The comparison of Faster R-CNN and Atrous Faster R-CNN in different distance and light condition

K Srijakkot, I Kanjanasurat, N Wiriyakrieng and C Benjangkaprasert

Department of Computer Engineering, Faculty of Engineering,
King Mongkut’s Institute of Technology Ladkrabang, Bangkok, Thailand.

corresponding author’s e-mail: chawalit.be@kmitl.ac.th

Abstract. This paper presents the comparison of Faster R-CNN and Atrous Faster R-CNN, which detection model, in the different distance and light condition. Also, the dataset for model training is COCO, and the classification model is residual network. The parameter for decision the performance of the model is Mean Average Precision (mAP). The results from an object resolution at 1024x768 of Faster R-CNN at 3 meters in the evening achieved mAP 1.000. Besides, the mAP at 5 meters and 8 meters were 0.798 and 0.760, respectively. The same resolution as previous, the results of Atrous Faster R-CNN at 3 meters in the evening presented mAP 1.000. Also, the mAP at 5 meters and 8 meters were 1.000 and 0.960, respectively. In addition, Atrous Faster R-CNN had better accuracy than Faster R-CNN with appropriate range and brightness from the period of the day for real-life usage.

1. Introduction

The technology of surveillance camera was widely used to detect abnormal situations from indoors and outdoors [1]. Also, this technology could record movement, detected any object, and identified them. The properties of a surveillance camera can apply the detection and identify the ability for further usages. In the past, the computer could understand which the detected images were detected by using the Convolutional Neural Network (CNN) [2] to transform the image into the feature maps. Object detection using a method called R-CNN. Even though R-CNN [3] could detect and use Support Vector Machine (SVM) in classification types of objects, the efficiency of detection was not enough for the real situation. R-CNN would do a selective search before feeding them into the CNN that used long time-consuming and got about two thousand feature maps. Thus, the new process changes could skip selective search processes to feed a single input image to the CNN, which the result from the CNN would be only one feature map from one input image. This method was called Fast R-CNN [4] that would reduce processing time. It was choosing a suitable object detection model for an application that used out-door surveillance cameras to detect invaders. Distance and light-condition were needed to consider. R. Gavrilescu et al. [5] who used Faster R-CNN model to detect traffic indicators and evaluated the model with different light-condition in each period of the day to test the performance of the Faster R-CNN in the application. This study was training and testing at a speed of 15 fps on a set of a dataset containing 3,000 images. The results indicated that Faster R-CNN could be able to use in real-life situations but could not show the objections distance in the test. B. Liu et al. [6] presented object detection based on Faster R-CNN and used a different dataset to represent detection and classification efficiency of Faster R-CNN. The study could classify objects such as human, cat, car. However, it only concerned about detection efficiency, which the condition of detection objects

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

distance and brightness were not included in the results. Although the model of Faster R-CNN [7] was efficient for object detection, it could not detect a small object well because location or spatial information were lost in the deeper layers. Atrous Convolution Neural Networks [8] allow to enlarge larger field-of-view can keep without increasing the number of parameters or the amount of computation, and this solution might be useful for recognizing objects from a different distance in the experiment. However, Atrous Convolution could detect small side objects, but it needs to exchange with increasing processing time.

In this paper, light-condition of environment where a camera was installed would be used to evaluate the faster R-CNN model, Atrous faster R-CNN model and also test in different distance-condition. The only factor was not enough to summarize the effectiveness of whether it is suitable for use in real-life. We should add more factor for reliability.

2. Methodology

2.1 Faster Region Convolution Neural Network

The bottleneck in R-CNN [3] and Fast R-CNN [4] process was selective search, which slow down and time-consuming in the process. After images were already calculated with CNN instead of running a separate discriminating exploration algorithm on the feature map to classify the region proposals liked Fast R-CNN. Faster R-CNN reused the same CNN results to predict region proposals and then the region proposals were reshaped using Region of Interest (RoI) pooling layer which was used to classify the images within the proposed region and prognosticate the offset values for the bounding boxes. The process is shown in figure 1.

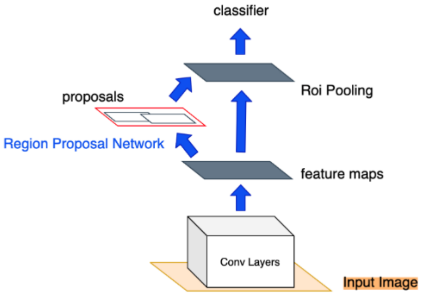


Figure 1. Block diagram of Faster R-CNN**.**

2.2 Atrous Faster Region Convolution Neural Network

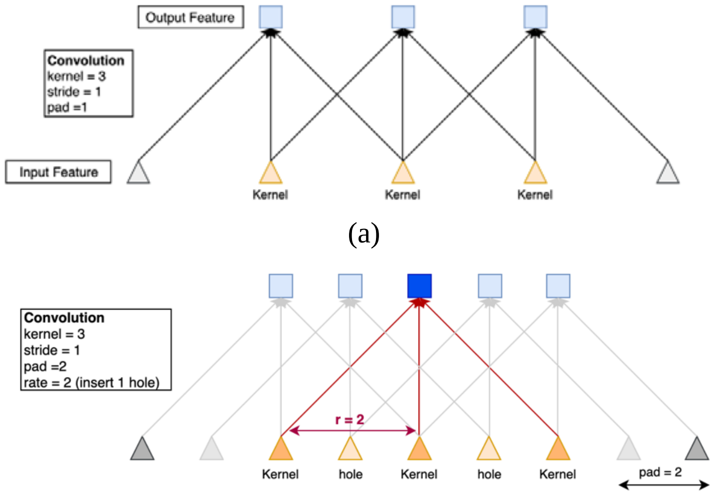
Atrous Convolution [9] was a technique which commonly used in wavelet transform and was recently applied in convolutions for deep learning model in a semantic segmentation task. The same operation was later called dilated convolution. The Atrous convolution allowed us to enlarge the fields of filters to incorporate broader context so an efficient mechanism would control the filed-of-view and find the best trade-off between accurate localization and context assimilation without increasing resource of computation. In figure 2, we can see that when rate r = 2, the input signal was sampled every 2 inputs for convolution. Thus, at the output, the 5 outputs make the output feature map larger.

2.3 Mean Average Precision (mAP)

Mean Average Precision or mAP [10,11] was the method of checking the efficiency of the model. The mAP would find from all average precision and bring them to be average. Then, we would get the value that was mAP. The equation of mean average precision is defined as an equation (1)

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{1}$$

where AP is average precision and N is the number of classes.



(a)

(b)

Figure 2. Atrous convolution. (a) A low resolution input feature map was extracted feature with standard convolution. (b) A high resolution was extracted feature with Atrous convolution with rate r = 2.

The AP [11] can be calculated from the shape of the precision curve, and is defined as an equation (2)

$$AP = \frac{1}{M_p}\sum_{i=1}^{M_p} Pr(i) \qquad (2)$$

where Pr is the precision at each recall level r which taking the maximum precision measured and M_p is the number of image samples.

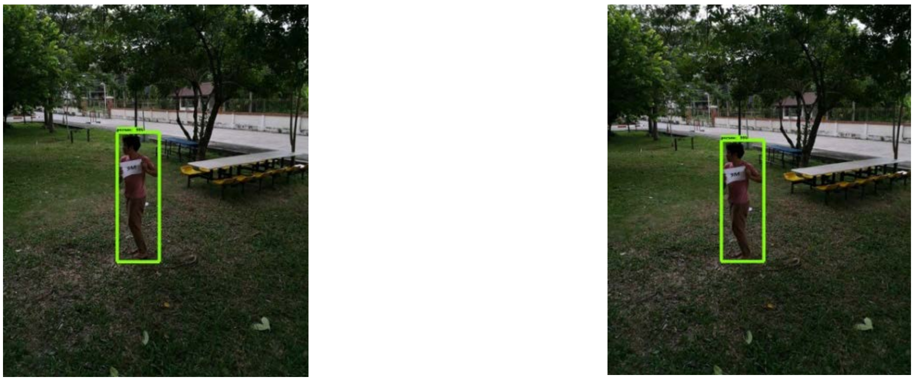
3. Implementation and Results

In this paper, the dataset for model training, which the COCO model, was run on Faster R-CNN and Atrous Faster R-CNN. The classification model was the residual network (ResNet). We then collected the test set data, composed of 30 human images in each range and brightness (light condition). In addition, we also collected data with a different resolution of images. The resolutions chosen test in the experiment have size 640x480, 800x600, and 1024x768. The test sets were collected for testing efficiency of the model to compare between suitable for use in real-life.

The results of human detection at 3 meters with every image resolution in figure 3 were shown in table 1. The values of mAP of human detection in the morning, afternoon and evening are shown in table 1 is 1.000 in every period of the day and every image resolution. The mAP of Atrous Faster R-CNN was equal to Faster R-CNN in this distance.

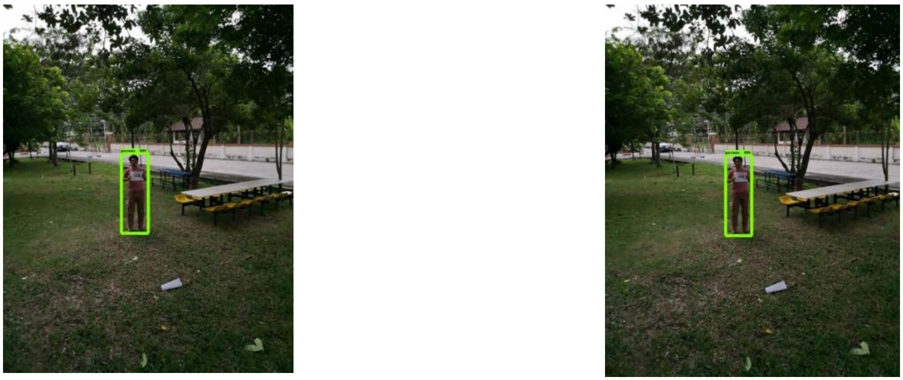
Table 1. Comparison results of an object at a distance of 3 meters with Faster R-CNN ResNet50 and Faster R-CNN ResNet50 + Atrous.

Person range 3 meters	Faster R-CNN			Faster R-CNN + Atrous		
	640x480	800x600	1024x768	640x480	800x600	1024x768
mAP in the morning	1.000	1.000	1.000	1.000	1.000	1.000
mAP in the afternoon	1.000	1.000	1.000	1.000	1.000	1.000
mAP in the evening	1.000	1.000	1.000	1.000	1.000	1.000



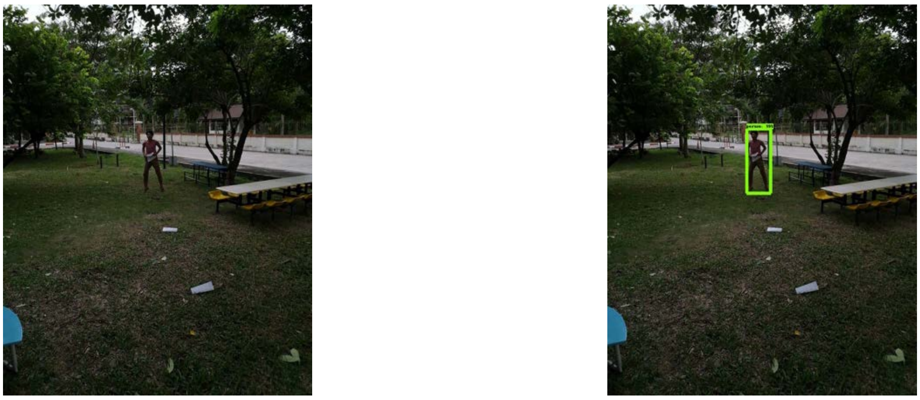
(a) 3 meters with Faster R-CNN (b) 3 meters with Faster R-CNN + Atrous

Figure 3. Example of object detection for the human at 3 meters in the evening.



(a) 5 meters with Faster R-CNN (b) 5 meters with Faster R-CNN + Atrous

Figure 4. Example of object detection for the human at 5 meters in the evening.



(a) 8 meters with Faster R-CNN (b) 8 meters with Faster R-CNN + Atrous

Figure 5. Example of object detection for the human at 8 meters in the evening.

The results of human detection at 5 meters in figure 4 with every image resolution were shown in table 2. The mAP of Atrous Faster R-CNN was better than Faster R-CNN 0.110 in the morning, 0.001 in the afternoon and 0.202 in the evening with image resolution at 1024x768.

The results of human detection at 8 meters in figure 5 with every image resolution were shown in table 3. The mAP of Atrous Faster R-CNN was better than Faster R-CNN 0.078 in the morning, 0.122 in the afternoon and 0.200 in the evening with image resolution at 1024x768.

Table 2. Comparison results of an object at a distance of 5 meters with Faster R-CNN ResNet**50** and Faster R-CNN ResNet**50** + Atrous.

Person Range 5 meters	Faster R-CNN			Faster R-CNN + Atrous		
	640x480	800x600	1024x768	640x480	800x600	1024x768
mAP in the morning	0.842	0.878	0.840	1.000	1.000	1.000
mAP in the afternoon	0.960	0.960	0.980	0.996	0.979	0.981
mAP in the evening	0.800	0.800	0.798	0.998	1.000	1.000

Table 3. Comparison results of an object at a distance of 8 meters with Faster R-CNN ResNet50 and Faster R-CNN ResNet50 + Atrous.

Person Range 8 meters	Faster R-CNN			Faster R-CNN + Atrous		
	640x480	800x600	1024x768	640x480	800x600	1024x768
mAP in the morning	0.758	0.758	0.913	0.796	0.834	0.991
mAP in the afternoon	0.858	0.880	0.878	0.996	1.000	1.000
mAP in the evening	0.720	0.720	0.760	0.920	0.960	0.960

4. Discussion

As the results shown that the distance factor was more significant to detect the model than the brightness of the image at every image resolution. At 3 meters of distance, the efficiency of Faster R-CNN and Atrous Faster R-CNN model was the same in every period of the day. However, the Atrous Convolution would start to be effective at 5 m. At 5 meters and 8 meters of distance, the efficiency of Atrous Faster R-CNN was still better than Faster R-CNN model. However, the further distance causes less of the mAP. For the brightness factor, mAP would decrease as brightness decrease, and the result appeared in the morning and evening. It can not summarize which period of the day has more than brightness from another one. Because the result in table 2 at 5 meters of a distance and resolution at 1024x768, the mAP in the afternoon has mAP more than in the morning but the result in table 3 at 8 meters of a distance, the mAP in the morning has mAP more than in the afternoon. Besides the mAP of Atrous Faster R-CNN model was more than mAP of Faster R-CNN model for every brightness range.

A various image resolution in our experiment shown in table 1 which test in a distance at 3 meters, the mAP of every image resolution has the same value, but in table 2 and table 3 which an object was placed further, the image resolution will affect to the mAP in both models.

5. Conclusion

In this paper, we used two models for object detection, and it was tested by using human images. Then, the comparison between the two models by using two factors, including distance and light condition. The models were trained from COCO dataset, and object types were classified by ResNet. The results of testing between Faster R-CNN model and Atrous Faster R-CNN model used mAP value to compare the efficiency of models. The results of mAP with image quality at 1024x768 of Faster R-CNN and Atrous Faster R-CNN at 3 meters in the evening were 1.000. As the same image quality, the results of Faster R-CNN at 5 meters and 8 meters in the evening achieved mAP 0.798, 0.760 and less than Atrous Faster R-CNN 0.184 and 0.200, respectively. The last experiment tested model predictions with different image resolution. Each image resolution at 3 meters has the same value of mAP. However, the value of the mAP with a poor image resolution at 5 and 8 meters had been decreased. In the future, we will use Atrous Faster R-CNN to detect the intruders who are smaller than human.

References

[1] Akiyama T, Kobayashi Y, Kishigami J and Muto K 2018 CNN-Based Boat Detection Model for Alert System Using Surveillance Video Camera *Proc. IEEE Global Conference on Consumer Electronics* pp 669-670

[2] Kannojia S P, and Jaiswal G 2018 Ensemble of Hybrid CNN-ELM Model for Image Classification *Proc. Int. Conf. on Signal Processing and Integrated Networks* pp 538-541

[3] Zhang W, Wang S, Thachan S, Chen J, and Qian Y 2018 Deconv R-CNN for Small Object Detection on Remote Sensing Images *Proc. IEEE International Geoscience and Remote Sensing Symposium* pp 2483-2486

[4] Hsu S C, Wang Y W, and Huang C L 2018 Human Object Identification for Human-Robot Interaction by using Fast R-CNN *Proc. IEEE Int. Conf. on Robotic Computing* pp 201-204

[5] Gavrilescu R, Zet C, Fosalau C, Skoczylas M and Cotovanu D 2018 Faster R-CNN: an Approach to Real-Time Object Detection *Proc. Int. Conf. and Exposition on Electrical and Power Engineering* pp 165-168

[6] Liu B, Zhao W and Sun Q 2017 Study of Object Detection Based On Faster R-CNN *Proc. Chinese Automation Congress* pp 6233-6236

[7] Wang G, and Ma X 2018 Traffic police gesture recognition using RGB-D and Faster R-CNN *Proc. Int. Conf. on Intelligent Informatics and Biomedical Sciences* pp 78-81

[8] Guan T, and Zhu H 2017 Atrous Faster R-CNN for Small Scale Object Detection *Proc. Int. Conf. on Multimedia and Image Processing* pp 16-21

[9] Chen L C, Papandreou G, Kokkinos I, Murphy K, and Yuille A L 2018 DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs *IEEE Transactions on pattern analysis and machine intelligence* pp 834-848

[10] Li K, Huang Z, Cheng Y C, and Lee C H 2014 A Maximal Figure-of-Merit Learning Approach to Maximizing Mean Average Precision with Deep Neural Network based Classifiers *IEEE International Conference on Acoustic, Speech and Signal Processing* pp 4503-4507

[11] Kim I and Lee C H 2011 Optimization of Average Precision with Maximal Figure-of-Merit Learning *Proc. IEEE International Workshop on Machine Learning for Signal Processing* pp 1-6