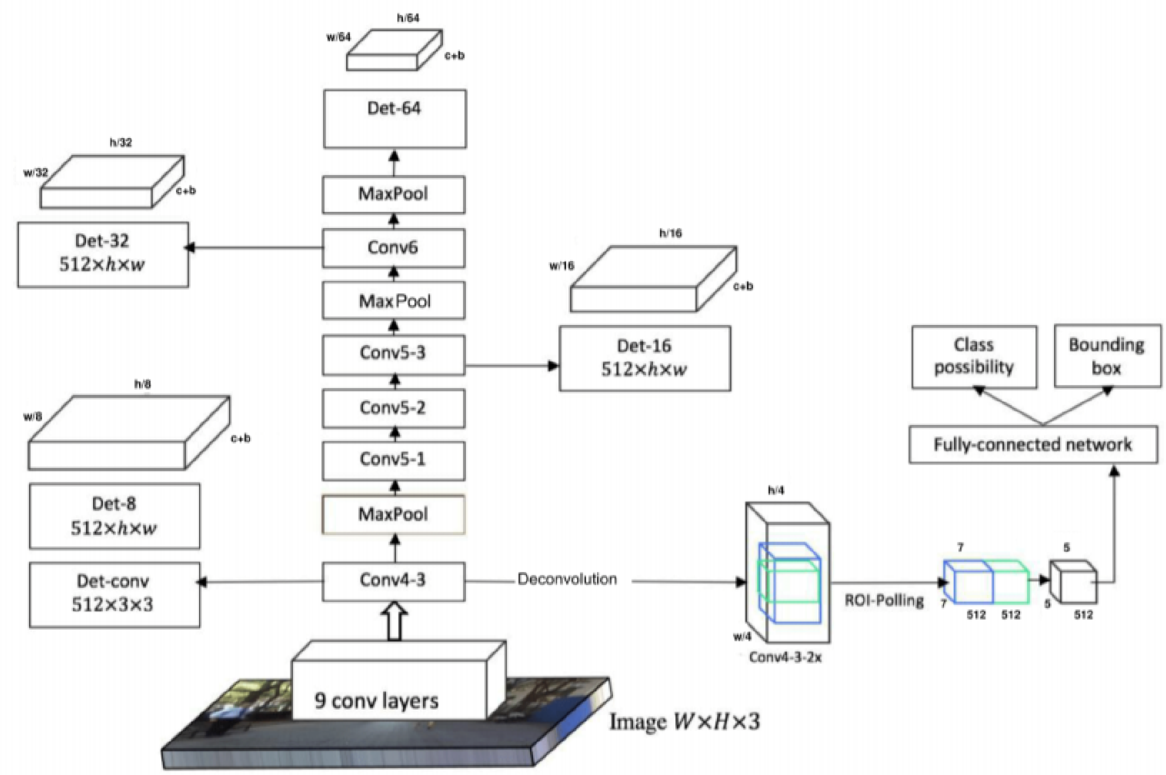


Figure 1: Convolutional neural network architecture for object detection in RGB domain.

lower layers of convolution of the main trunk which may cause instability due to the sub-network being closest to the main trunk) followed by a detection layer which maps the convolution map into a fixed feature vector. These feature vectors are fed to two fully connected networks which output the class label and the proposal bounding box coordinates. The other three proposal sub-networks have similar detection layers which output the class label and its corresponding bounding box coordinates. The proposals are given by the sliding window paradigm. The proposal network itself can be used as a detection network but a separate detection network is introduced to improve performance. As shown in Figure.2, the IR sensor detects distance values for each of the 16 independent segments. Those values are mapped back to the image captured by the stereo camera. The IR sensor detects object distances based on the reflected infrared. If an object is further than the effective range of the sensor, no value will be returned. In such situations, we will use the depth value from the stereo camera as distance value. Otherwise, our fusion algorithm will merge the values from the two sensors and fairly give higher weight to the IR sensor. Using this approach, we can get fused and reliable depth information for those 16 segments.

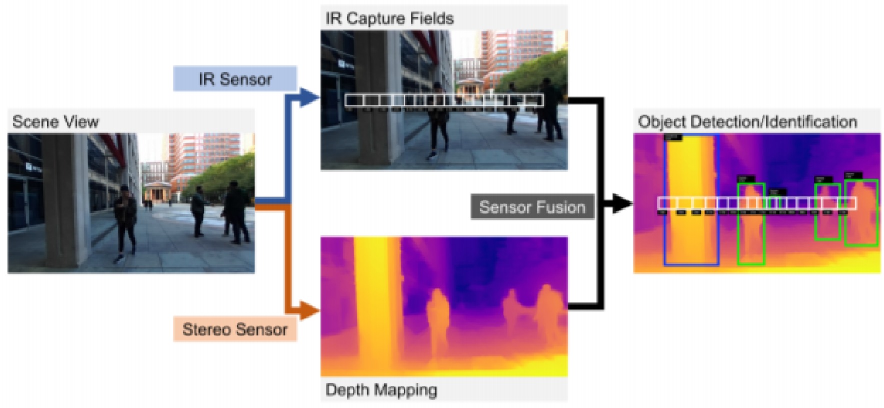


Figure 2: Sensor fusion framework.

# 3 Results

In this section, we used our sensor fusion framework as shown in Figure.2 for object detection in indoor and outdoor environments. We carried the sensor setup with a portable battery source and saved the RGB image data, depth values from stereo camera and distance values from the IR sensor simultaneously for several 10 second clips. Figure.3 and Figure.4 show the raw RGB image (on the upper left), and the distance array detected by the IR sensor, bounded in 16 boxes horizontally. The size of each box in the image is calculated by the physical parameters of the IR sensor and the stereo camera. Those distance values are then fused with the depth map from the stereo camera as shown in the lower left. We use a more consistent color map to render depth image where light yellow corresponds to objects closer to the sensor and dark purple corresponds to objects that are further away. Our object detection network not only can detect distant objects, but as shown in the blue bounding box in the lower right of Figure.4, it can detect nearby obstacles like the structural column as shown in the figure. The column was detected by the IR sensor even it was not one of the objects in our training set.

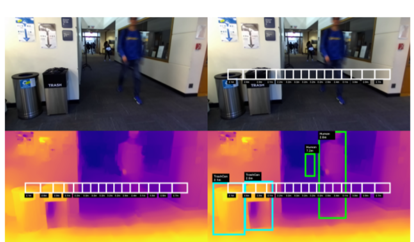

Figure 3: Indoor object detection using sensor fusion.

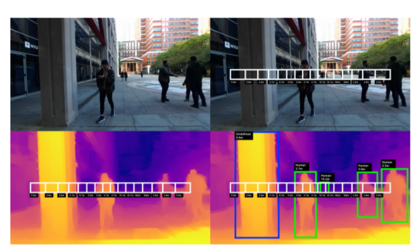

Figure 4: Outdoor object detection using sensor fusion.