Applying Convolutional Neural Network for Detecting Highlight Football Events

Tuan Hoang Viet Le[3], Hoang Thien Van[3], Hai Son Tran[4], Phat Kieu Nguyen[5], Thuy Thanh Nguyen[6], and Thai Hoang Le[1,2(✉)]

[1] Faculty of Information Technology, University of Science, Ho Chi Minh City, Vietnam
lhthai@fit.hcmus.edu.vn
[2] Vietnam National University, Ho Chi Minh City, Vietnam
[3] Faculty of Information Technology, Saigon International University, Ho Chi Minh City, Vietnam
{lehoangviettuan,vanthienhoang}@siu.edu.vn
[4] Faculty of Information Technology, University of Education, Ho Chi Minh City, Vietnam
haits@hcmup.edu.vn
[5] Department of Academic Affairs, Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam
nkphat@ntt.edu.vn
[6] VNU University of Technology, Ha Noi, Vietnam
nguyenthanhthuy@vnu.edu.vn

Abstract. Automatic detection of videos with outstanding situations is a practical issue that needs to be studied in many events of different fields with common length and frequency of occurrence for instance: meetings, musicals, sports events that the user uploads regularly, one of the concerned areas is the highlights in football videos. The matches of the annual top leagues and between nations within federations form a huge database in need of different purposes in which requires the specific model assisting in extracting outstanding situations. Besides, building a reliable and accurate model requires an appropriate approach, a large amount of training data (diverse, accurate, clear data), which need to be assigned label correspondingly. The Convolutional Neural Network (CNN) was chosen as an approach to help building a smart system as a foundation, combined with a proposed new method for synthesizing results called the adaptive threshold towards specific data, along with the optimization model to draw a reliable conclusion. The work was proceeded on a video data set of the top four teams of the English Premier League (ELP) 2018–2019 and a randomly selected dataset on the Internet.

Keywords: Key frame · Highlight football events · Convolutional Neural Network (CNN) · Highlight football events classification using CNN · Wrongly-Validated Dataset Re-training (WVDR)

P. Cong Vinh and A. Rakib (Eds.): ICCASA 2021, LNICST 409, pp. 300–313, 2021.
https://doi.org/10.1007/978-3-030-93179-7_23

1 Introduction

Currently, football leagues in general and professional tournaments, in particular, draw great attention from fans, experts in many different fields even from world top companies that recognize football as a promising business field. Their great need is to gather information, collect statistics or simply review the highlights of the football matches. Hundreds of corner kicks, thousands of goals, millions of fouls take place worldwide every weekend when the tournaments occur, hence data explosion is considered a challenge, and working on its solution has highly practical meaning. From this real challenge, the problem of detecting outstanding situations in football videos is proposed as follows: the input as a full video of any match, the output as scenes containing prominent elements. (Corner, Fault, Goal [1]). The overall architecture of an outstanding situation detection system includes: video pre-processing [2–4], extracting and labeling features [5–7], building machine learning models from extracted features [8–10], using it to spot outstanding situations in the input video [11–13].

Therefore, within the scope of this paper, we propose a model based on a convolutional neural network developed from work [14] that was built from the group of authors, in conjunction with adaptive threshold and wrongly-validated dataset retraining and a method for synthesizing the results. The experimental results 95,8% in our EPL dataset showed the feasibility of a proposed model.

2 Background and Related Work

2.1 Background

Machine Learning research has been under serious concern of which Deep Learning is an expansion. To get high degree of accuracy of result in recognition of the vision, speech, and audio; processing and translation of natural language, or filtering of the social network, deep learning architectures have been broadly used.

Yann LeCun introduced an architecture network in 1988 named Convolutional neural networks (CNN), which now has been used in various ranges of activities in problem recognition like translation, medical diagnostics, image classification, and recognition, etc. To get worthwhile information, the inputs are filtered by the layers, which are composed of the CNN. Though no feature selection is needed, useful information extraction relies on automatically filters adjusting. In terms of dealing with problems in image classification, CNN is a better choice in comparison to Normal Neural networks.

In this paper, the authors built an application for training and validating the CNN model by using ConvNetJS. So far, the ConvNetJS has been used within one own browser as an open JavaScript library for training Deep Learning models.

Figure 1. An overview of Convolutional Neural Network architecture for image classification.

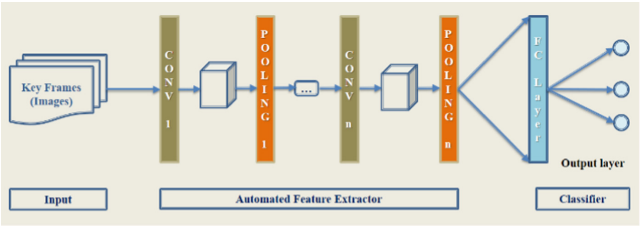

Fig. 1. CNN image classifier model

2.2 Related Work

Identifying highlights in video clips, especially football, is deeply and widely noticed. A demonstration [15], inspired by the model of two-stream CNN, DilateRNN, and long short-term memory (LSTM) units, the authors conducted classification of a data set of SoccerNet. The authors also further investigated the inclusive performance of their best model and achieved significantly improved performance 0.8%–13.6% compared to state of the art, and up to 30.1% accuracy gain in comparison to the baselines. On the other hand [16], the research of ball localization is successfully carried out using several sub-modules to tackle the limitation of using one fixed model. Hence, the authors present an optimized ensemble algorithm for effective and efficient ball tracking. The main module in their proposed design is to localize the ball when it is freely moving on the football pitch with higher accuracy. In [17], the authors also refer to a novel soccer video event detection algorithm based on self-attention. It extracts keyframes through the self-attention mechanism and then obtains the characteristics of time window level through the NetVLAD network. Finally, each video clip is classified into 4 types of events (goals, red/yellow card, substitutions, and others). The experimental results show that with the introduction of the self-attention mechanism, the classification accuracy on the SoccerNet data set has improved from 67.2% to 74.3%.

3 Proposal Methods for Detecting Highlight Football Events

3.1 CNN Model Architecture

As mentioned above, this article proposes a CNN model based on works from [14] to classifying the event of the input video after extracting keyframes, combined with a practical-based evaluation method called “Adaptive threshold” and “Wrongly-validated dataset Re-training” for classifying events.

The model was run on two devices with specification: Local computer (EN1): Intel Core i7 - 4400, 8 GB RAM, Windows 10 pro 64bit; and the second Weak AWS Cloud Computer (EN2): Intel Xeon E5–2686, 8 GB RAM, Windows Server2016 64bit, both surfing Google Chrome browser.

Figure 2. Architectural representation of the model that the authors use to solve the mentioned problem. Each keyframe following through the CNN model leads to a

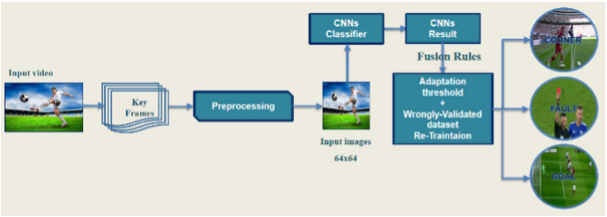

Fig. 2. CNN model architecture

conclusion of belonging or not to the set of highlights. We then apply the proposed evaluation methods and will further present them in Sect. 3.2.

3.2 CNN Model

CNN model, whose architecture is shown as Fig. 3, used to train on 64 × 64 English Premier League 2018 2019 image dataset (Dataset EPL20182019). This dataset will be mentioned in Sect. 4.3.

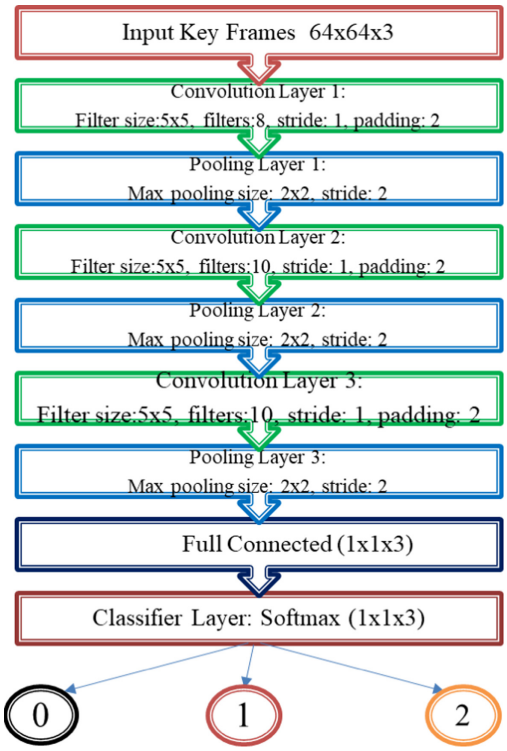

Fig. 3. CNN architecture

Figure 3. An overview architecture of the CNN Model.
With

0 is CORNER
1 is FAULT
2 is GOAL.

The proposed model has the pre-proceeded image set as input data, groups it according to the defined prominent events, then labels them. The proposed model has 3 layers of Conv (The first, second and third layer includes 5, 10, 10 filters each correspondingly), 3 Pooling layers, 1 layer Full connected, and 1 layer Softmax with 3 output results.

3.3 Adaptive Threshold

According to practical experiments, the authors propose a threshold where KF is determined through the training and testing process by the CNN model presented above, it is possible to classify the input data. The term adaptive threshold is developed on the formula:

$$1 - Max(V_{R_C}, V_{R_F}, V_{R_G}) < 0,5 \tag{1}$$

Call $V_{R_x} = \{V_{R_C}; V_{R_F}; V_{R_G}\}$
If (1) is TRUE then result R_x.
If (1) is FAULT then result R_0.
With V_{R_x} is the result of KF after running the CNN model.

R_G means the result of KF is Goal.
R_F means the result of KF is Fault.
R_C means the result of KF is Corner.
R_O means the result of KF is different.

The adaptive threshold, hereby, is determined by:

$$V_{R_{\bar{x}}} = \frac{\sum r_{x1}, r_{x2} \ldots r_{xn}}{n} \tag{2}$$

With R_{xi} mean KFs meet Eq. 1.
$V_{R_{\bar{x}}}$ mean adaptive threshold based on the fact.
Using the above formula, the authors expect to shorten the identification time as well as optimize the outcome through the actual features of the situations occurring in football.

3.4 Wrongly-Validated Dataset Re-training

A wide variety of colors, objects, angles of rotation… in the above features, governs the authors' offer to the *Wrongly-validated dataset Re-training* to improve the structure's accuracy.

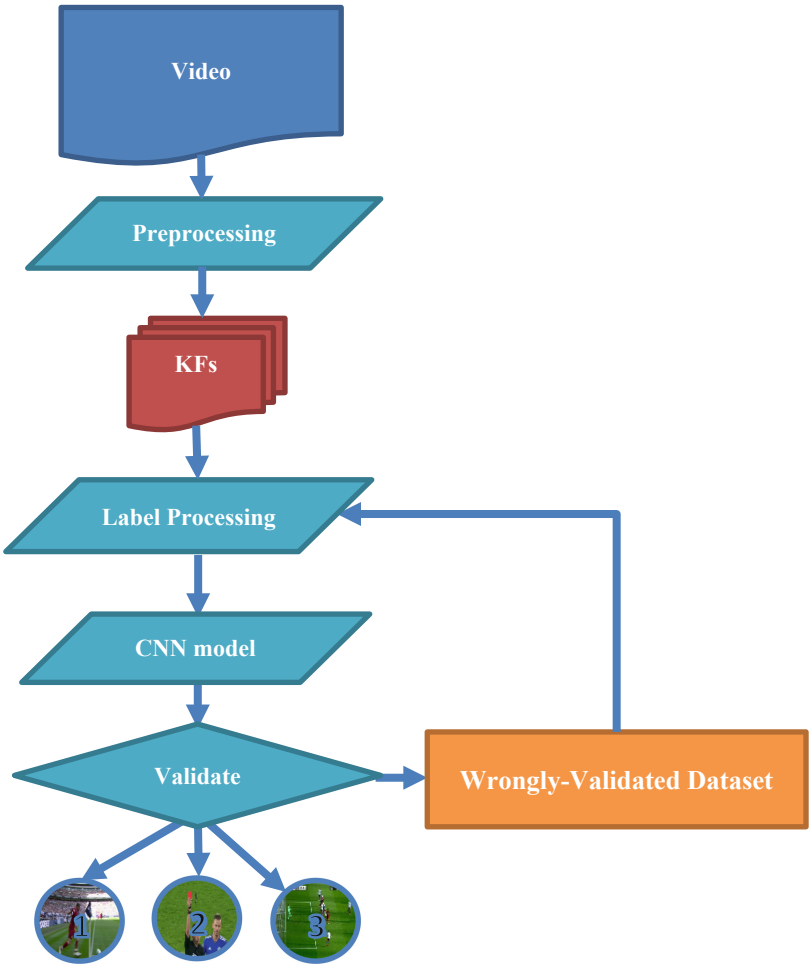

Fig. 4. The flow of wrongly-validated dataset re-training method

With

0 is CORNER
1 is FAULT
2 is GOAL.

Figure 4. The approach and process of the proposed method and are described below:

- **Step 1:** Input videos are pre-proceeded through the conversion system into image frames.
- **Step 2:** Proceed to label the frames by the specialist on the football field.
- **Step 3:** Put the standardized and labeled data into the proposed CNN model to conduct training, then classify the data for the validating set.

- **Step 4:** Proceed to record information of separate validated samples giving false classification results from the validating set (built independently and completely new presented in Sect. 4.3).
- **Step 5:** Synthesize the wrong samples into a set.
- **Step 6:** Apply a label pre-assigned by the specialist into the proposed model for retraining. The results obtained are positive, improved, and presented in Sect. 4.4.

4 Experimental Results

4.1 Dataset

Our primary dataset for the project is Top 4 English Premier League 2018–2019 (ELP20182019) and randomly collected data on the Internet. The dataset contains labeled 600 KFs, which were divided into a training set of 400 images, testing set of 200 images. Each image was marked 0 for Corner, 1 for Fault, and 2 for Goal by a specialist with years of experience in the field of referee in football match. We built 02 datasets from the collected.

Fig. 5. Dataset EPL20182019

Figure 5. The dataset description containing 600 ELP20182019 KFs with dimension 64 × 64.

Fig. 6. Dataset Random_Internet

Figure 6. The dataset 2 description of 200 Random_Collected KFs with dimension 64 × 64.

4.2 Environment for Experiments

Pre-proceeding Image
We scaled each labeled images in Dataset into 64 × 64 size. And then we divided them into two parts of Training and Testing as mentioned previously (Fig. 6). As a result, we have constructed two datasets.

Training CNN Model
With the above CNN architecture, we trained the model with the following parameters:

- Learning Rate: 0.01, Momentum: 0.9,
- Weight decay: 0.0001, Batch size: 1,
- Method for training is Adadelta.

CNN Model Accuracy Evaluation
From our perspective, without Wrongly-Validated dataset Re-training, the assessment of the accuracy of the classification model is essential. To evaluate the proposed model, we have selected any 200 KFs images (consist of 66, 66, 68 labeled Corner, Fault, and Goal images respectively) for proposed model testing, which results in 95.2%. By using Wrongly-Validated dataset Re-training, its accuracy evaluation increases to 95.8% noticeably, which will be presented in detail in the next section.

4.3 CNN Components Accuracy Evaluation

Assessment of the accuracy of the classification is vital because it allows predicting the accuracy of the classification results of future data and participating in the comparison of different classifiers. The CNN classifiers proposed used the Hold-out method for evaluating.

Fig. 7. The structure and flow of operations of the dataset

Figure 7. An overview of the Hold-out method which we had used to evaluate the accuracy of the CNN model.

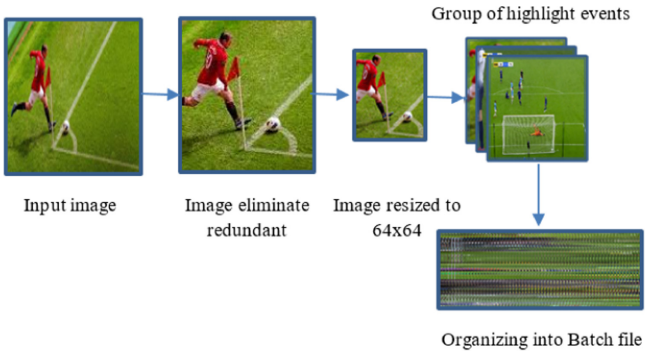

Fig. 8. Keyframe images process flow

Figure 8. The way that the data is proceed in which images are extracted from the video, then stripped of redundant details (do not contain features that influence model decision-making). After being labeled, images will be resized to 64 × 64, and proceed to convert to binary files from the image files into the data set and these binary files form batches of data. This is appropriate input for the proposed CNN model to achieve sufficient performance.

Training Process

Experimenting with the proposed model, the authors used ConvNetJS which is a Javascript library to train deep learning models entirely in the browser, software requirement, compiler, installation, or GPU. This library was designed and built by Karpathy and is used in many research projects because of its high applicability [18] [19]. This is the library realizing the ConvNet model proposed by Yann Lecun.

Fig. 9. Experimental progress of the CNN model

Figure 9. During the training and evaluating of the group model, the authors applied the K-Fold evaluation method to conduct experiments.

Evaluation of Validating Process

Table 1. The result of the proposed model in the different environment

Result	Environment	
	EN1	EN2
Training time (minutes)	50	104
Max accuracy on the validate set (%)	95.2	93.6
Classification loss (%)	0.0007	0.0006
L2 Weight loss (%)	0.1346	0.1377

Table 1 shows the result of the proposed in two different environments presented above. It may be observed that the EN2 needs more training time to achieve high results. Besides, the proposed model has proven to work on low-end computers.

In addition, we also prepared a new dataset named EPL20202021 to experiment with the method Wrongly-Validated dataset Re-training of the data organized as follows.

Table 2. Structure of EPL20202021 dataset

Dataset	Corner	Fault	Goal
EPL20202021	500	500	1000

Table 2 shows a structure with a total of 2000 validate samples, we obtained 100 samples that the system misrepresented, then used to retrain the model, and obtained more positive results presented in Sect. 4.4.

4.4 Experimental Results

The Result of CNN Model Testing with Different Environment

The results obtained when experimenting with 2 datasets EPL20182019 on 2 different environments introducced in Sect. 4.2.

Table 3. Result of CNN model different environments

Dataset	EN1	EN2
EPL20182019	95,2%	93,6%
Ramdom_Data	93,4%	91,8%

Table 3 shows that, noticed by the authors, the accuracy differences between the two environments are not significant, demonstrating the compactness, lightness, and feasibility of the proposed model. In addition, the model also achieved a quite high

accuracy with the data set randomly selected on the Internet with the feature of images from the camera, which shows good adaptability to different types of data of the model.

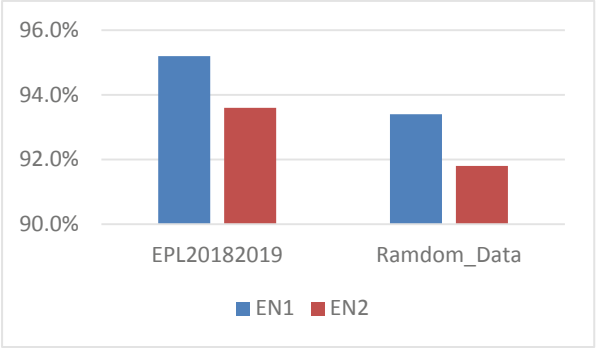

Fig. 10. Result of testing on different environments

Figure 10. The chart depicts the classification results of the proposed model on two different data sets

Table 4. Result of CNN model on different Dataset

Dataset	TOP4_ELP20182019		RANDOM_COLLECT	
Result	Amount	(%)	Amount	(%)
True	190	95	93	93
False	10	5	7	7

Table 4 shows the classification results of the proposed model, the rate of correct recognition, and the number of samples that the system detected wrong. Through a quick review, the authors found that the errors come from the complexity of the GOAL, CORNER situations with a large number of objects and the turbulence of the angle and color in the test cases.

Table 5. Result of CNN model with Adaptive threshold

Fusion rule	CNN Model results			
	Dataset1		Dataset2	
Result	Time(s)	Acc(%)	Time(s)	Acc(%)
Before	55	95.2	63	93.4
After	50	95.2	60	93.4

Table 5. As a result of applying the Adaptive threshold method, the authors found that the training time of the model decreased slightly but still achieved the initial accuracy of the proposed model, is a sign that the effectiveness of the proposed method, as well as suggestions for directions to improve the model in the future.

Table 6. Result of CNN model with Wrongly-Validated dataset Re-training method

WVDR	CNN Model results	
Result	Dataset1 Acc(%)	Dataset2 Acc(%)
Before	95.2	93.4
After	95.8	93.6

Table 6. Presenting the improvement of the model's accuracy through retraining the samples that the system recognizes as wrong, through statistics from the validation process, and then aggregated into a new training set. Use the assigned label for this set to train the proposed model to gain more knowledge about difficult false positives.

4.5 Discussion

With the aim of constructing a model with the ability to recognize outstanding situations in football as well as in other sports, then extract outstanding points to serve the needs of a substantial number of fans, experts and also case analysts. The authors also encountered difficulties such as camera angle and distance between stadiums; The dissimilar quality of pictures obtained from broadcasters; Instability in important features such as the color of the player's outfit, accessories; also The stadium design (goal, flagpole, net color…).

Moreover, the challenge of the precision of the standout situation is also noticeable since validity must be taken into account, and the appearance of VAR (Video Assitant Referee) is also noise the statement by the negation of the validity of the situation.

5 Conclusions

In this research, we proposed the CNN model and methods for synthesizing the results of the components of the model which we are called "Adaptive threshold" and "Wrongly-Validated dataset Re-training".

- Propose a suitable CNN network architecture for the highlighted situations as mentioned in football matches.

- Use a combination of the evaluation "Adaptive threshold" and the optimal processing method Wrongly-Validated dataset Re-training
- Validate CNN model with combined evaluation methods to detect two datasets of highlight football events. The accuracy results of 95.2% and up to 95.8% showed the feasibility of the proposed model when it combined these rules.

Acknowledgment. This project would not be posible without the financial means from Sai Gon International University (SIU). Many thanks to my specialist Mr. Nguyen Vo Thuan Thanh, major in Physical Education, for providing expert advice and labeling the dataset. And finally, thanks to numerous friends who endured this process with me, offering me lots of support and effort.

References

1. Vietnam Football Federation: Law of football. The duc the thao Ha Noi Publisher (2013)
2. Shambharkar, P.G., Doja, M.N.: Movie trailer classification using deer hunting optimization based deep convolutional neural network in video sequences. Multimed. Tools Appl. **79**(29), 21197–21222 (2020). https://doi.org/10.1007/s11042-020-08922-6
3. Bastani, F., Madden, S.: MultiScope: efficient video pre-processing for exploratory video analytics. arXiv preprint arXiv:2103.14695 (2021)
4. Del Campo, F.A., et al.: Influence of image pre-processing to improve the accuracy in a convolutional neural network. Int. J. Combin. Optim. Probl. Inform. **11**(1), 88–96 (2020)
5. Stoeve, M., et al.: From the laboratory to the field: IMU-based shot and pass detection in football training and game scenarios using deep learning. Sensors **21**(9), 3071 (2021)
6. Jackman, S.: Football shot detection using convolutional neural networks (2019)
7. Shi, S.: Comparison of player tracking-by-detection algorithms in football videos (2020)
8. Viet, V.H., et al.: Multiple kernel learning and optical flow for action recognition in RGB-D video. In: 2015 Seventh International Conference on Knowledge and Systems Engineering (KSE). IEEE (2015)
9. Bottino, A.G., Hesamian, S.: Deep learning model for 2D tracking and 3D pose tracking of football players (2020)
10. Russo, M.A., Kurnianggoro, L., Jo, K.-H.: Classification of sports videos with combination of deep learning models and transfer learning. In: 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE). IEEE (2019)
11. Sheng, B., et al.: GreenSea: visual soccer analysis using broad learning system. IEEE Trans. Cybern. **51**, 1463–1477 (2020)
12. Tran, D.-S., et al.: Real-time hand gesture spotting and recognition using RGB-D camera and 3D convolutional neural network. Appl. Sci. **10**(2), 722 (2020)
13. Venkatesh, S., Ramachandra, R., Bours, P.: Video based deception detection using deep recurrent convolutional neural network. In: Nain, N., Vipparthi, S., Raman, B. (eds.) CVIP 2019. Communications in Computer and Information Science, vol. 1148, pp. 163–169. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-4018-9_15
14. Tran, H.S., Le, T.H., Nguyen, T.T.: The degree of skin burns images recognition using convolutional neural network. Indian J. Sci. Technol. **9**(45), 1–6 (2016)
15. Mahaseni, B., Faizal, E.R.M., Raj, R.G.: Spotting football events using two-stream convolutional neural network and dilated recurrent neural network. IEEE Access **9**, 61929–61942 (2021)

16. Perera, D.S., et al.: Ball localization and player tracking using real time object detection. In: International Conference on Advances in Computing and Technology (ICACT–2020) Proceedings. ISSN 2756-9160 (Nov 2020)
17. Ma, S., et al.: Event detection in soccer video based on self-attention. In: 2020 IEEE 6th International Conference on Computer and Communications (ICCC). IEEE (2020)
18. Vo, A.T., Tran, H.S., Le, T.H.: Advertisement image classification using convolutional neural network. In: 2017 9th International Conference on Knowledge and Systems Engineering (KSE). IEEE (2017)
19. Kieu, P.N., et al.: Applying multi-CNN model for detecting abnormal problem on chest x-ray images. In: 2018 10th International Conference on Knowledge and Systems Engineering (KSE). IEEE (2018)