❐   1330

Badminton player detection using faster region convolutional neural network

Nur Azmina Rahmad[1], Nur Anis Jasmin Sufri[2], Nurul Hamizah Muzamil[3],
Muhammad Amir As'ari[4]

[1,2,3,4]School of Biomedical Engineering and Health Sciences, Faculty of Engineering,
Universiti Teknologi Malaysia, Malaysia
[4]Sport Innovation and Technology Center (SITC), Institute of Human Centered Engineering (IHCE),
Universiti Teknologi Malaysia, Malaysia

Article Info

ABSTRACT

Nowadays, coaches and sport analyst are concerning about sport performance analysis through sport video match. However, they still used conventional method which is through manual observation of the full video that is very troublesome because they might miss some meaningful information presence in the video. Several previous studies have discussed about tracking ball movements, identification of player based on jersey color and number as well as player movement detection in various type of sport such as soccer and volleyball but not in badminton. Therefore, this study focused on developing an automated system using Faster Region Convolutional Neural Network (Faster R-CNN) to track the position of the badminton player from the sport broadcast video. In preparing the dataset for training and testing, several broadcast videos were converted into image frames before labelling the region which indicate the players. After that, several different trained Faster R-CNN detectors were produced from the dataset before tested with different set of videos to evaluate the detector performance. In evaluating the performance of each detector model, the average precision was obtained from precision recall graph. As a result, this study revealed that the detector successfully detects the player when the detector is being fed with more generalized dataset.

Corresponding Author:

Nur Azmina Rahmad
School of Biomedical Engineering and Health Sciences, Faculty of Engineering,
Universiti Teknologi Malaysia,
Johor Bahru, Malaysia.
Email: amir-asari@biomedical.utm.my

1. INTRODUCTION

Video is one of the unstructured data [1] that contains meaningful understanding that only can be obtain through analyzing. Video sport analysis get more attention nowadays because by analyzing the previous sport game, it inspires the coaches to rearrange the players' position or enhance the performance of the players. The task of analyzing sport can be generalized into movement classification [2-3], player detection [4] and activity classification [5]. Several previous works proposed the identification and tracking of players in sport video [6] while other discovered the content characterization in sport program by applying several techniques on extracted low-level features such as audio, video and text captions [7] as method in sport video analysis. Most of the researches in sport video analysis focus on three issues which are shot classification, highlight extraction and ball and player tracking [8-9]. Many analysts urge to use video as medium in their observation and summarize the tactical analysis of the athlete. However, the conventional method which is manual video notation is very time consuming [10-11].

In analyzing a video, a large capacity is required to compute a large dataset. Deep learning method has been said to be the most promising method compared to the other conventional machine learning method [12]. It is a subfield of supervised learning [13] and a well-known method compared to the last technique which is shallow learning. With the capability to have large network as it consists of three components which are input, output and two or more of hidden layers, it becomes preferable method used in many research [10, 14]. Then deep learning goes through evolution process and produce a model known as Convolutional Neural Network (CNN). It was derived from Artificial Neural Network (ANN), but differ from ANN as the present of convolution operation in convolutional layer [15]. R-CNN born after CNN brings new technique in image processing which it introduced region of

interest (ROI) instead of full image [16]. This model has successfully fulfilled the demands of less training time [17]. Consequently, Salvador et al. take initiative to combine these two models and produce Fast R-CNN [16]. This approach minimizes the training time because the network identifies ROI after convolution. The latest evolution was Faster R-CNN where some modification has been done from the previous model where Faster R-CNN has an ability to extracts both image and region feature efficiently [16]. Nowadays, CNN has been used widely in analyzing task since the computational cost can be reduced as well as it gives result near real-time [18-19].

As mentioned before, Faster R-CNN model is one of the deep learning method which success in detecting object that invariant to rotation, pose, scale and illumination condition. There are several work that have been done in analyzing the sport video content such as soccer [6], basketball and volleyball [8]. However, to our knowledge, the detection of players based on faster R-CNN in broadcast video of badminton match has not been studied yet. Therefore, the badminton player detection from broadcast video using faster R-CNN was introduced in this study. The proposed algorithm will identify the badminton player in badminton match which could provide information about the movement of player and facilitate the coaches in monitoring the current performance and directly boost the capabilities of player.

2. RESEARCH MATERIAL AND METHOD
2.1. Research Material
Focusing on the player detection that is one of the important component in interpreting the sport video, in this study, the broadcast video of badminton match was obtained from the YouTube database and the software called Virtual dub was used to extract the sequence of image frames from the obtained video. MATLAB 2017a is used in developing the faster R-CNN deep learning algorithm for both training and testing.

2.2. Research Flowchart
Figure 1 shows the flow of the research for a better understanding. Each sub process is presented in Section 2.21 to Section 2.2.
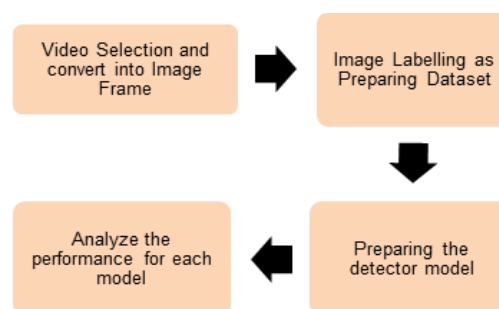

Figure 1. Flowchart for developing Faster R-CNN

2.2.1 Video Selection
There are many broadcast sport videos available online. Medium such as YouTube, blogs and official sport webpage enriched with broadcast video from previous game to the latest one. In this study, the application of Faster R-CNN on the badminton broadcast video was executed using three videos of badminton match: 1) Men Single Final Badminton Asia Championship 2017 Figure 2; 2) All England Badminton Tournament Men's Single Final 2011 Figure 3; and 3) Men's Badminton Doubles Gold Medal Olympic 2012 Figure 4. Each video has resolution of 360 resolutions.

Figure 2. Example of image frame from video 1

Figure 3. Example of image frame from video 2

Figure 4. Example of image frame from video 3

2.2.2　Image Labelling as Dataset Preparation

Analyzing the continuous video is a tough task as it required specific software that compatible to treat video as input. As an alternative, the extraction of image frames from video has been done by using software called Virtual dub. This software provides variety of choices for user to choose either to extract the image of full video or make some selection of interested scene in the video.

Only 100 image frames from each video were selected for labelling session. By using Training Image Labeler in MATLAB Application, the badminton players are being labelled with a square box regardless the referee and spectators as shown in Figure 5-7. The significance of this step is to build a dataset of video frames which will be used for the training and testing. In evaluating the effectiveness of Faster R-CNN with different condition, the data of each video was arranged according to several cases which later will be described in the next section.

Figure 5. Labelled image from broadcast video 1

2.2.3　Preparing The Detector Model

After the dataset is prepared, it is used to feed the proposed R-CNN during the training process in producing the trained R-CNN detector model. In training session, six different type of trained models were

created and then the trained detector is being tested with the testing dataset according to six different cases as shown in the following Table 1.

Figure 6. Labelled image from broadcast video 2

Figure 7. Labelled image from broadcast video 3

Table 1. List of Model and Description

Case	Type of trained R-CNN detector model	Testing Dataset
1	Model 1: R-CNN was trained with video 1	Video 1 (a single match)
2	Model 2: R-CNN was trained with video 2	Video 2 (a single match)
3	Model 3: R-CNN was trained with combination of video 1 and 2	Combination of video 1 and 2 (both single matches)
4	Model 4: R-CNN was trained with video 3	Video 3 (Double match)
5	Model 5: R-CNN was trained with combination of video 1 and 3	Combination Single Match and Double Match
6	Model 6: R-CNN was trained with combination of video 1, 2 and 3	Combination both Single Matches and Double Match

2.2.4 Analyze the Performance for Each Model

After that, the aforementioned trained models were tested with several testing videos combination according to Table 1 in evaluating the capability of the detector in tracking the position of the player. After the testing process is done, the image frame with square box indicate the score how much confident the detector able to detect the player can be produced. In analyzing the performance of the detector in each case, the precision recall graphs were generated before the average precision were calculated.

3.   RESULTS AND DISCUSSION

Precision recall (PR) graph was evaluated to demonstrate the performance of each detector. The average precision obtained from PR graph is analyzed to determine which model has the best performance where it able to spot the badminton player excellently in every testing videos. Figure 8 to Figure 10 indicate the score pointing to the badminton players. In general, Faster R-CNN successively able to detect the player when it was tested with the same video data that was used to train the detector. Figure 11 demonstrate the performance of detector for all six cases stated in Table 1. Apparently, detector capable to detect player better when it has been trained and tested on the same video compared to when it trained and test on different video. Furthermore, detector can trace player perfectly in most of the video when it was trained with more generalized video data. From the graph in Figure 11, it is obviously seems that detector in Case 1, Case 2 and Case 4 produce high precision when training and testing with same video dataset compared to testing on different video. For instance, the detector model in Case 1 manage to get high average precision when it is tested with video 1 which is the same video used to train the detector while produce lower average precision when it is tested with other video combination. This is similar to other cases such as Case 2 and Case 4.

Besides, the performance of Faster R-CNN improving when it trained with combination between different set of videos. As demonstrate in Figure 11, the performance of trained detector for Case 3, Case 5 and Case 6 have been improved (when it is tested with almost on all video combination) compare to the previous Case 1, Case 2 and Case 4. Faster R-CNN has best performance in detecting player when it has been trained with combination of all videos as proved in Case 6. This is because Faster R-CNN has variety of data with different condition and more generalized information. From Figure 11, it can be seen that the trained detector capable to detect the player perfectly almost in all condition since it has learned variety of features. Hence, detector require many data including varies color, texture and condition so that it can be excellent in detecting player from different source of the broadcast video.

Figure 8. Detected player from video 1

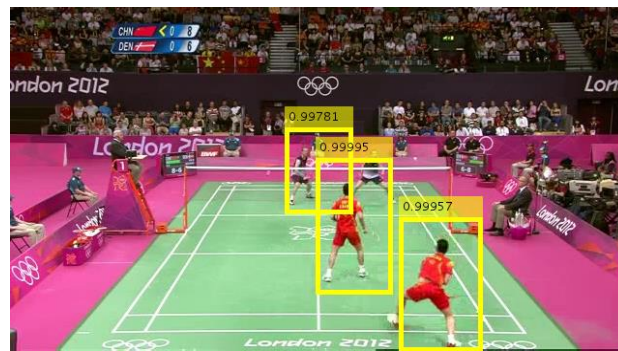Figure 9. Detected player from video 2
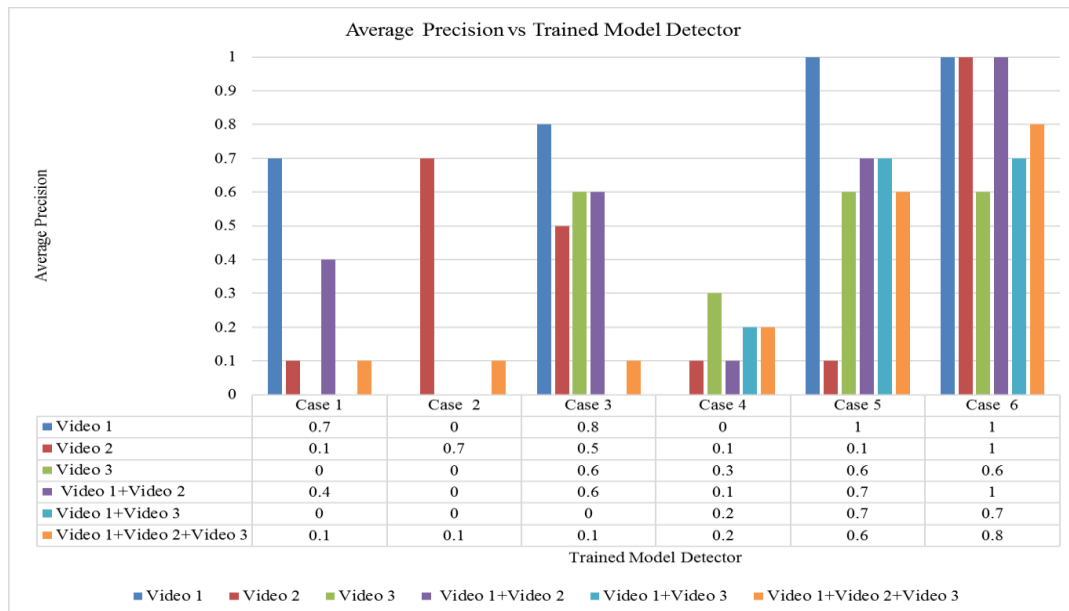

Figure 10. Detected player from video 3

Figure 11. Average Precision vs Trained Model Detector

4. CONCLUSION AND FURTHER WORK

As a summary, this study has been presented a new technique for automatic player detection from broadcast video via Faster R-CNN. By applying variation of badminton video including combination between single match and double match to the detector, it would be beneficial to coaches and sport analyst in

improving their player's performance. Overall, most of the model may have excellent performance based on analyzing average precision from precision recall graph. Though, there also has several models might not be able to perform well in detecting the player due to some factors that has been discussed in Chapter 3. As stated above, there is no study regarding the automatic player detection for broadcast badminton video using deep learning. Thus, this approach has potential to detect players in other sport video or non-sport video as well. In future, it is advised that some improvement should be done for this system. The detector should be validated by training and testing with more diverse condition such as low quality and high definition video to ensure the system is totally intelligent enough to distinguish player from background. Furthermore, it is worthwhile when the computational time can be reduced since the proposed system require quite long training duration. Lastly, it is an opportunity for researchers to explore the appropriate technique to detect player from live video rather than using image frames.

REFERENCES

[1] Srinivasa KG, Anupindi S, Sharath R, Chaitanya SK. *Analysis of facial expressiveness captured in reaction to videos.* 7th IEEE Int. Adv. Comput. Conf. IACC 2017. 2017; 664–670.
[2] Sudhir G, Leet M, Jain AK. Automatic classification of tennis video for high-level content-based retrieval. *IEEE. 1997;*81-90.Li S, Lihong H. Resreach of background segmentation method in sports video. *TELKOMNIKA Indonesian Journal of Electrical Engineering.* 2014;12(6): 4274-4282.
[3] Turchini F, Seidenari L, Bimbo AD. Understanding and localizing activities from correspondences of clustered trajectories. *Computer Visual Image Understanding.* 2017;159(1): 128–142.
[4] Fani M, Yazdi M, Clausi DA, Wong A. Soccer video structure analysis by parallel feature fusion network and hidden-to-observable transferring markov model, *IEEE Access.* 2017; 3536(1):1-11.
[5] Lu WL, Ting JA, Murphy KP, Little JJ. *Identifying players in broadcast sports videos using conditional random fields.* Proc. IEEE CVPR 2011. CO. 2011; 2:3249–3256.
[6] Adami N, Leonardi R, Migliorati P. Overview of multimodal techniques for the characterization of sport programs. *Visual. Communnication. Image Process.* 2003; 5150(1): 1296–1306.
[7] Chen HT, Chen BS, Lee SY. Physics-based ball tracking in volleyball videos with its applications to set type recognition and action detection. *IEEE International Conference. Acoustic Speech Signal Process.* HI. 2007;1: 1097–1100.
[8] Tao ZH. A tracking algorithm of moving target in sports video. *TELKOMNIKA Indonesian Journal of Electrical Engineering.* 2014; 12(10): 7463-7470.
[9] Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Fei-Fei L. *Large-scale video classification with convolutional neural networks. CVPR 2014 IEEE Conference.* OH. 2014; CVPR.2014.223: 1725–1732.
[10] Girshick R. Fast r-cnn. *ICCV.* 2015; abs/1504.08083(1): 1440–1448.
[11] Rahmad NA, As'Ari MA, Ghazali NF, Shahar N, Sufri NAJ. A survey of video based action recognition in sports. *Indonesian Journal of Electrical Engineering and Computer Science*, 2018; 11(3): 987-993.
[12] Basheer IA, Hajmeer M. Artificial neural networks : fundamentals , computing , design , and application. *Journal of Microbiological methods.* 2000; 43(1): 3–31.
[13] Xu Z, Yang Y, Hauptmann Itee AG. A discriminative CNN video representation for event detection. *CVPR.* 2014; abs/1411.4006(1): 1-11.
[14] Ji S, Xu W, Yang M, Yu K. 3D Convolutional neural networks for human action recognition.*IEEE. Pattern Analysis and mAchine Intelligence.* 2013; 35(1): 221–231.
[15] Salvador A, Giro-I-Nieto X, Marques F, Satoh S, Faster R-cnn features for instance search. *CoRR.* 2016; abs/1604.08893(1): 394–401.
[16] Le THN, Zheng Y, Zhu C, Luu K, Savvides M. *Multiple scale Faster-RCNN approach to driver's cell-phone usage and hands on steering wheel detection.* 2016 IEEE CVPRW. NV; 2016: 46-53.
[17] Azam S, Rafique A, Jeon M. *Vehicle pose detection using region based convolutional neural network.* ICCAIS. Ansan; 2016:194–198,
[18] Saha S, Singh G, Sapienza M, Torr PHS, Cuzzolin F. Deep learning for detecting multiple space-time action tubes in videos. *CoRR.* 2016; abs/1608.01529(1): 1–13.
[19] Chen CM, Chen LH. *Novel framework for sports video analysis : a basketball case study.* 2014 IEEE ICIP. Paris; 2014: 961–965.