Turkish Journal of Computer and Mathematics Education *Vol.13 No.03 (2022), 377-387*

Research Article

An offside soccer detection system using ontology and deep learning

Mohammed Yassine Kazi Tani [a], Lamia Fatiha Kazi Tani [b], Abdelghani Ghomari [c]

LabRI-SBA Lab., Ecole Superieure en Informatique Sidi Bel Abbes, Algeria [a]
RIIR Laboratory, Computer Science Department, University of Oran 1 Ahmed Ben Bella,Oran, Algeria [b,c]
y.kazitani@esi-sba.dz [a]
lamiakazitani@yahoo.fr [b]
ghomari65@yahoo.fr [c]

Article History: *Received: 18 April 2022; Accepted: 25 May 2022; Published online: 06 June 2022*

Abstract: Nowadays, the Soccer events detection domain has become a more critical issue that attracts many researchers due to the enormous volume of available soccer video data worldwide. Consequently, it was a complicated task to recognize events using the video object detection process. This challenge leads us to propose an approach based on deep learning supplied by the ontology paradigm. This article develops a soccer offside detection system divided into two parts: applying deep learning algorithms to extract both visual and audio low-level features like balls, players, referee whistle sound, Etc. The second one considers these results and runs some ontology SWRL rules to identify events like offside or not offside players. Our final experiments demonstrate that the proposed approach reached better results than the other ones in the state-of-the-art.

Keywords: Deep learning; Soccer offside event detection; CNN, RNN, Mask R-CNN, visual and audio features extraction, Ontology paradigm; SWRL rules

1. Introduction

In recent years, multimedia has been used in various fields, such as e-learning, telemedicine, surveillance, soccer, Etc. However, several existing multimedia documents are in different forms and extensions, like text, sound, graphics, and video. In this paper, we will focus on video documents. Generally, the primary research problem in the machine learning domain concerns the semantic gap between low-level features and their semantic interpretation. In the soccer domain, this problem can translate low-level features like colours and sounds on goal, penalty, red or yellow card, offside, Etc. Many recent works in state-of-the-art, try to handle this problem. In [1], the authors propose an approach for identifying major complex events like "Ball possession" and "Kicking the ball" in soccer videos, starting from object detection and spatial relations between objects. In the aim of exploiting the advantages of Convolution Neural Network (CNN) and the ability of Recurrent Neural Network (RNN), a deep neural network system to detect soccer video events was proposed [2]. Another approach based on DTW and Interval Type-2 Fuzzy Logic Systems employing the Big Bang Big Crunch (BB-BC) algorithm for video activity detection and classification of critical events from the large-scale data of soccer videos was proposed in [3]. Another approach has been proposed in [4] and developed a system that uses C3D (Convolution 3-dimensional) to exploit spatiotemporal relation for detecting events like goal attempt completely, penalty kick, corner, shoot. A novel soccer video event detection algorithm based on self-attention has been proposed in [5]. The main idea is to extract keyframes through the self-attention mechanism, to detect events like goals, red/yellow cards, substitutions, and others. In [6], the authors propose an efficient deep learning-based framework for soccer events recognition in IoT-enabled FinTech. The proposed framework performs event recognition in three steps: Firstly, image frames extraction and refinement. Secondly, frame-level features extraction. Finally, event recognition using (MLSTM) network.

Our main contributions of this paper are presented as follow:

First, in this paper, we propose a complete ontology that regroups all concepts needed for semantic interpretations of visual and audio low-level features extracted in the soccer domain. Secondly, our focus on the only offside event leads us to improve our previous works [7] in the soccer domain by adding audio information like our background in the surveillance domain [8,9] and getting better precision and recall results. Furthermore, we based our background research on recent works (most of the articles cited are recent). Finally, our used deep learning algorithm combined with ontology paradigm and specified SWRL rules give us a promising result compared with state-of-the-art ones.

This paper is organized as follows: Sect. 2 concerns related works. In Sect. 3, we present our soccer ontology. Section 4 focuses on a comparative study with other works in state-of-the-art. Our offside detection system architecture is illustrated in Sect. 5. Final results of the system developed here are given in Sect. 6. The final section provides some concluding remarks and future works.

2. State of the art

In the soccer domain, an ontology paradigm can be helpful as a solution to support the indexing process and events recognition. In [10], authors developed a Falcon-S solution, representing a semantic web application for semantically indexing soccer images and using ontology for driven searching and browsing mechanisms. Another approach proposed in [11] developed a system capable of standardizing the concept related to the soccer domain. For this purpose, low- and high-level concepts were created and validated by 60 soccer experts, using Protégé (as a development tool) and OWL (as language). Ontology-based information extraction and retrieval system and their application in the soccer domain have been presented in [12]. The author proposed three issues in semantic search, namely, usability, scalability, and retrieval, to develop a keyword-based semantic retrieval system.

However, first of all, the works cited above are very old, and no work uses the strength of the deep learning algorithms. Secondly, the description of the ontology concepts for the soccer domain is not complete and no part of the audio was presented. Finally, no SWRL (semantic web rule language) inference rules are used for having new knowledge descriptions.

Other approaches in the background and related works based their system on deep learning for video processing in the soccer domain. In [13], the author proposed a model that presents an essential element for pass analysis named the concept of value. Several factors have been used to evaluate the impact of passes and decision-making. For example, the expected value added to the possession (expected pass EPV added) or the probabilities of the pass reaching a given location, the expected value of the successful pass (reward), and the expected value in case of turnover (risk). Another approach proposed in [14] presents a new deep learning model for 2D ball detection and tracking (DLBT) in soccer videos. Authors describe many issues for moving objects blob detection, classification of an image patch into three classes (ball, player, and background), and ball track validation. In [15], The authors present a system capable of estimating entire probability surfaces of potential passes in soccer based on high-frequency spatiotemporal data and fully convolutional neural network architecture. Another approach for long soccer video summarizing has been presented in [16]. The authors based their system on a three-dimensional Convolutional Neural Network (3D-CNN) and Long Short-Term Memory (LSTM) – Recurrent Neural Network (RNN). In [17], the authors propose an approach that generates story clips with complete temporal context, based on a replay detection model. the system finally detects events in long soccer games with a single pass through the video. Individual passe between players' predictions [18] has been presented in the 5th Workshop on Machine Learning and Data Mining for Sports Analytics. The authors based their system on a convolutional architecture to predict the receiver of the pass between all players. Another approach in state of the art [19] proposes a system for soccer events detection based on deep learning. The authors used a model to distinguish between red and yellow cards images. For this purpose, three modules are used: the variational autoencoder (VAE) module, the image classification module, and the fine-grain image classification module. In [20], the authors proposed a model for predicting and classifying events in a soccer game. Based on deep learning architecture, the system used football and player tracking for class event detection like penalties. In [21], the authors evaluate the state of the art of predicting the outcome of the football match by reviewing their methods. After that, the authors tried to create a prediction system based on the strength of their analyses and used Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTMs). Another approach in state of the art [22] developed a system called GreenSea, based on a broad learning system (BLS) for soccer game analysis, tactics, and training. In [23], to develop a system capable of managing player tracking, shot segmentation, and soccer event detection, the authors propose a soccer dataset called Soccer Dataset for Shot, Event, and Tracking (SSET). A live text of football matches as helpful information for event segmentation has been used in [24]. The authors based their model on deep learning for soccer event detection and segmentation purpose. Another approach in state of the art [25] develops a model based on Spatio-temporal features of video frames for automatic recognition of essential events in soccer broadcast videos. the authors use the local visual content of video frames in a low-dimensional transformed sparse space. In [26], the authors propose a system that manages at the same time the two tasks of Soccer video scene and event classification and also improves the efficiency of video processing. A new Soccer Video Scene and Event Dataset (SVSED) with six categories has been proposed.

However, all these approaches extract only video features, and all the potentials generated by audio features combined with video ones, are not used.

Another group of state-of-the-art works based their approaches on audio features extraction for soccer video processing. In [27], the authors present and evaluate different approaches based on neural networks and combine visual features with audio features to detect (spot) and classify events in soccer videos like (goals, substitution, and cards). Another approach in the state of the art [28] proposed a novel Hough transform-based whistle detection algorithm to analyze the semantic content of soccer video by extracting both audio and visual features. In the goal

of producing highlights automatically, a novel system that uses both audio and visual features has been proposed in [29]. The model manages the three tasks of play-back detection, soccer event recognition, and commentator emotion classification. In [30], the authors propose an approach based on multi-modal features (audio and video) for soccer event classification purposes. The Soccernet benchmark dataset, which contains annotated events for 500 soccer game videos from the Big Five European leagues, is used for the system evaluation. Another approach in state of the art [31] proposed a system based on the normalization step that defines a gain fixed with time to scale each ugr and a mixing step is responsible for summing all the individual channels together. In [32], the authors propose a model that automatically highlights soccer matches. The system is based on both event features that better represent the game and audio features that help detect the excitement generated by the game.

However, all the articles cited above do not experiment with the offside event that represents a very important point in soccer domain. In [32], the authors mentioned an offside event, but no details about its experiments were presented.

After a deep analysis of all the problems noted above, we extended our previous work and introduced in this paper an innovative approach by creating an offside detection system based on both video and audio deep learning algorithm supplied by ontology paradigm that manages video and audio features at the same time. For this purpose, a group of video and audio SWRL rules was created to handle results obtained from the deep learning module and make a decision if the event of offside exists or not in the video clips with start and end frames.

3. Our ontology description

In order to overcome all the existing concepts in the soccer domain, the development of our semantic ontology was based on the description of all the essential things that have a relationship with the soccer domain. For this purpose, we extend our previous works presented in [7] by proposing a new and more complete architecture by adding new concepts, data-property and object-property, and also inserting audio parts.

3.1. Ontology concepts

For the representation and categorisation of the soccer domain, this part of our semantic ontology describes all the essential concepts needed in the soccer domain and presents them as generalisation/specialisation relationships. In order to reach this complete representation, our ontology concept part was divided into four categories representing Video_Objects, Video_Actions, Video_Sequences, and adding a new part as Audio_Objects as presented in Fig. 1 below.

Event or actions in the soccer domain like offside can be created by interactions between the different objects in the video sequences. According to their nature, the video object category represents all the objects viewed by the camera in soccer matches. These objects can be divided into three main categories: soccer equipment, humans, stadium objects. The soccer equipment part regroups objects like a ball, referee whistle, time panel; when the humans part contains objects like players, referee, manager. The third part concerns stadium objects like a soccer field and stadium stands. Table 1, illustrates the generalisation/specialisation relationship between all objects.

The audio object category regroups the different sounds detected in the soccer domain. This category is divided into two main parts: Equipment sound and human sound. Equipment sound regroups all those generated by equipment like referee whistle, fans objects sound. Human sounds contain all those created by humans, like manager voice, player voice. All the details of the audio_object category are presented in table 2 below.

Table.1. Video_Object_Hierarchy

L1	L2	L3	L4
Video_Object	Soccer_ Equipment	Ball / Referee_ Whistle / Referee_Assistant_Flag / Time_Panel / Etc…	
	Humans	Player	Center_Back / Goal_Keeper / Playmaker / Centre-forward / Sweeper / Winger / Etc…
		Referee	Official-referee / Assi_Referee / Etc…
		Manager	
	Stadium_Objects	Soccer_Field	Centre-circle / Centre-spot / Corner-flag / Corner-spot / Crossbar / Goal-lines / Halfway-line / Penalty-area / Penalty-spot / Corner lines / Corner point / Side-Line / Etc…
		Stadium_Stands	Stairs / Chairs / Etc…

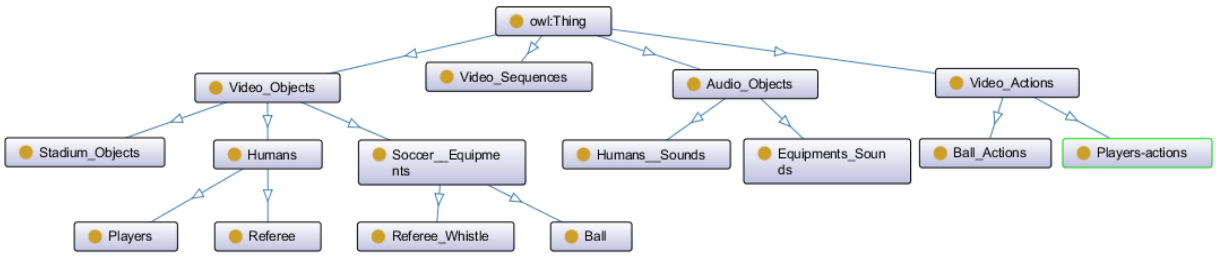

Fig. 1. Illustration of some classes of our soccer concept ontology using OntoGraf plugins.

The video action category contains all the actions that appear in the soccer domain. Two main categories can be raised here: Ball_actions and Player_actions. Ball actions concern those related to balls like ball-control or assist-pass. Players_actions regroup those in relationship with a player as, for example, a goal, hattrick, or offside. Table 3 summarizes all the actions generalisation/specialisation relationship. The video sequence category regroups all streams indexed by our system with offside events or no offside events.

Table.2. Audio_Objects_Hierarchy

L1	L2	L3
Audio_Objects	Equipments_Sounds	Referee_Whistle_Sound / Fans_Objects_Sound / Presentator's_Micro_Sounds / Etc…
	Humans_Sounds	Manager_Voice / Player_Voice / Etc…

Table.3. Video_Actions_Hierarchy

L1	L2	L3
Video-actions	Ball-actions	Assist-pass / Back-pass / Ball-control / Bicycle-kick / Corner-kick / Etc…
	Players-actions	Goal / Corner / Counter-attack / Dribbling / Equalizer / Foul / Goal Attempt / Hat-trick / Kick-off / Offside / Etc…

3.2. Ontology Data Property

The Data_Property represents a part of ontology that describes the features information related to individual's concepts. In our case study, this part includes all the properties extracted from deep learning algorithms and used in SWRL rules. We divide our data property into three main categories: detected_object_properties, Frame_properties and video_sequences properties. Table 4 describes all sub-classes of our data_property hierarchy.

Table.4. Data_Property_Hierarchy

L1	L2	L3
Top-DataProperty	Detected_Objects_Properties	Bottom_Left_Point_X / Bottom_Right_Point_Y / Detected_In_Frame / Started_F / Etc…
	Frame_Properties	Number_Frame / Number_Of_Pixel / Etc…
	Video_Sequence_Properties	Number_Of_Frame / Started_F_Goal_Event / Etc…

3.3. Ontology Objects Property

The Object_Property represents the relationship between all the concepts of our ontology. These relationships are used in the different SWRL rules to detect the event of Offside or Not Offside. The Object_Property is divided into two parts related to the nature of the relationship: detected_Object_With_Frame or Frame_With_Video_Sequence, as illustrated in Table 5.

Table.5. Objects_Property_Hierarchy

L1	L2	L3
Top-Object _Property	Detected_Objects_With_Frame	Player_Detected_In / Assistant-referee_Detected_In / Official-referee_Detected_In / Etc…
	Frame_With_Video_Sequence	Belong / VS_Contained_F / Etc…

4. Comparative Study of our ontology with background

Using a set of metrics, Table IV illustrates a comparison study of our semantic ontology with other ones presented in the state of the art. our ontology development is based on the idea to extend its use for further research works in the soccer domain and not only for Offside event detection. We also created our ontology to overcome all weaknesses presented in Table 6.

Table.6. Comparison of our ontology with background

Metrics/Works	**[7]**	**[10]**	**[11]**	**[12]**	**Our**
Consistency	OK	OK	OK	OK	OK
Formalism	OK	OK	OK	OK	OK
Conceptualisation	OK	OK	OK	OK	OK
Large Coverage	OK	OK			OK
SWRL Use	OK				OK
Audio Part					OK

As demonstrated in table 6, our ontology is more complete than the others in the background and tries to overcome all concepts related to the soccer domain. Moreover, the development of our ontology was created to respect consistency, formalism, and conceptualization metrics, as presented in section III. Furthermore, our ontology used SWRL rules for inferring new knowledge as an offside event, as described in section IV. Finally, our ontology includes an audio part that is missed in the other ontologies.

5. Our offside detection system architecture

To develop a robust offside detection system, the ontology approach proposed here, supported by SWRL rules, represents the core module in the global system architecture, as shown in Fig. 2. Our ontology ensures the reception of low-level features extracted from video and audio deep learning modules and processes these data to decide if offside events exist in the video sequences.

As shown in Fig. 2, the indexing process initiates when the video and audio analysis module extract the different features from the video sequence using a Mask R-CNN [33] and Log-Mel spectrogram [34] with 2D-ResNet [35] models. The ontology considers these features as an input, and create DataProperty and ObjectProperty, frame, video sequence, etc (Ontology Population step). Finally, once all data filled, then, the reasoner of our ontology using a set of SWRL rules, infer this video sequence into the appropriate video event class (offside or not offside) with start and end event frames.

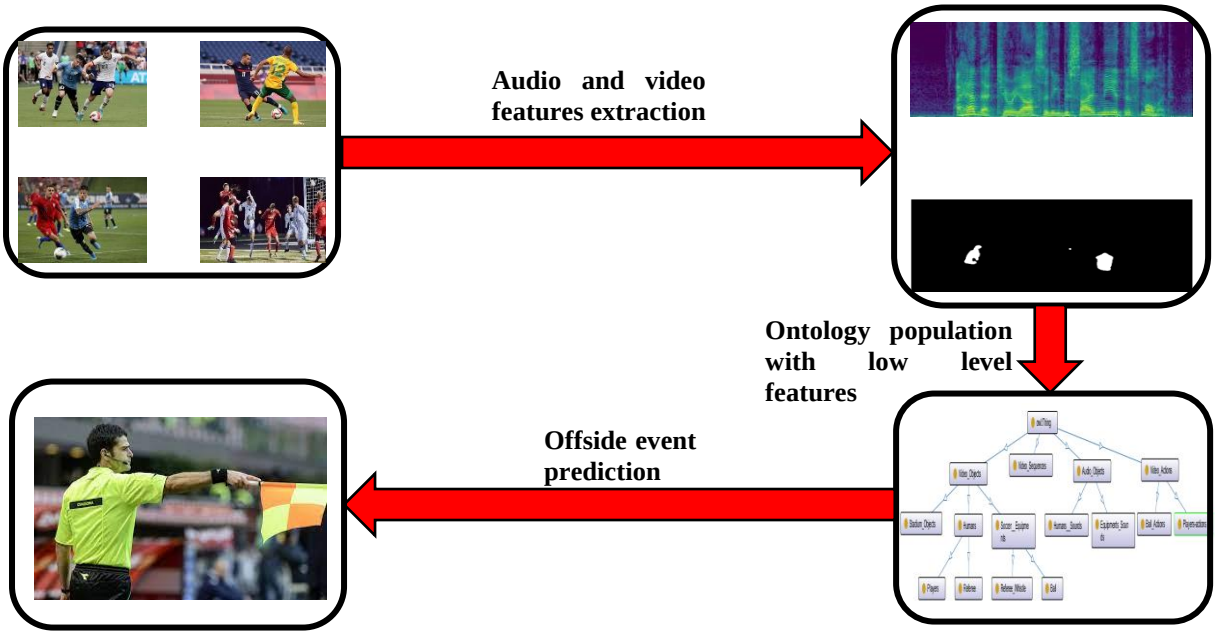

Fig. 2. Our soccer offside detection system.

5.1. Video extraction features module

The motivation of this work is to create a system that use both Deep Convolutional Neural Network and ontology paradigm to detect offside event in soccer domain. In soccer, the position of the ball and players is fundamental to study the event detection. Indeed, tracking the ball and different players over multiple successive frames can be used to detect the existence of this one. In image processing domain, Convolutional Neural Network (CNN) that represents a special type of deep neural network that is suited. In fact, convolution allows us to extract appropriate features from the input videos (frames); then pooling reduces dimensionality of the feature maps but keep the most important information; where forming fully connected layers allows us to ensure connections to all activations in the previous layer. In this work, we apply Mask-R-CNN that represents an extension extension of Faster R-CNN that concerns semantic segmentation, object localization, and object instance segmentation of natural images. This method achieves detection of objects to output bounding boxes, labels and masks like presented in Fig. 3. We start with features extraction using ResNet-101 convolutional networks architectures as a backbone. Then a network head for bounding-box recognition where for each RoI (Region of Interest), we apply mask prediction.

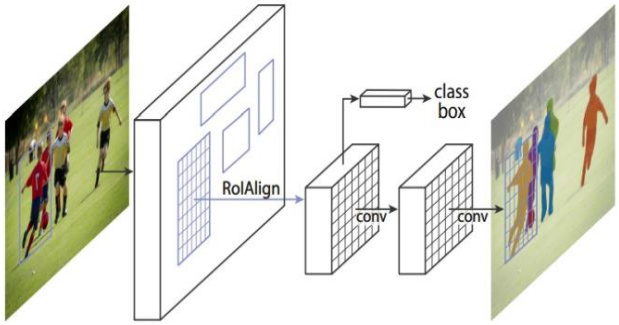

Fig. 3. The Framework of Mask R-CNN [33]

Mask R-CNN adds a third branch that outputs the object mask. Here, the goal is to classify individual objects and localize each one using a bounding box, and semantic segmentation in order to classify each pixel into a fixed set of categories without differentiating object instances as presented in Fig. 4.

Fig. 4. Application of Mask R-CNN [33]

5.2. Audio extraction features module

In the audio processing domain, the best way to perform an audio stream represents the transformation of the audio stream into a Log-Mel spectrogram and analysing it with the 2D-ResNet model. In this paper, we follow this idea and develop an audio extraction features module like described in Fig. 2. In the same way, as we train our visual models, audio is extracted first, from video in wave-form. After that, Log-Mel spectrograms is generated and represent the input of CNN plus Linear Classifier model (2D-Resnet). After a training set, the output of this audio module represents the class of audio in our domain (a referee whistle for example). Fig. 5, below. Represents the processus of sound classification step by step.

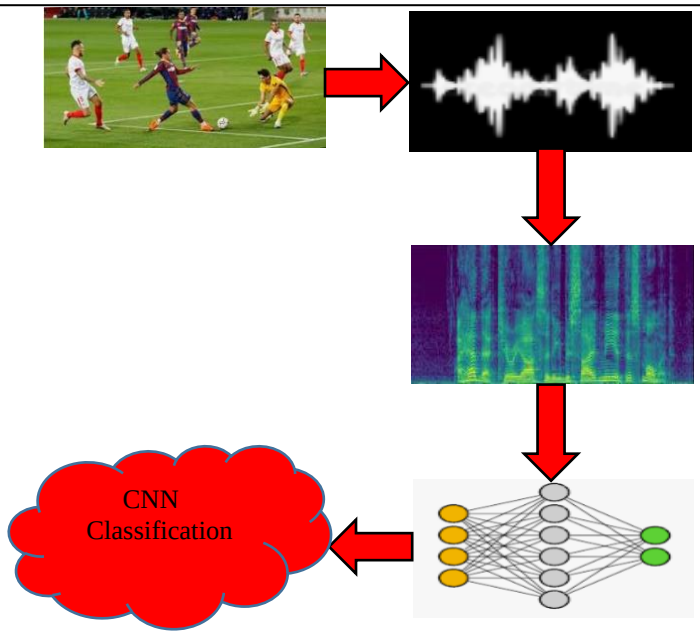

Fig. 5. Audio classification processus

5.3. Ontology module

The ontology module created with Protege2000 editor [36] receives all image and audio features from both previous modules to create all individuals, data properties, and object properties. After that, a pellet reasoner plugin [37] executes adequate SWRL rules for inferring if this video sequence can be annotated with an offside event or not. For this purpose, we divide our SWRL rules into three categories: ball and player tracking, player position, and video sequences event.

The first category represents the tracking process of the ball and the players for inferring the nature of players (striker or defender) and passing/receiving ball data. For example, Rule 1 below presents an example of this category.

Rule 1:

Video_sequence (? V1), Frame(? F1), Frame(? F2), Ball (? B), Player (? P1), Player (? P2), Player (? P3), Player_Dectected_In (?P1, ?V1), Player_Dectected_In (?P2, ?V1), Player_Dectected_In (?P3, ?V1), B_Dectected_In (?B, ?V1), Frame_Detected_In(?F1, ?V1), Frame_Detected_In(?F2, ?V1), Center_Started_Position(?P1, ?a) , Center_Finished_Position(?P2, ?b), Center_Started_Position(?P3, ?c), Center_Finished_Position(?P3, ?d), Center_Started_Position(?B, ?e), Center_Finished_Position(?B, ?f), Player_Started_Position_Frame(?P1, ?F1), Player_Started_Position_Frame(?P3, ?F1), Ball_Started_Position_Frame(?B, ?F1), Player_Ended_Position_Frame (?P2, ?F2), Player_Ended_Position_Frame (?P3, ?F2), Ball_Ended_Position_Frame (?B, ?F2), swrlb:add(?g, ?e, 25), swrlb:subtract(?h, ?e, 25), swrlb:greaterThan(?a, ?h), swrlb:lessThan(?a, ?g), swrlb:greaterThan(?c, ?g), swrlb:lessThan(?c, ?h), swrlb:add(?i, ?f, 25), swrlb:subtract(?j, ?f, 25), swrlb:greaterThan(?b, ?j), swrlb:lessThan(?b, ?i), swrlb:greaterThan(?c, ?i), swrlb:lessThan(?c, ?j) → Striker (? P1), Striker (? P2), Defender (? P3), Passing_Ball(?P1, ?B1), Receiving_Ball(?P2, ?B1), Passing_Ball_Frame(?F1), Receiving_Ball_Frame(?F2).

The second category represents the inferring process of the position of each player. The input of this category is taken from the output of the first category. Rule 2 illustrates an example of this process.

Rule 2:

Video_sequence (? V1), Frame(? F1), Frame(? F2), Frame(? F3), Ball (? B), Striker (? P1), Striker (? P2), Defender (? P3), Referee_Whistle_Sound(?S1), Player_Dectected_In (?P1, ?V1), Player_Dectected_In (?P2, ?V1), Player_Dectected_In (?P3, ?V1), B_Dectected_In (?B, ?V1), Frame_Detected_In(?F1, ?V1), Frame_Detected_In(?F2, ?V1), Frame_Detected_In(?F3, ?V1), Started_Sound_Frame(?F2), Ended_Sound_Frame(?F3), Passing_Ball(?P1, ?B1), Receiving_Ball(?P2, ?B1), Passing_Ball_Frame(?F1), Receiving_Ball_Frame(?F3), X_Position(?P1, ?a), X_Position(?P2, ?b), X_Position(?P3, ?c), swrlb:greaterThan(?c, ?a), swrlb:lessThan(?c, ?b) → Offside(?P2) , Started_Player_Offside(?F1), Ended_Player_Offside(?F3)

The last category of SWRL rules represents the events detection part (Offside or Not_Offside with start and end frames). Rule 3 demonstrates an example of this category.

Rule 3:

Video_sequence (? V1), Striker (? P2), Frame(? F1), Frame(? F2), Player_Dectected_In (?S2, ?V1), Frame_Detected_In(?F1, ?V1), Frame_Detected_In(?F2, ?V1), Offside(?P2) , Started_Player_Offside(?F1), Ended_Player_Offside(?F2) → Offside (?V1), Started_Offside_Event(?F1), Ended_Offside_Event(?F2)

Rule 1, 2, and 3 present only an example of SWRL rules used. In fact, more than 150 rules are used to handle all possible positions of players and ball.

6. Experiment results

With the aim of experimenting with the efficiency of our Offside detection system, we selected four soccer videos from the web for more than six hours. We developed an application in the Java environment that handles all the steps of our indexing and retrieval system. The process started with the selection of the different videos and ended with the indexing results. The tests were performed on a machine with an Intel Core I7 CPU and 32 GB RAM, under Windows 11. We considered three types of evaluations to check the performance of our system:

6.1 Evaluation based on the offside events

The first type of evaluation was based on the number of events returned by the OVIS system; it is carried out by many metrics, such as Precision, Recall, F-measure. We considered these measures as follows:

$$\text{Precision} = \frac{\text{Number of detected videos that contain the offside event}}{\text{Number of videos indexed with the offside event}}$$

$$\text{Recall} = \frac{\text{Number of detected videos that contain the offside event}}{\text{Number of all videos in database that contain the offside event}}$$

$$\text{F} - \text{mesure} = 2 * \frac{(\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})}$$

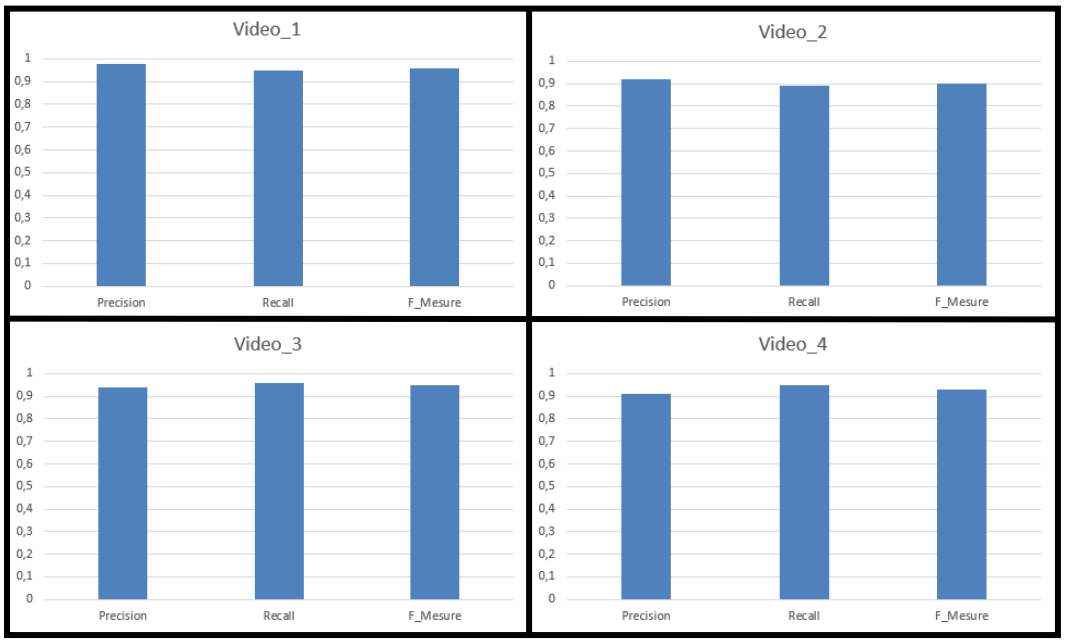

Fig. 6. Offside events evaluation results

In Fig. 6 above, we summarized the statistics of the obtained results from the four videos experimented. On one hand, offside detection event provides excellent results in term of Precision, Recall and F-measure. However,

the F-measure metric expresses the relation between precision/recall. Consequently, the F-measure metric provides excellent result and reach 90% of measure in the four videos. Therefore, these results mean that the combination of both video and audio features, supplied by ontology paradigm, represents a verry powerful approach of offside detection event.

6.2 Evaluation based on frame timing

The second type of evaluation represents frame timing using an overlay (in frame number) of 12 frames to evaluate the performance of our system. To meet this end, three kinds of situations are considered:

• Too Early: Our system detects offside events before they really start (ground truth) with 12 frames.

• On Time: Our system detects offside events exactly when they start.

• Too Late: Our system detects offside events after they really start with 12 frames.

For this purpose, another set of four videos are used from the web.

Fig. 7, represents the frame timing events of the fourth videos that concern offside event (Ground Truth and results). The ground truth is generated manually by watching the entire four videos.

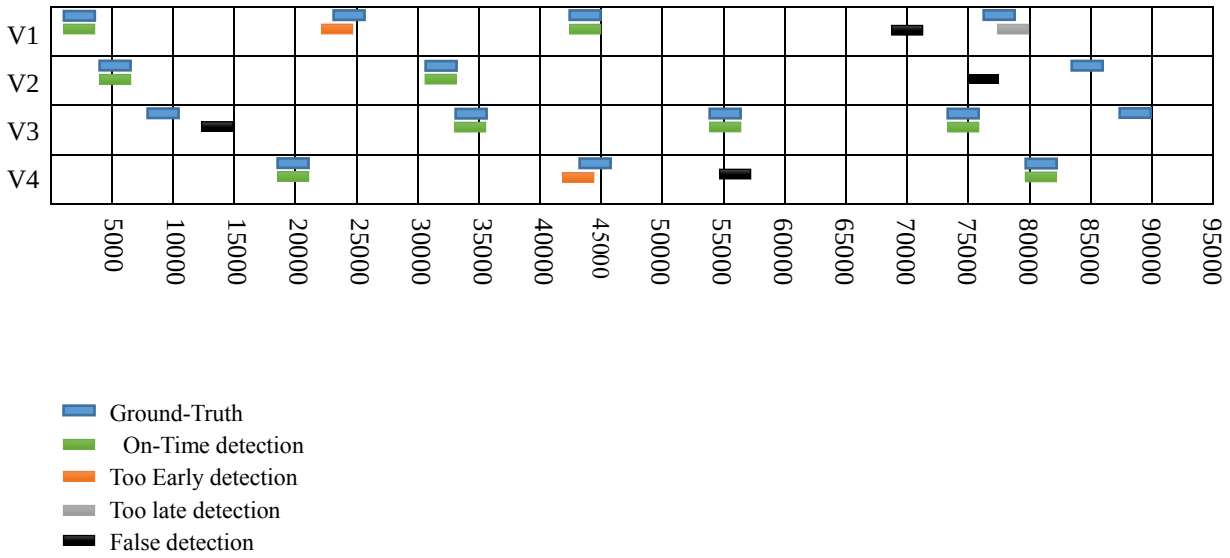

Fig. 7. Frame timing evaluation results

Fig.7 above illustrates the results for the frame timing evaluation obtained from the output of our system in comparison with the ground truth. As described, the first advantage is the detection of at least two offside events on time situation (Our system detects offside events exactly when they start), without exceeding the overlay of 12 frames. These results, demonstrate another strength of our approach, that the other works in the state of the art don't perform.

7.Conclusion

Offside detection events become a more and more attractive challenge in the soccer domain. Therefore, many researchers in the field lead their research intending to present the best approach for offside detection purposes. In this paper, we present our offside detection system based on both audio and video features handled by deep learning algorithms. Thus, these features represent the input ontology paradigm that processes low-level data and decides if the video clips contain offside events or not. Moreover, in comparison with the state of the art, our system infers also the started frame and ended frame of each offside event detected. Consequently, the results of the experiment described in this paper demonstrate the efficiency of our approach. In future work, we will extend our system to detect other events in the same field of soccer like a goal for example, or other domains like scream events in surveillance.

References

[1] Abdullah Khan, Beatrice Lazzerini, Gaetano Calabrese, and Luciano Serafini. Soccer Event Detection. In International Conference on Image Processing and Pattern Recognition (IPPR), April 2018.

[2] H. Jiang, Y. Lu, and J. Xue. Automatic soccer video event detection based on a deep neural network combined cnn and rnn. In Tools with Artificial Intelligence (ICTAI), 2016 IEEE 28th International Conference on, pages 490–494. IEEE, 2016.

[3] W. Song and H. Hagras, "A type-2 fuzzy logic system for event detection in soccer videos," *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2017, pp. 1-6.

[4] Khan MZ, Hassan MA, A Farooq and Khan MU, "Deep CNN Based Data-Driven Recognition of Cricket Batting Shots", *2018 International Conference on Applied and Engineering Mathematics (ICAEM)*, pp. 67-71, 2018 Sep 4.

[5] S. Ma, E. Shao, X. Xie and W. Liu, "Event Detection in Soccer Video Based on Self-Attention," *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*, 2020, pp. 1852-1856.

[6] K. Muhammad *et al*., "AI-Driven Salient Soccer Events Recognition Framework for Next Generation IoT-Enabled Environments," in *IEEE Internet of Things Journal*, 2021.

[7] Tani, L. F. K., Ghomari, A., & Tani, M. Y. K. (2019). A semi-automatic soccer video annotation system based on Ontology paradigm. In 2019 10th International Conference on Information and Communication Systems (pp. 88–93).

[8] Tani, M. Y. K., Ghomari, A., Youcef, L. D., Lablack, A., & Bilasco, I. M. (2017, July 18–20). An audio indexing and retrieval approach using a video surveillance ontology. In Computing conference (pp. 258–261). IEEE.

[9] M. Y. Kazi Tani, A. Ghomari, A. Lablack and I. M. Bilasco, "OVIS: ontology video surveillance indexing and retrieval system", *Int. J. Multimed. Inf. Retr.*, vol. 6, no. 4, 2017.

[10] Wu, H., Cheng, G., Qu, Y. 2006 Falcon-S: An ontology-based approach to searching objects and images in the Soccer domain. Supplemental Proceedings of ISWC, Nov. 2006.

[11] P. Abreu, M. Faria, L. P. Reis and J. Gargarnta, "Knowledge representation in soccer domain: An ontology development," 5th Iberian Conference on Information Systems and Technologies, 2010, pp. 1-6.

[12] Soner Kara, Özgür Alan, Orkunt Sabuncu, Samet Akpınar, Nihan K. Cicekli, Ferda N. Alpaslan, An ontology-based retrieval system using semantic indexing, Information Systems, Volume 37, Issue 4, 2012, Pages 294-305.

[13] Javier Fernández, Luke Bornn, and Dan Cervone. 2019. Decomposing the Immeasurable Sport: A Deep Learning Expected Possession Value Framework for Soccer. In MIT Sloan Sports Analytics Conference.

[14] P.R. Kamble, A.G. Keskar, K.M. Bhurchandi, A deep learning ball tracking system in soccer videos, Opto-Electronics Review, Volume 27, Issue 1, 2019, Pages 58-69.

[15] Fernández J., Bornn L. (2021) SoccerMap: A Deep Learning Architecture for Visually-Interpretable Analysis in Soccer. In: Dong Y., Ifrim G., Mladenić D., Saunders C., Van Hoecke S. (eds) Machine Learning and Knowledge Discovery in Databases. Applied Data Science and Demo Track. ECML PKDD 2020. Lecture Notes in Computer Science, vol 12461. Springer, Cham. https://doi.org/10.1007/978-3-030-67670-4_30.

[16] R. Agyeman, R. Muhammad and G. S. Choi, "Soccer Video Summarization Using Deep Learning," 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), 2019, pp. 270-273, doi: 10.1109/MIPR.2019.00055.

[17] Yu J., Lei A., Hu Y. (2019) Soccer Video Event Detection Based on Deep Learning. In: Kompatsiaris I., Huet B., Mezaris V., Gurrin C., Cheng WH., Vrochidis S. (eds) MultiMedia Modeling. MMM 2019. Lecture Notes in Computer Science, vol 11296. Springer, Cham. https://doi.org/10.1007/978-3-030-05716-9_31.

[18] Hubáček O., Šourek G., Železný F. (2019) Deep Learning from Spatial Relations for Soccer Pass Prediction. In: Brefeld U., Davis J., Van Haaren J., Zimmermann A. (eds) Machine Learning and Data Mining for Sports Analytics. MLSA 2018. Lecture Notes in Computer Science, vol 11330. Springer, Cham. https://doi.org/10.1007/978-3-030-17274-9_14.

[19] Ali Karimi, Ramin Toosi, and Mohammad Ali Akhaee. 2021. Soccer Event Detection Using Deep Learning. arXiv preprint arXiv:2102.04331 (2021).

[20] T Thamaraimanalan, D Naveena, M Ramya, & M Madhubala. Prediction and Classification of Fouls in Soccer Game using Deep Learning Irish Interdisciplinary Journal of Science & Research, volume 4, p. 66 – 78.

[21] Jain S., Tiwari E., Sardar P. (2021) Soccer Result Prediction Using Deep Learning and Neural Networks. In: Hemanth J., Bestak R., Chen J.IZ. (eds) Intelligent Data Communication Technologies and Internet of Things. Lecture Notes on Data Engineering and Communications Technologies, vol 57. Springer, Singapore. https://doi.org/10.1007/978-981-15-9509-7_57

[22] B. Sheng, P. Li, Y. Zhang, L. Mao and C. L. P. Chen, "GreenSea: Visual Soccer Analysis Using Broad Learning System," in *IEEE Transactions on Cybernetics*, vol. 51, no. 3, pp. 1463-1477, March 2021, doi: 10.1109/TCYB.2020.2988792.

[23] Feng, N., Song, Z., Yu, J. et al. SSET: a dataset for shot segmentation, event detection, player tracking in soccer videos. Multimed Tools Appl 79, 28971–28992 (2020). https://doi.org/10.1007/s11042-020-09414-3.

[24] K. Tang, Y. Bao, Z. Zhao, L. Zhu, Y. Lin and Y. Peng, "AutoHighlight : Automatic Highlights Detection and Segmentation in Soccer Matches," *2018 IEEE International Conference on Big Data (Big Data)*, 2018, pp. 4619-4624, doi: 10.1109/BigData.2018.8621906.

[25] Fakhar, B., Rashidy Kanan, H. & Behrad, A. Event detection in soccer videos using unsupervised learning of Spatio-temporal features based on pooled spatial pyramid model. Multimed Tools Appl 78, 16995–17025 (2019). https://doi.org/10.1007/s11042-018-7083-1

[26] Y. Hong, C. Ling and Z. Ye, "End-to-end soccer video scene and event classification with deep transfer learning," *2018 International Conference on Intelligent Systems and Computer Vision (ISCV)*, 2018, pp. 1-4, doi: 10.1109/ISACV.2018.8369043.

[27] Nergard Rongved, O.A.; Stige, M.; Hicks, S.A.; Thambawita, V.L.; Midoglu, C.; Zouganeli, E.; Johansen, D.; Riegler, M.A.; Halvorsen, P. Automated Event Detection and Classification in Soccer: The Potential of Using Multiple Modalities. Mach. Learn. Knowl. Extr. 2021.

[28] Zengkai Wang, Semantic analysis based on fusion of audio/visual features for soccer video, Procedia Computer Science, Volume 183, 2021, Pages 563-571, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2021.02.098.

[29] X. Gao *et al*., "Automatic Key Moment Extraction and Highlights Generation Based on Comprehensive Soccer Video Understanding," *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 2020, pp. 1-6, doi: 10.1109/ICMEW46912.2020.9106051.

[30] Vanderplaetse, Bastien and Dupont, Stephane, Improved Soccer Action Spotting Using Both Audio and Video Streams. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. June, 2020.

[31] N. Stefanakis, Y. Mastorakis, A. Alexandridis, and A. Mouchtaris, "Automating Mixing of User-Generated Audio Recordings from the Same Event," J. Audio Eng. Soc., vol. 67, no. 4, pp. 201-212, (2019 April.). doi: https://doi.org/10.17743/jaes.2019.0008.

[32] Melissa Sanabria, Sherly, Frédéric Precioso, and Thomas Menguy. 2019. A Deep Architecture for Multimodal Summarization of Soccer Games. In Proceedings Proceedings of the 2nd International Workshop on Multimedia Content Analysis in Sports (MMSports '19). Association for Computing Machinery, New York, NY, USA, 16–24. DOI: https://doi.org/10.1145/3347318.3355524.

[33] K. He, G. Gkioxari, P. Dollar, R. Girshick. (2017). Mask r-cnn. arXiv:1703.06870.

[34] Purwins, H.; Li, B.; Virtanen, T.; Schlüter, J.; Chang, S.; Sainath, T. Deep Learning for Audio Signal Processing. IEEE J. Sel. Top. Signal Process. 2019, 13, 206–219.

[35] He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

[36] Protege. The protege project. http://protege.stanford.edu

[37] Sirin EB, Parsia B, Cuenca Grau B, Kalyanpur A, Katz Y (2003) Pellet: a practical OWL-DL reasoner. J Web Semantics 5:51–53.