Telling "what is where", object detection is a fundamental problem in computer vision and has a broad range of applications such as video surveillance and autonomous driving. One major challenge of object detection comes from large intra-category appearance variations which are caused by factors including subcategory, viewpoint, deformation and occlusion. Large appearance variations make it difficult to model an object category properly such that the object category is well distinguished from other object categories as well as backgrounds. Appearance variations resulting from subcategories and viewpoints are usually handled by clustering object examples of a category into groups each of which represents one subcategory or viewpoint and then learning a detector for each group. For dealing with appearance variations due to deformations and occlusions, part-based detection methods have demonstrated their promise by integrating a set of part detectors to form a part detector ensemble. This thesis studies how to learn detector ensembles to better address deformations and occlusions, particularly the latter situation, for generic object detection as well as pedestrian detection.

For generic object detection, two approaches are proposed to handle deformations and occlusions respectively based on a classic part detector ensemble, deformable part model (DPM). The former discovers a set of non-rectangular parts which can well fit object structures to replace the original rectangular parts in the DPM. The discovered non-rectangular parts can better capture the appearance of local regions and structural deformations of objects. The latter discovers a set of representative and discriminative occlusion patterns which share the same set of parts from a DPM trained on fully visible object examples. The discovered occlusion patterns are themselves DPMs, and when properly tuned, can be applied directly or combined with state-of-the-art detectors, e.g. Faster R-CNN for improving detection performance and achieving part-level occlusion reasoning.

For pedestrian detection, two approaches are developed to improve two modules of a commonly used framework of learning a part detector ensemble respectively for handling occlusions. The first approach focuses on how to integrate part detectors properly to reduce negative effects from unreliable and irrelevant part detectors on heavily occluded pedestrian detection. The second approach aims to learn reliable part detectors jointly by sharing a set of decision trees among the part detectors to exploit part correlations and also reduce the computational cost of applying these part detectors. Experimental results on pedestrian detection benchmark datasets show promising performance of the two approaches for detecting partially occluded pedestrians, especially heavily occluded ones.