Available online www.ejaet.com

European Journal of Advances in Engineering and Technology, 2020, 7(2):30-33

Research Article ISSN: 2394 - 658X

AI Referee with Mask R-CNN

Kyu Jung Choi and Chae-Bong Sohn*

Dept.of Electronics and Communications Engineering, Kwangwoon University, Seoul, Korea - 01897
cbsohn@kw.ac.kr

ABSTRACT

In this paper, Object detection is a fundamental field of computer vision and has received much attention in recent years and has made great development. As the development progressed, there were many cases applied to various fields like Sports, surveillance, autonomous driving. This paper describes the algorithm of object detection and describes the papers to which it is applied. In particular, the three-second rule of basketball will be heard as an example. If the attacker or defender without a ball is in the paint zone for 3 seconds, the offense is 3 seconds.

Key words: Object detection, Mask R-CNN, Sports, AI Referee, 3 second rule

---

1. INTRODUCTION

As the object detection algorithm has actively been studied, continuous developments such as Fast R-CNN [2] are being made starting from the existing R-CNN [1]. Object detection has purpose to detect location of the object in the image and classify as a class, so this is mainly applied in the field of computer vision, such as image retrieval, security and surveillance, autonomous driving. In this paper, we apply an object detector to detect the position of players in basketball.

In this paper, we aimed to recognize players who have entered the specific zone (paint zone) regardless of the ball and judged as a foul when the time is over 3 seconds. However, there are some problems. As a result of seeing the basketball game on the wide screen, the images of the players are very small. This can cause difficulty in recognizing one player or occluding the foot when the players are overlapped, making it difficult to determine the position. We must figure out exactly how to make a decision. Therefore, we decide to use Mask R-CNN [3] which has higher accuracy than other object detection.

In the next section, from R-CNN to Mask R-CNN will be described. In section 3, we will present experimental setup environment and the conclusion of that experiment are given in section 4.

2. RELATED WORKS

Faster R-CNN

R-CNN searches and classifies candidates of objects extracted through Selective Search through the network. However, the initial R-CNN model has a large amount of computation because it uses all the results extracted through selective Search to CNN for detection [5]. Classify over Support Vector Machine can also cause R-CNN to slow down.

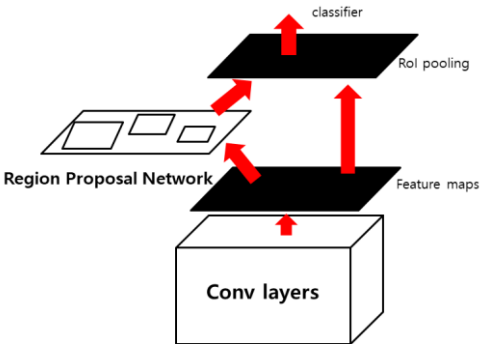

Fig. 1 Architecture of Faster R-CNN

Faster R-CNN saves time by using region proposal process instead of Selective Search algorithm to solve this speed problem. This builds a model with a structure for learning Region Proposal Network(RPN) and instead of performing convolution operations for every Region Proposal, ROI Pooling applies a CNN only once to an input image and extracts features to identify objects by pooling. Figure 1 shows the structure of Faster R-CNN.

Mask R-CNN

In this paper, Mask R-CNN is used to for detection and segmentation. Mask R-CNN adds a binary mask that determines whether a pixel is an object within a section in the existing Faster R-CNN and uses ROI Pooling. In the existing Faster R-CNN, because the model was only for object detection, the loss of location information was not a problem during ROI Pooling.

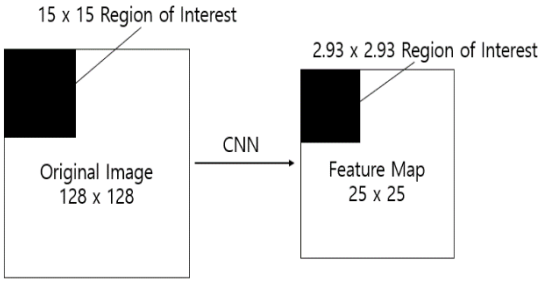

Fig. 2 Losing object position information in ROI Pooling

In Figure 1, we first output the feature map (25x25) by entering the original image (128x128) into the CNN. This indicates that the size of the input image changes at a rate of 0.1953125/ Here, the ROI (Bounding Box) of 15x15 size is also changed in the same proportion and becomes approximately 2.93x2.93. In the ROI Pooling of existing Faster R-CNN, it rounds up to use a 3x3 ROI. At this time, the difference generated affects the final performance. In Mask R-CNN, the 2D Linear Interpolation method reduces the decimal point error that occurs in the ROIPooling area while passing through the CNN to extract the exact pixel position. This is called ROI Align. Table 1 compares the characteristics and performance of ROI Align used in ROI Pooling, ROI warp and Mask R-CNN in Faster R-CNN.

Table -1 Performance comparison table (ROI Pooling, ROI Warp, ROI Align)

	align	**blinear**	**agg**	AP	AP_{50}	AP_{70}
ROI Align			max	26.9	48.8	26.4
ROI Pooling		✓	max	27.2	49.2	27.1
		✓	ave	27.1	48.9	27.1
ROI Warp	✓	✓	max	30.2	51.0	31.8
	✓	✓	ave	30.3	51.2	31.5

We want to apply Mask R-CNN like this to basketball. Although few cases of object detection have been studied at the level of simple players. In addition to simply detection players, we have proposed a kind of artificial intelligence referee that can recognize 3 seconds rule, which is one of the rules of basketball, by recognizing only players within a certain area such as paint zone. Controversies have continued in the existing sports world, such as bias determination and nausea, and there are cases where artificial intelligence referees have already been introduced in sports such as baseball. Since basketball has been the case, the artificial intelligence referee in this paper can be very helpful. The learning data was divided into 2 classes, paint zone and player, and the learning data was created for more precise recognition by learning in the form of polygons rather than simple rectangular boxes. As a result, we could not detect perfectly segmented detection due to the lack of training data, but the players in the paint zone succeeded in recognizing.

3. EXPERIMENT

The data used in this paper was edited about 30 frames of NBA game on November 7, 2016. In the process of creating a Ground Truth, it was classified into 2 classes, paint zone and player, and the Ground Truth was created so that the recognized player could be a player who has a bright paint zone, At this time, since the camera that shoots the actual game is several tens, only the players who are clearly in the paint zone are recognized based on the feet of the players from the angle of view shown in the video. For example, in the data below, one player is clearly in the paint zone but is

not recognized because the player's feet are not visible and can be recognized by simply stepping on the paint zone line. However, the recognition of the players and paint zones is not nearly as perfect as the existing Mask R-CNN.

Fig. 3 Detection of Players in Paint Zone

4. CONCLUSION

In this task, we used Mask R-CNN to test artificial intelligence in basketball game. As you can see in the picture, the mAP detected the player in 97.8% of the paint area, so the mask did not split perfectly. Finally, since it was recognized when in a certain position, it correctly judged the foul for 3 seconds.

The problem with this paper was that the mask results were incomplete, resulting in very unstable results when the players overlap. Solving this problem and based on paper, not only the 3 second rule but also the judge other rules can recognize the positions of the players and use them as a strategy analyser.

Acknowledgements

The present research has been conducted by the research grant of Kwangwoon University in 2019.

REFERENCES

[1]. 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. p. 580-587.

[2]. Girshick, Ross. "Fast r-cnn." *Proceedings of the IEEE international conference on computer vision*. 2015. p. 1440-1448.

[3]. He, Kaiming, et al. "Mask r-cnn." *Proceedings of the IEEE international conference on computer vision*. 2017. p. 2961-2969.

[4]. Burić, Matija, MiranPobar, and Marina Ivašić-Kos. "Object detection in sports videos." *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. IEEE, 2018. p. 1034-1039.

[5]. Uijlings, Jasper RR, et al. "Selective search for object recognition." *International journal of computer vision* 104.2 (2013): 154-171.

[6]. 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.

[7]. Liu, Wei, et al. "Ssd: Single shot multibox detector." *European conference on computer vision*. Springer, Cham, 2016. P. 21-37.

[8]. 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. P. 779-788.

[9]. Redmon, Joseph, and Ali Farhadi. "Redmon, Joseph, and Ali Farhadi. "YOLO9000: better, faster, stronger." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017. P.7263-7271.

[10]. Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." *arXiv preprint arXiv:1502.03167* (2015).

[11]. Ivasic-Kos, Marina, and MiranPobar. "Multi-label Classification of Movie Posters into Genres with Rakel Ensemble Method." *International Conference on Innovative Techniques and Applications of Artificial Intelligence*. Springer, Cham, 2017. P. 370-383.
[12]. Haralick, Robert M., and Linda G. Shapiro. "Image segmentation techniques." *Computer vision, graphics, and image processing* 29.1 (1985): 100-132.
[13]. He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. P. 770-778.
[14]. Szegedy, Christian, et al. "Going deeper with convolutions." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015. P. 1-9.
[15]. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*. 2012. P. 1097-1105.