Weakly Supervised Object Detection

Detecting objects in images is important for robots because it provides them with rich knowledge about the parts of the environment with which they may interact. Accurate localisation of these objects is especially important if a robot intends to grasp or manipulate the item. While state-of-the-art object detection methods have achieved impressive results for classifying, localising and even precisely segmenting multiple objects in individual images [1], a significant drawback is that they rely on large training datasets containing detailed annotations. Even for bounding boxes, the annotation task is tedious and prevents a robot from quickly learning new objects in new environments. To alleviate this requirement, weakly supervised methods learn tasks without explicit annotation. That is, annotation is only required at the image level instead of the pixel level. For object detection, it has been shown that just a list of object categories per image, without segmentation masks, bounding boxes or centroids, is sufficient for learning [2].

The aim of this project is to develop a pipeline to enable a robot to quickly learn object detection by leveraging weak supervision. Specifically, a method will be developed to acquire data from a hand-held RGB camera and with a robot to train a state-of-the-art weakly supervised object detector.

Tasks

- Review the literature of weakly supervised object detection and from the analysis select the most promising methods that can be applied within this project then replicate the results of the chosen method(s) using the same training and testing datasets.
- Apply the chosen method(s) to the Autonomous Robot Indoor Dataset (ARID) [3] and the Object Clutter Indoor Dataset (OCID) [4].
- Acquire a new training dataset from a hand-held RGB camera with overlapping instances and classes of ARID and OCID. Implement a GUI that enables a human to very quickly annotate the training images. Finally, train the model(s) and evaluate their accuracy for object detection using ARID and OCID as ground truth (test set). Importantly, investigate how many and how diverse the training set needs to be to achieve reasonable performance.
- Implement a patrolling behaviour on a mobile robot so that it collects images from different viewpoints in an office environment. Using the developed GUI, label the robot acquired dataset, then train and evaluate the performance of the object detection method(s).

Workload split

- Research and theory: 30%
- Programming and implementation: 50%
- Writing: 20%

Contact
Timothy Patten, patten@acin.tuwien.ac.at

References