University Defence Research Collaboration in Signal Processing

Edinburgh Consortium White Paper

Improving Reliability of Object Detectors

Introduction

Many video and image processing tasks rely on the ability to accurately and consistently detect objects such as people within an image or video stream. These tasks form part of a wide range of defence and civilian applications, such as automated surveillance of targeted and wide areas, navigation for autonomous vehicles, driver assistance, augmented reality entertainment and many more situations. These applications are becoming ubiquitous.

Object detectors typically work by processing the pixels within a region of interest in order to extract relevant features including shape, colour, or texture. These features are then classified using a machine learning technique such as support vector machines (SVMs), decision trees (Adaboost) or deep learning approaches such as convolutional neural nets (CNNs). These methods all use an object model, learned during a training phase, which expresses a representation of how the object is expected to appear. The features calculated from a new sample are then compared to the object model. This produces a binary yes/no decision about whether the expected object is present in that region. As well as a binary decision, the classification stage can produce a confidence score. Using additional training samples, this can be 'squashed' or converted to a probability value which expresses the classifier's belief that an object (such as a person) is present. This allows comparisons between different classifiers to be made, and probabilistic detection scores can be passed to subsequent higher-level data interpretation algorithms as part of a larger task.

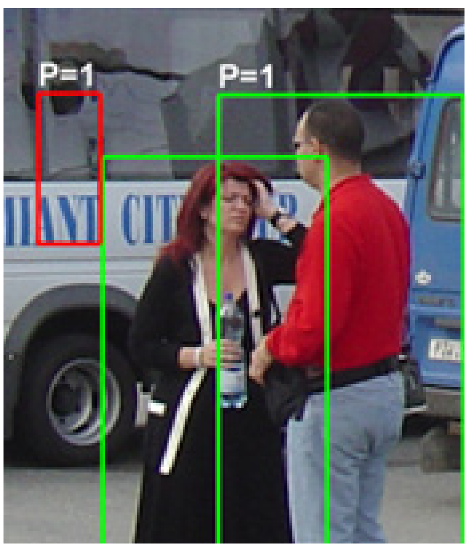

Figure 1: An overconfident person detector algorithm. True positives (green boxes) and false positives (red boxes) are both detected with 100% confidence.

Many of the recent advances in detection capability have focused on improving the accuracy of the binary decision, attempting to reduce false positive detections (background wrongly identified as an object of interest) and false negatives (vice versa). As part of this, the confidence score output is often overconfident; state-of-the-art detectors are significantly more confident about the presence or absence of an object in a region than they should be, and areas of extreme uncertainty (around 50% probability that an object is present at all) are significantly under-represented. This makes object detectors unreliable; if a higher-level task or human operator trusts or relies on the probabilistic detection, and a detector returns a true positive and a false positive with 99% confidence in both cases, the detection algorithm is of less use; see Figure 1 for an example. This is a particular problem in scenarios where some level of assurance is needed (e.g. mine countermeasures).

Method

A desirable goal in this case is to improve the reliability of existing state-of-the-art object detectors while maintaining accuracy (the number of correct and incorrect detections of objects). We define *reliability* as a measure of *how closely the confidence expressed by a detector in its result matches the real-world outcome of that prediction*. At the same time, we must keep in mind computational complexity, and ensure that increasing reliability does not mean that image processing takes an unacceptably long time.

Improved classification techniques such as the Gaussian Process Classifiers (GPCs) can be used to achieve these aims. These model the distribution of features in the two classes of interest (object and background) using Gaussian probability distributions and have been shown to more reliably identify ambiguous regions which contain uncertain detections, while achieving similar levels of accuracy [1]. They also directly generate probabilistic classifications. However, GPCs require many more computations than equivalent techniques (Adaboost or SVMs). To mitigate this, we can use faster classifiers (Adaboost) to process the whole image and generate preliminary confidence scores. High-scoring regions which may contain objects of interest are then processed with the GPC to generate accurate, reliable detections in a fraction of the time taken if GPCs were used to process the entire image. In addition, we have accelerated GPC calculations using GPUs (Graphics Processing Units) to allow further reductions in processing time. [2]

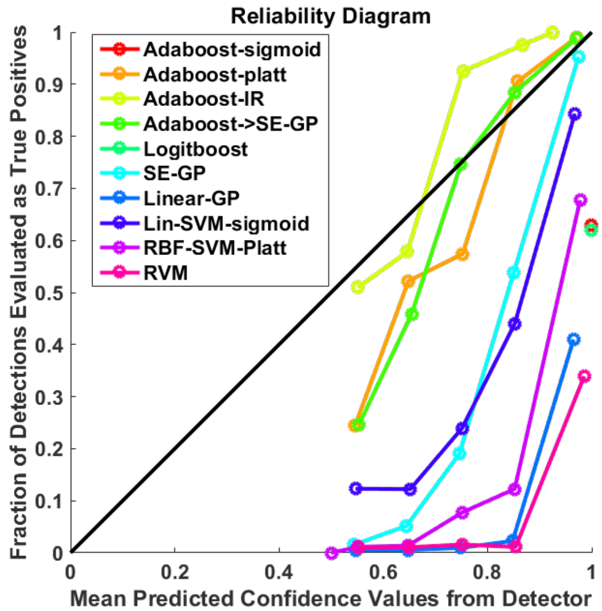

Figure 2 Reliability Diagram: Classifiers closer to the black line produce detection confidence scores which match ground-truth data (presence or absence of objects) more reliably.

Results

The reliability diagram in Figure 2 shows a comparison of several state-of-the-art classifiers. An ideal classification algorithm, lying on the black line, would be *well-calibrated*; if an ideal classifier processes a number of regions of interest then, for example: 70% of the time that the detector produces a confidence level of 0.7, the region will actually contain a target (*a true positive detection*). (Conversely, a classifier with poor reliability may only identify a true detection 50% of the time it produces a score of 0.7 (*over-confident*) or, say, 90% of the time (*under-confident*)."Detectors which lie closest to the line are therefore more reliable, and our approach (using Adaboost followed by GPC) performs best here.

In conclusion, reliable object detection in video and imagery is a challenge with many civilian and security applications. However, applying a fast and relatively accurate detector (Adaboost) followed by using a slower, more introspective GP classifier allows a substantially faster detection rate coupled with a significant gain in accuracy and reliability. This can be utilised effectively throughout many sensing modalities to perform monitoring or surveillance tasks which rely on object detection.

References and Further Reading:

[1] C. G. Blair, J. Thompson, and N. M. Robertson, "Introspective Classification for Pedestrian Detection," in Sensor Signal Processing for Defence (SSPD 2014), 2014.

[2] C. Blair, J. Thompson, and N. M. Robertson, "GPU-Accelerated Gaussian Processes for Object Detection," in Sensor Signal Processing for Defence (SSPD 2015), 2015.