This is a summary for the article “Traffic sign detection method based on Faster R-CNN” by Wu et al. in J. Phys.: Conf. Ser. Like the title suggests, the paper utilizes Faster R-CNN to detect and classify road traffic signs using the German Traffic Sign Detection Benchmark (GTSDB) dataset for evaluation of performance.

The authors stress the importance of traffic sign recognition systems to autonomous vehicles. Many solutions exist for detection of traffic signs that are based on traffic sign colors, shapes, and deep learning. Color-based methods can differentiate traffic signs from background using color space information to segment images, and provide real-time performance, but perform poorly in suboptimal conditions influenced by light and weather condition variances. Detection methods based on shapes of signs usually utilize the Hough Transform algorithm to detect shapes of interest, but involve complex calculations making them not suitable for real-time performance. Shape-based methods are also limited to few shapes that can be easily optimized for.

Traffic sign detection algorithms based on deep learning models can extract features automatically, which are highly robust to changes and are not dependent on heavy domain knowledge of the researchers, unlike the other methods. Faster R-CNN, the model used in this work, is one of the deep learning models that is used for object detection and is based on improving the previous model, Fast R-CNN, which in turn was an improvement over R-CNN.

Faster R-CNNs utilize a Region Proposal Network (RPN) which is used to extract regions of interest from the images. The proposed regions are generated by CNNs, passed to the RPN and to a classifier simultaneously where the classifier has a Region of Interest (RoI) pooling to pool together the detected regions for finding meaningful regions that may be of interest to the model being developed for. Regression is used to calculate bounding boxes for the proposed RoIs, before any classification can take place. At the last stage, a Softmax function is used for classification and boundary regression.

The proposed model’s performance is evaluated on GTSDB, a popular benchmarking dataset for detection of road signs. The GTSDB dataset has 900 images, all with 1360*800 pixels resolutions, and includes 1213 instances of traffic signs. The signs are divided into four categories; prohibitory, danger, mandatory, and other. For training, the VGG16 network, which is a publicly available model trained on the ImageNet dataset, is used for pre-training the proposed model to initialize the CNN part of the network. Mean Average Precision (mAP) is used to evaluate the performance of the model. The authors describe the results as good, and state that the model performs with mAP of 91.75%. They also note that the model’s results indicate robustness to different lighting, motion blur, sign occlusion, and size sign variances.