A probabilistic challenge for object detection

To safely operate in the real world, robots need to evaluate how confident they are about what they see. A new competition challenges computer vision algorithms to not just detect and localize objects, but also report how certain they are.

Visual object detection provides answers to two questions: what is in an image and where is it? Object detection is often an important part of the perception system of robots or autonomous systems such as driverless cars. It provides crucial information about the robot's surroundings and has significant influence on the performance of the robot in its environment. For example, driverless cars need object detection to be aware of other cars, pedestrians, cyclists and other obstacles on the road. Future domestic service robots and robots in healthcare will have to be able to detect a large range of household objects in order to properly fulfil their tasks.

Failures or mistakes in object detection could lead to catastrophic outcomes that not only risk the success of the robot's mission, but potentially endanger human lives. We can distinguish four major types of object detection failures: failing to detect an object; assigning a wrong class label to a detected object; badly localizing an object — for example, only detecting parts of it; and erroneously detecting non-existing objects. Despite a lot of progress over the past years, today's state-of-the-art object detectors are still prone to all of these failure types. Why is that? One reason is that existing detectors are not built to express uncertainty, which often leads to overconfident detections.

State-of-the-art visual object detectors are based on deep convolutional neural networks (CNNs) that localize objects in an image by predicting a bounding box and providing a class label with a confidence score, or a full label distribution, for every detected object in the image.

For safe and trusted operation in robots or autonomous systems, CNNs must express meaningful uncertainty information. Object detectors will have to quantify uncertainty for both the reported labels and bounding boxes (pictured). However, while state-of-the-art object detectors report at least an uncalibrated indicator of label uncertainty via label distributions or label scores, they currently do not report spatial uncertainty. Evaluating the quality of the label or spatial uncertainties is not even within the scope of typical computer vision benchmark measures and competitions.

To overcome this gap, we defined the new task of probabilistic object detection,
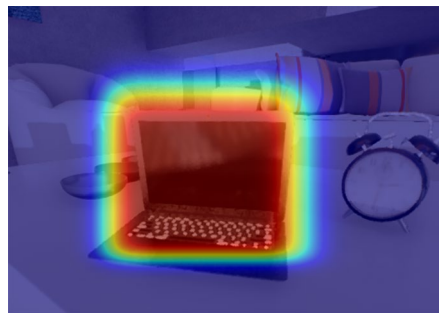
The output of a probabilistic object detector encodes its spatial uncertainty: red indicates the detector is very certain these pixels belong to the object (laptop), blue highly certain they do not belong to the object. Reproduced from ref. [1].

along with a new evaluation measure[1] and a challenging dataset[2]. To motivate the international research community to address this new problem, we organized a competition and called for participants around the world to benchmark their approaches. We presented the results of the competition during a workshop we organized at this year's Computer Vision and Pattern Recognition (CVPR) conference.

We created a new dataset consisting of video footage recorded in a high-fidelity simulation: we simulated three different domestic robots operating in three different indoor environments, under daytime and night-time conditions. From the resulting 18 video sequences, we extracted 56,513 frames with their ground truth annotations (camera poses, object classes, bounding boxes and object IDs). In addition, we created further validation and development test datasets consisting of 145,195 images.

We received 111 submissions from 18 participating teams from around the world. These submissions were evaluated automatically on our evaluation server hosted via CodaLab. We invited the best four participants to write a short paper and present their approaches and results at our workshop at CVPR. From the results, papers and presentations, we concluded that there currently exist only very few approaches that allow object detectors to express spatial uncertainty, and that those methods are still in their infancy.

As was evident by the evaluation results, uncertainty techniques like Monte Carlo dropout show promise when combined with existing object detectors, but a lot of research work is required to develop mature methods for probabilistic object detection. The best methods only reached scores of 22% in our specifically developed performance measure[1], leaving a lot of room for improvement in the future.

We believe competitions are a fantastic way of highlighting important research problems, enable measurable progress, and help ensuring research is comparable and replicable. Large competitions like ILSVRC (ImageNet Large Scale Visual Recognition Challenge) and COCO (Common Objects in Context) have supported much of the remarkable progress in computer vision and deep learning over the past years, and we aim to recreate this success for robotic vision. We are organizing various challenges and are working on a new competition that focuses on robotic scene understanding and active vision (more information will appear on our website).

The next round of our probabilistic object detection challenge will run later this year, and the results will be presented in a workshop at the IEEE International Conference on Intelligent Robots and Systems (IROS) in November. We expect to see submissions based on more advanced and mature approaches achieving higher scores. ❑

Niko Sünderhauf[1,2]*, Feras Dayoub[1,2], David Hall[1,2], John Skinner[1,2], Haoyang Zhang[1,2], Gustavo Carneiro[1,3] and Peter Corke[1,2]

[1]Australian Centre for Robotic Vision, Brisbane, Queensland, Australia. [2]Queensland University of Technology, Brisbane, Queensland, Australia. [3]University of Adelaide, Adelaide, South Australia, Australia.
*e-mail: niko.suenderhauf@qut.edu.au

References
1. Hall, D. et al. Preprint at https://arxiv.org/abs/1811.10800 (2019).
2. Skinner, J. Hall, D., Zhang, H., Dayoub, F. & Sünderhauf, N. Preprint at https://arxiv.org/abs/1903.07840 (2019).