IOP Conference Series: Materials Science and Engineering

**PAPER** • **OPEN ACCESS**

# Automated vehicle detection in satellite images using deep learning

To cite this article: Ahmad Mansour *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **610** 012027

## Recent citations

# Automated vehicle detection in satellite images using deep learning

**Ahmad Mansour[1], Ahmed Hassan, Wessam M Hussein and Ehab Said**

[1]Department of Mechatronic Engineering, Military Technical College, 11766, Cairo, Egypt

[1] E-mail: ahmad.mansour_44@yahoo.com

**Abstract**: Automatic detection of small objects such as vehicles in satellite images is a very challenging task, due to the complexity of the background, vehicles colors, the large size of ground sample distance (GSD) for satellite images and jamming caused by buildings and trees. Many methods were proposed for this task by using handcrafted features (such as a Histogram of an Oriented Gradient, Local Binary Pattern, Scale-Invariant Feature Transform, etc.) along with support vector machine classifier, however, Convolutional Neural Networks (CNN) have proved to be potentially more effective. In this paper, we use two advanced deep learning frameworks, Faster Region CNN (Faster R-CNN) and Single Shot Multi-Box (SSD) based on (CNN) with Inception-V2 as a feature map generator instead of VGG-16, to detect vehicles through Transfer Learning, and making an experimental analysis comparison between the two models. Experimental results on the test dataset demonstrate the effectiveness and efficiency of the proposed methods.

## 1. Introduction

Vehicles fleets in the world increase continually specially in cities. A lot of cities use field equipment like fixed location cameras or motion sensors on the traffic lights to monitor vehicles, recently cameras mounted on UAV used to provide a wider field of view [1-2]. Keeping up with this rapid increase of vehicles number researchers face a really challenging task, since the fixed location cameras and motion sensors are not widespread enough. The enumeration of this huge number of vehicles is required to improve traffic management, detect fuel demands in certain locations and to estimate the emissions in crowded areas in order to know the pollution percentage. Enumeration overtime periods also is important to estimate future Possible overcrowding to plan transportation infrastructure.

Satellite image represents an area on the earth captured by the satellite camera, which stored in groups of pixels arranged in matrices[3-4**]**, the value of each pixel represents the reflected amount of electromagnetic wave by certain location on earth, the area represented in one pixel called the spatial resolution of the image, and also known as the ground sample distance(GSD).

The captured reflected waves are separated into number of bands; this number of bands represents the spectral resolution of the image. The satellite temporal resolution indicates the periodic ability to shoot the same area [5]. Satellite images provide a good source for vehicles monitoring task, but it still faces some challenges such as the small size of detected object relative to the ground sample distance (GSD) of the image [6-8], the complexity and variety of backgrounds, different shapes and colors of vehicles, and the interference with trees and buildings.

In the last few years, optical sensors technology has been evolved exponentially in terms of quantity, quality, and applications. The internet is replete with Images, but the image itself is not useful without performing a proper analysis to extract useful knowledge from it. Hand-crafted features and classification are the most suitable methods to detect vehicle [9-11], however the hand-crafted features don’t serve a general solution, and the classifiers need some modifications to fit the different features, also using shallow neural network is used for vehicle detection [12], but the performance was not at the desired quality and the quality of the extracted features was not good enough. Handling this huge amount of data requires a new technique which is capable of performing fast, précised and reliable. With the progress of deep learning deep neural network (DNN) that has achieved fast progress in various processes such as object detection, and classification based on a convolutional neural network [13-14].

Convolutional neural networks (CNNs) have appeared on the stage as an application to visual tasks since the 1980s [15], but they still a bunch of scattered applications, they were dormant till the mid-2000s, when the developments in computing capabilities and the growth of large amounts of labeled data began Improving the algorithms contributed to (CNNs) advancement and brought them to the forefront of neural networks and still in rapid progression since then[16-17].

Many algorithms and architectures were proposed for object detection based on CNN. In recent years there were two series mainly for object detection based on deep learning. The first series is the combination of region proposal and CNN classification in the two-stage object detection framework, which is represented by RCNN [21], including SPP-NET [22], Fast R-CNN [23], and Faster R-CNN [24]. The second series is the object detection framework with a single stage, using a single convolution neural network which is represented by SSD (Single Shot Multi-Box Detector) [25] and YOLO (You Only Look Once) [26].

The presented study makes a comparison in the performance and evaluates Faster R-CNN and SSD in terms of accuracy and processing time while studying vehicle detection from satellite images.
The rest of the paper is organized as follows: Section 2 discusses related work about vehicle detection from satellite images. Section 3 provides an overview of Faster R-CNN model and SSD model. Section 4 presents the results of the algorithms for vehicle detection from satellite images. Finally, Section 5 presents the conclusion which highlights the main results.

## 2. Related work

A lot of updates appear all the time, and as a result, a lot of vehicle detection methods using deep learning appears, in a trial to reach better results in accuracy and processing time. Hybrid DNN (HDNN) [13], which presented by **Xueyun** Chen in 2014 depending on dividing the feature maps of the final convolutional layer and the max-pooling layer into multiple blocks of different sizes, in a try to detect variable-scale vehicles to increase the accuracy, but on other hand it increased the processing time. Another model presented by the same team in the same year using Parallel Deep Convolutional Neural Networks to detect vehicles in satellite images [18], by dividing the DNN into several parallel branches in the same size, and without using direct connections in lateral direction between branches, this results in slightly drop in accuracy with increasing the speed. Another idea presented in 2007 to detect vehicles in high-resolution satellite images by **XiaoyingJin**, Morphological Shared-weight Neural Networks [19] by using GIS road vector map, but this method constrained the detected vehicles to the roads map.

In October 2016 another method presented Called Fast Vehicle Detection which is presented to replace the fully connected layers in traditional CNNs with convolutional layers in order to decrease the number of parameters and enhance the processing time [20].

## 3. Theoretical overview faster R-CNN and SSD

### *3.1. Faster R-CNN*

The Faster R-CNN model is divided into three networks. The first network is the shared convolutional network (base network) for feature extraction to generate good features from the images, which uses a pre-trained CNN for the task of classification (e.g. ImageNet) [27]. This technique is very commonly used in the context of Transfer Learning, especially for training a classifier on a small dataset using the weights of a network trained on a bigger dataset. The base network used in this paper is the Inception-V2 for enhancing extracted features [12]. The second network is the region proposal network (RPN) for producing region proposals with multiple scales and aspect ratios from the input image. The third network is the Fast R-CNN detector, its input is the regions of interest (ROIs) that comes from RPN. Then the ROI pooling layer extracts for each ROI a feature vector. These features are then fed into Fully Connected (FC) layers before two regression and SoftMax layers follow to calculate the location of bounding boxes and classified objects. In the experiments, the probabilities of the classes will have a vehicle and background. The used loss function for learning is the same loss function used in Faster R-CNN [24].

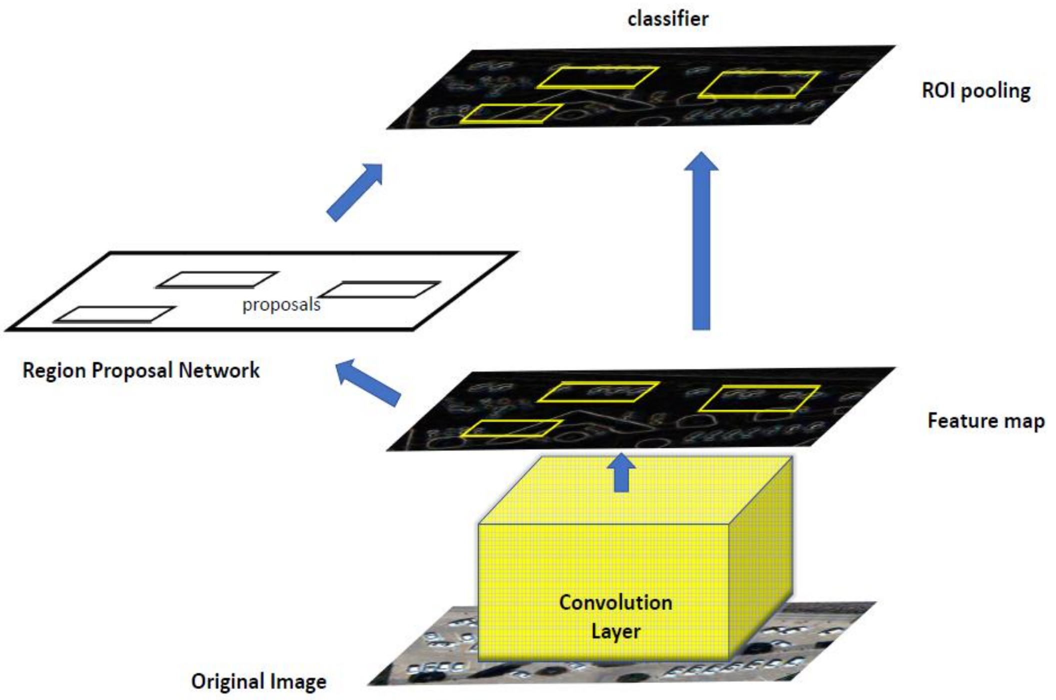

**Figure 1.** Faster R-CNN architecture

### *3.2. SSD: Single shot multi-box detector.*

SSD is used to detect objects in real-time, it reduces the processing time by eliminating the need of the region proposal network, to enhance the accuracy SSD applies some improvements like multi-scale features and default boxes, The SSD object detection consists of 2 parts:

- Extract feature map.
- Apply convolution filters to detect objects.

Ordinary SSD uses VGG16 [27] to extract feature maps. But in this paper, we replace this model with Inception-V2 [28] to extract feature maps trying to enhance accuracy.Instead of region proposal network, SSD use Multi-Box Detector, after obtaining the feature maps as a layer of size (m × n) (number of locations) with p channels, a (3×3 conv) is applied to this (m ×n ×p) layer. For each location, we have a number of bounding boxes k. These bounding boxes in different sizes and different aspect ratios. The idea behind this, to find the fittest box to each kind of objects. For each box, we will compute the number of class scores c and 4 offsets for the bounding box shape, as a result, we have ((c+4) (k ×m ×n)) outputs.

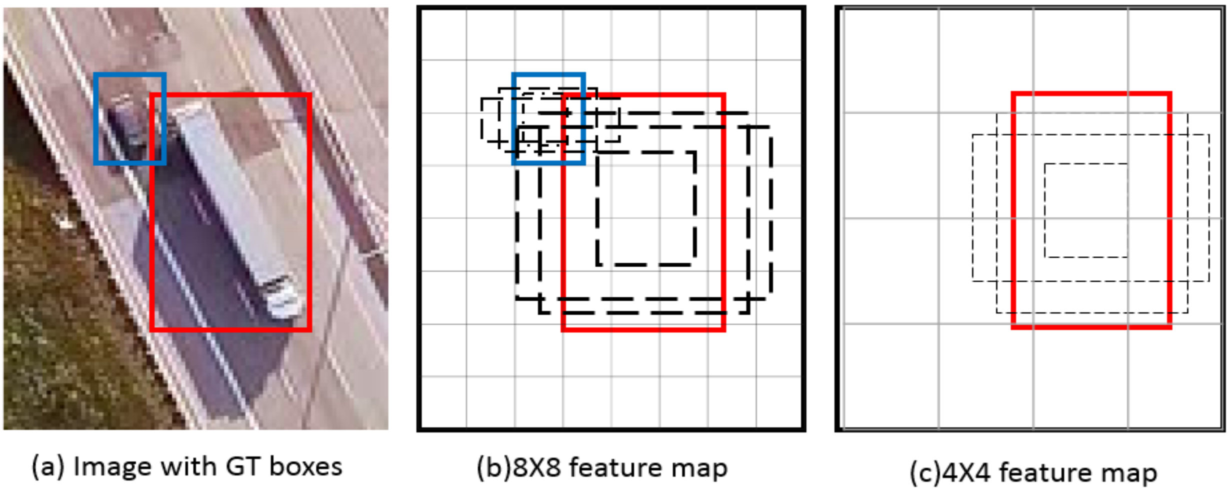

**Figure 2.** SSD architecture

## 4. Implementation

The two proposed models used for vehicles detection in this paper implementation has been achieved by the open-source framework Tensor-flow. From TensorFlow Object Detection API [29]. The chosen models are Faster R-CNN and SSD with the same base network Inception-V2.

### *4.1. Faster R-CNN inception-V2*

The training phase of Faster R-CNN the used Optimizer was Momentum Optimizer with momentum set of 0.9. A dynamic learning rate is set to start with 0.0001 till the training reaches 50K step the next 50K step learning rate is 0.00001 and the final 20K step learning rate is 0.000001 the total number of steps is 120K step. For data augmentations, applying horizontal flip and random rotation90 on the training dataset. And setting the batch size to one

### *4.2. SSD inception-V2*

In the other training phase of SSD the used Optimizer was RMS_PROP_ Optimizer with momentum and decay: of 0.9. Learning rate starts with 0.004 with decay factor 0.95 the total number of steps is 180K step. Using the horizontal flip and random rotation90 augmentation operations. And setting the batch size to one. NVIDIA GeForce GTX 1060 with 8GB of memory was used in the experiments

### *4.3. The used datasets*

The training and testing Datasets used in this paper are collected from satellite images taken from different locations around the world from Google Earth and another satellite samples such as JF-2 and WORLD-VIEW satellites, To perform the experimental part We trained the two models understudy to the training set which contains 324 images and 6843 instances of labeled vehicles. The test one contains 73 images and 1604 instances of vehicles. We tried to collect vehicle images from different environments and different spatial resolution satellites to increase the validity and reality of our attempts, all the datasets images spatial resolution is less than 1 meter.

## 5. Results

The mean average precision (MAP) used for evaluating the performance of the two proposed detection models, and comparing between them. MAP computes the average precisions value over a certain interval from Recall value= 0 to 1 Precision is the fraction of relevant instances among the retrieved instances. The recall is described as the fraction of retrieved related instances over the total value of related instances. In the Executed comparison the time consumed in testing one image with Faster R-CNN Inception-V2 was 2358.15ms and for the same one with SSD Inception-V2 was 1650.67ms. AP for testing images with Faster R-CNN Inception-V2 was 89.21% and with SSD Inception-V2 was 84.21%. From the previous results, it is noticed that Faster R-CNN Inception-V2 gives better accuracy than SSD Inception-V2. but the SSD Inception-V2 performs in a shorter time.

.

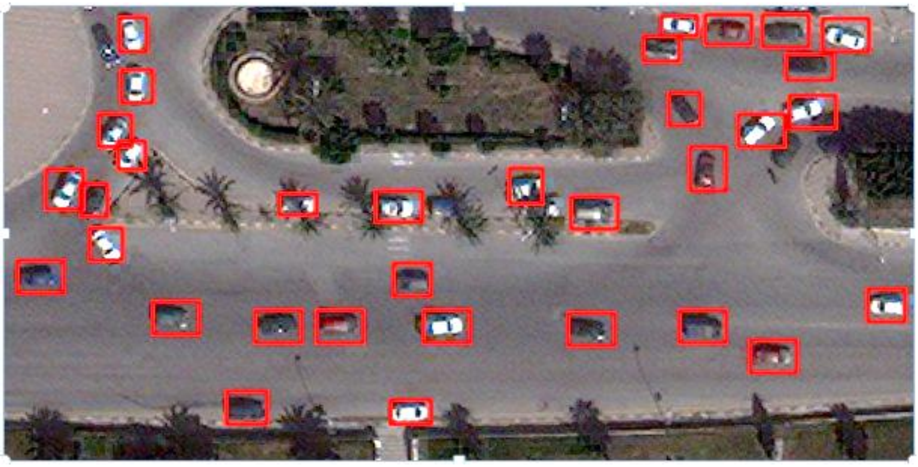

**Figure 3.** Vehicle detection by Faster R-CNN

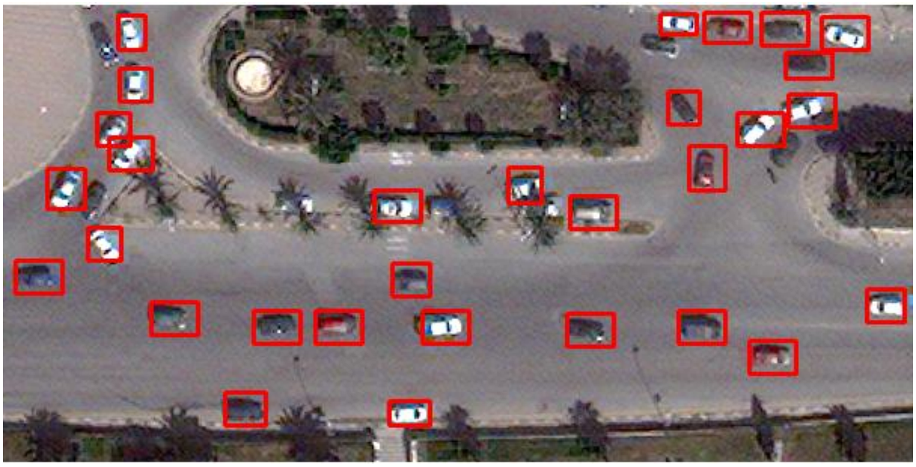

**Figure 4.** Vehicle detection by SSD

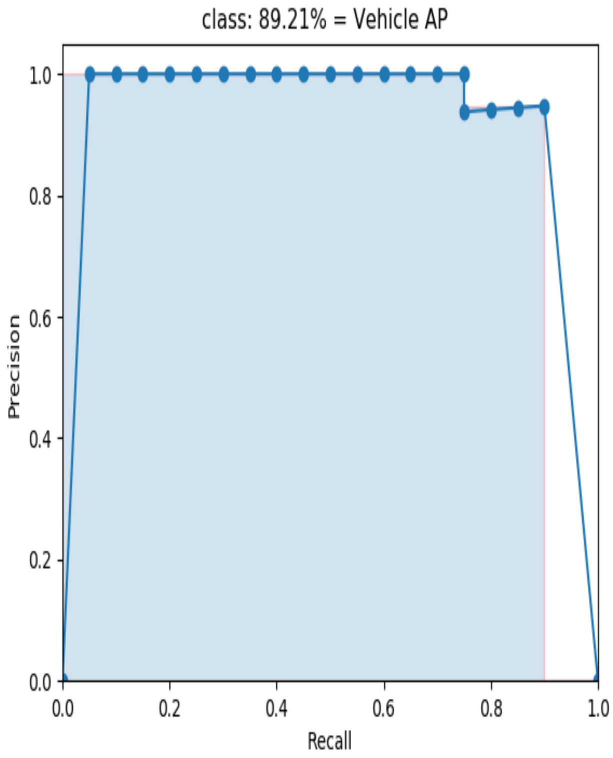

**Figure 5.** PR curve for Faster R-CNN model

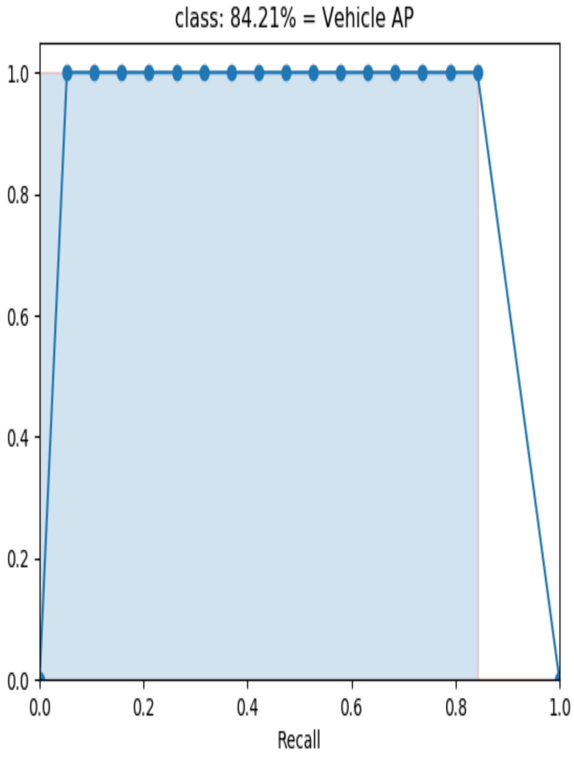

**Figure 6.** PR curve for SSD model

## 6. Conclusion

In this paper, the two states of the art algorithms for object detection (Faster RCNN and SSD) applied to detect vehicles in satellite images through Transfer Learning and making an experimental analysis comparison between them. We construct vehicle dataset collected by Google Earth and other satellite samples such as JF-2 and WORLD-VIEW satellites. The Inception-V2 used as a base network to enhance the accuracy of detection. Enlarge and increase the variety of training data by using Augmentation techniques. Mean average precision (MAP) used for performance evaluation. Based on the results obtained, Faster R-CNN Inception-V2 gives better accuracy than SSD Inception-V2. but the SSD Inception-V2 performs in a shorter time for image detection. The study will extend for general vehicle detection (bicycle, motorcycle, bus, truck).

## References

[1] on Benjdira, Bilel, et al. "Car Detection using Unmanned Aerial Vehicles: Comparison between Faster R-CNN and YOLOv3." arXiv preprint arXiv:1812.10968 (2018).

[2] Duarte, D., et al. "SATELLITE IMAGE CLASSIFICATION OF BUILDING DAMAGES USING AIRBORNE AND SATELLITE IMAGE SAMPLES IN A DEEP LEARNING APPROACH." ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences 4.2 (2018).

[3] (Dial, Gene, et al. "IKONOS satellite, imagery, and products." Remote sensing of Environment 88.1-2 (2003): 23-36.

[4] Asra, Ghassem. Theory and applications of optical remote sensing. Ed. Ghassem Asrar. New York: Wiley, 1989.

[5] Landgrebe, David A. Signal theory methods in multispectral remote sensing. Vol. 29. John Wiley & Sons, 2005.
[6] Zheng, Hong, Li Pan, and Li Li. "A morphological neural network approach for vehicle detection from high-resolution satellite imagery." International Conference on Neural Information Processing. Springer, Berlin, Heidelberg, 2006.
[7] Chen, Xueyun, et al. "Vehicle detection in satellite images by parallel deep convolutional neural networks." 2013 2nd IAPR Asian Conference on Pattern Recognition. IEEE, 2013.
[8] Hinz, Stefan. "Detection and counting of cars in aerial images." Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429). Vol. 3. IEEE, 2003.
[9] Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." international Conference on computer vision & Pattern Recognition (CVPR'05). Vol. 1. IEEE Computer Society, 2005.
[10] (Ojala, Timo, Matti Pietikäinen, and Topi Mäenpää. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns." IEEE Transactions on Pattern Analysis & Machine Intelligence 7 (2002): 971-987.
[11] Lowe, David G. "Distinctive image features from scale-invariant key points." International journal of computer vision 60.2 (2004): 91-110.
[12] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.
[13] Chen, Xueyun, et al. "Vehicle detection in satellite images by hybrid deep convolutional neural networks." IEEE Geoscience and remote sensing letters 11.10 (2014): 1797-1801.
[14] Cheng, Gong, Junwei Han, and Xiaoqiang Lu. "Remote sensing image scene classification: Benchmark and state of the art." Proceedings of the IEEE 105.10 (2017): 1865-1883.
[15] Herault, Jeanny, and Christian Jutten. "Space or time adaptive signal processing by neural network models." AIP conference proceedings. Vol. 151. No. 1. AIP, 1986.
[16] Dong, Chao, Chen Change Loy, and Xiaoou Tang. "Accelerating the super-resolution convolutional neural network." European conference on computer vision. Springer, Cham, 2016.
[17] Kim, Yoon. "Convolutional neural networks for sentence classification." arXiv preprint arXiv:1408.5882 (2014).
[18] Chen, Xueyun, et al. "Vehicle detection in satellite images by parallel deep convolutional neural networks." 2013 2nd IAPR Asian Conference on Pattern Recognition. IEEE, 2013.
[19] networks (Jin, Xiaoying, and Curt H. Davis. "Vehicle detection from high-resolution satellite imagery using morphological shared-weight neural networks." Image and Vision Computing 25.9 (2007): 1422-1431.
[20] Hu, Jingao, et al. "Fast Vehicle Detection in Satellite Images Using Fully Convolutional Network." Chinese Conference on Intelligent Visual Surveillance. Springer, Singapore, 2016.
[21] Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.
[22] He, Kaiming, et al. "Spatial pyramid pooling in deep convolutional networks for visual recognition." IEEE transactions on pattern analysis and machine intelligence 37.9 (2015): 1904-1916.
[23] Girshick, Ross. "Fast r-cnn." Proceedings of the IEEE international conference on computer vision. 2015.
[24] Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." Advances in neural information processing systems. 2015.
[25] Liu, Wei, et al. "Ssd: Single shot multibox detector." European conference on computer vision. Springer, Cham, 2016.
[26] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of

the IEEE conference on computer vision and pattern recognition. 2016.

[27] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[28] Szegedy, Christian, et al. "Rethinking the inception architecture for computer vision." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[29] Huang, Jonathan, et al. "Speed/accuracy trade-offs for modern convolutional object detectors." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017