Object Detection: How Noise Patterns Affect the Performance of Convolutional Neural Networks

Simon B. Jensen
Software Development & Technology
IT University of Copenhagen
`sije@itu.dk`

1   Problem Formulation

In the last few years Object Detection challenges such as the Pascal Visual Object Classes Challenge (Pascal VOC)[1] and the ImageNet Large Scale Visual Recognition Competition (ILSVRC)[2] have seen tremendous increases in detection performances due to new development in Deep Learning methods. More specifically development in Convolutional Neural Networks (CNNs) - a specific branch of Deep Learning. The performance of the Object Detecting CNNs are now at a point where they have even surpassed human capabilities on these specific challenges. However, one can argue that the images of the challenges are not realistic as the images are ideally cropped, non-blurry and mostly only contain a few specific objects, for which the CNNs can be optimized for. The human visual system is still far superior when the settings are more realistic and non-ideal. What happens for example if we add different noise patterns to the images in the Object Detection Challenges? By adding noise patterns to the images we try to simulate a realistic and non-ideal setting. A non-ideal setting could for instance be a slightly blurred image, an incomplete image or a damaged image etc. Which we as humans typically easily can deviate from and still be able to identify what is in the image. For this Individual Project we will investigate how different noise patterns influences the detection performance of state-of-the-art object detection system.

2   Image based data sets: Pascal VOC

The following section introduces and describes the Pascal VOC data set and it's role as a benchmark data set. The section can be skipped if the reader is already familiar with the data set.

There exists only a handful of image based data sets which are openly available to the public, and where the images and corresponding labels are both of high quality and there are a substantial amount of them. Thereby containing the key properties for serving as a valid benchmark data set. The Pascal VOC image data set is such a data set. Other such data sets are e.g. The ImageNet, COCO [3] and KITTI Vision [4]. What is significant by publicly available benchmark data sets is the fact that they, in opposition to strictly private or company owned data sets, indulges to good competition between competing factions and can serve as a fair basis for comparison. Competition and the possibility for comparison will typically motivate and provoke even faster development in an area. And this is exactly what the Pascal VOC data set has done, since the first Pascal VOC challenge started in 2005, for the area of Object Detection among others. It has since been held as a yearly competition, each year presenting new systems which pushed the boundaries for what is possible for computer vision systems. In this report we will be focusing on the Pascal VOC data set which is used for both experiments and evaluation. It should be noted that the Pascal VOC competition was last time held in 2012 and is no longer active. But it still serves as a valid basis for the tasks of this project.

---

[1] http://host.robots.ox.ac.uk/pascal/VOC/
[2] http://www.image-net.org/challenges/LSVRC/
[3] http://cocodataset.org/#home
[4] http://www.cvlibs.net/datasets/kitti/

2.1 The Pascal VOC data set and Object Detection

The Pascal VOC data set consists of numerous computer vision challenges, e.g. Image Classification, Object Detection, Instance Segmentation and even more advanced challenges such as Action Classification. In this project we focus on the Object Detection part of the data set. The Object Detection part of the data set consists of 20 different object classes. For example a person, a bird, an airplane, a chair etc. The goal of the challenge is to recognize both the class and the location of the object in the image. The location of an object is simply represented by the coordinates of a 2D bounding box. The images of the data set depict realistic scenes and have one or several object(s) of a specific class in them.

2.1.1 Statistics of Pascal VOC

3 State-of-the-art Object Detection systems: SSD and YOLO

The following section introduces and describes the Convolutional Neural Network based Object Detection systems, SSD [1] and YOLO[2], which are used and tested in during the project. The systems were, as of 2015-2016, among the state-of-the-art in Object Detection systems and are still today some of the top performing systems.

3.1 SSD: Single Shot Detection

3.2 YOLO: You Only Look Once Detection

4 Noise patterns and Object Detection

The following section introduces different noise patterns and explains the relevance of noise patterns in relation to object detection systems.

4.1 Why do noisy images matter?

First of all, let us dwell upon the question of why we are interested in testing how Object Detection system perform on noisy images? Let us consider the case of a vehicle which is driving around on a rainy day. When a human driver is sitting behind the steering wheel the rainy weather unquestionably means it becomes harder for the driver to detect objects he or she sees out of the windows. The rainy weather makes the outside world look different from how it usually looks on a lovely sunny day. Perhaps the traffic light lights and the other cars on the road have become blurry and perhaps a raindrop on the car window means a person in a red rain coat, in the distance, looks more like a big round blob of paint than a person. But the human visual perception system is nevertheless still able to deviate from these changes and we can perform the Object Detection task well enough for us to be able to safely navigate around in our cars, even in rainy weather. Other weather circumstances can likewise change how the world is looking. Consider for instance snowy weather, foggy weather etc. So the reason why we are interested in how Object Detection systems perform on noisy images is to figure out how robust the systems are towards small changes in how the input looks.

4.2 Noisy Patterns

Another relevant question to ask is what kind of noise patterns will serve as realistic? For us to able to answer this question we analyze and compare some different noise patterns.

References

[1] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott E. Reed, Cheng-Yang Fu, and Alexander C. Berg. SSD: single shot multibox detector. *CoRR*, 2015.

[2] Joseph Redmon, Santosh Kumar Divvala, Ross B. Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. *CoRR*, 2015.