This is a summary for the paper titled "Small Traffic Sign Detection and Recognition in High-Resolution Images", by You et al. published in the Cognitive Computing – ICCC 2019. In this work, the authors start off by stressing the importance of traffic sign detection and recognition systems which can help reduce traffic accidents. The traffic signs need to be detected in advance as speed of vehicle increases, rendering the traffic signs relatively small and difficult to detect. The authors come up with an end-to-end deep learning model which detects and recognizes small traffic signs in high-resolution images. Performance of the model is tested on two recently published benchmarks, Tsinghua-Tencent 100K (TT100K) and Chinese Traffic Sign Dataset (CTSD).

Convolutional Neural Network (CNN) methods achieve near-perfect results on commonly used benchmarks like the German Traffic Sign Detection Benchmark (GTSDB) and the German Traffic Sign Recognition Benchmark (GTSRB), due to the ability of CNNs to generate more discriminative features than previously used methods based on color segmentation and shape detection. However, these benchmarks lack in a few aspects for use in training a model for detection of small traffic signs. For starters, the GTSDB contains only 900 images and GTSRB contains only images of traffic signs. While these benchmarks maybe effectively utilized to train for traffic sign detection and classification in close distances, the reaction time required of drivers at those distances is very short.

Traffic signs in the TT100K are approximately 20*20 pixels in images with resolutions of 2048*2048, while CTSD has slightly larger proportions for the traffic signs, and half the resolutions for images. TT100K has 10,000 images collected from urban and suburban areas of 10 regions in China, with a total of 30,000 traffic sign instances consisting detailed annotations for each sign. CTSD contains 1100 images. The performance of the proposed method is evaluated on the TT100K dataset and CTSD is then used to evaluate the generality of the model.

The authors describe their proposed network as having an architecture like Fully Convolutional Networks (FCN), but with several modifications to reduce the parameters, thus improving performance. The first half of the network does feature extraction using VGG network. The first few layers of convolutions are performed with small kernels of sizes 1*1 and 3*3, but the last layer has a bigger kernel with a size of 6*6, which is used for learning more contextual information. The last two layers of convolutions are fused together, in what the authors refer to as the feature fusion strategy. This strategy is helpful for detecting small traffic signs, while also ensuring no drop in accuracy for detection of larger signs or accuracy-loss of the general classification occurs. For the second half of the network, another strategy that the authors propose for increasing detection recall rate is called the Hierarchical Multi-task Learning. This strategy takes into consideration the fact that larger numbers of convolutions produce smaller feature maps, which while conducive to classification tasks, has negative effects on detection of small traffic signs which rely on larger feature maps. Therefore, the detection tasks are based on relatively larger feature maps, while classification is carried out on higher feature maps. This strategy improves the accuracy and recall rates by 3% each.

For experimental results the authors choose to first test the performances of several classic models on the TT100K dataset for identifying the weaknesses of applying these models without the suggested strategies deployed in this paper. The tested models include YOLO, Fast-RCNN, Faster-RCNN, and R-FCN. The results show that general object-detection models perform poorly for purposes of small sign detection. Comparisons of recall and accuracy rates of the different models against traffic signs of different sizes in TT100K reveal that the proposed network outclasses all previous models with recall of the proposed model reaching as high as 94%, and an accuracy of 91%. The generality tests conducted with CTSD indicate rates of recall as 95% and accuracy of 97%.

In conclusion, the authors state that small traffic signs under different road conditions can be effectively detected by the proposed method. The limitation of the model is noted to be its time-performance as the model cannot meet real-time requirements, since the high resolution of images need a lot of processing power.