Highlights Extractor

CERTIFIED SPECIALIST
IN
MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

Exit Test 2
ICT Academy of Kerala

Abstract

Video analysis is extensively being used for several applications in different domains like sports, entertainment, healthcare, etc. In sports it helps to perform in-depth analysis of scenarios and human behaviors to learn from or plan strategies.

Here, a deep learning method is used to design a football event detection algorithm for football game video analysis. The algorithm can automatically detect and classify various events (goals, happy moments and loss moments) in football game videos. Among them, the three-dimensional convolution neural network is used for feature extraction, which can process multiple frames of images at the same time, so as to retain relevant information between frames. It uses a bidirectional recurrent network to integrate features from both positive and negative directions to obtain past and future contextual information to improve the effect of event detection.

Problem Statement

Description

Scenario: Sara is a football coach and her team has recently qualified for semi-finals of a most Prestigious match of the country. She wants to plan the strategy for the next game based on the video Analysis done on the games of the opponent team. Sara has identified three scenarios to be analyzed – ‘Goal moments’, ‘Happy moments’, ‘Loss moments’. The analysis would help to learn about the Techniques and strategy of the opponent team and hence decide on the strategy for her own team.

Requirements: An application to extract the identified scenarios – Goal moments, happy moments and Loss moments – from the input video. The extracted video segments would be of 30 s to 1 min duration and would be capturing the emotions of the people in the video as well. Extraction should happen in such a way that if "goal" is at "x" min in the video, the extracted segment should start from "x-y" min of the input video and end at "x+y" where "y" is the time in seconds to be considered before and after the shot.

Outcome expected

• Working prototype

o Code
o Documentation
o Video

Literature Survey

There are many studies, researches and papers to extract highlights or important events from a sports video.

"Event Detection Based Approach for Soccer Video Summarization Using Machine Learning" by Hossam M. Zawbaa, Nashwa El-Bendary, Aboul Ella Hassanien [**1**] is one of the early solutions which analysis soccer videos with the extraction of valuable semantics by efficient and effective processing of a combination of visual, audio and text information. This paper presents an approach for automatic soccer video summarization using machine learning techniques. In order to generate effective summaries for soccer videos, the proposed system initially segments the whole video stream into small video shots. Then, the system applies a support vector machine (SVM) algorithm for emphasizing important segments with logo appearance with addition to detecting the caption region providing information about the score of the game. Subsequently, the system uses K-means algorithm and Hough line transform for detecting vertical goal posts and Gabor filter for detecting goal net. Finally the system highlights the most important events during the match.

In **[2],** an article published for scene understanding on football matches from video feed using ML and traditional CV techniques they split the problem into different segments such as

1. Reference system and homography estimation (how to project players' position from camera-view to a 2D plane).
2. Object detection (aka what and where is the players/ball/referee).
3. Object tracking (aka how do I track entities across frames).
4. Player Identification (aka how do I recognize the players across frames).
5. Team Recognition (how do I figure what team a player plays for).

They use YOLOv3 for object detection. YOLOv3 (You Only Look Once, Version 3) is a real-time object detection algorithm that identifies specific objects in videos, live feeds, or images. Latest version of YOLOv7 [3] is the most powerful object detection algorithm as of now.

"Football Game Analysis Method with Deep Learning" by Nian Liu, Lu Liu and Zengjun Sun [4] proposes a model which is divided into two stages, in which the first stage is utilized to generate candidate event fragments. It divides the football video to be detected into a sequence of frames of a certain length and scans using a sliding window. Multiple frame sequences within a sliding window form a segment, and each segment is a prediction unit. The frame sequence features within the segment are obtained through a three-dimensional convolution neural network, which is used as the input of each time point of the bidirectional recurrent neural network and further integrated to generate the event prediction of the segment. The second stage is to further process the above results to remove all segments predicted as nonevents. The thresholds are set according to the detection effect of various events to filter out event fragments with higher probability values, obtain the start and end positions of the events through merging, classify and mark them, and finally output complete event fragments. This work had carried out comprehensive and systematic experiments to verify correctness of the proposed method.

Proposed Solution

HIGHLIGHTS EXTRACTOR

1. Requirement:

An application to extract the identified scenarios – Goal moments, Happy moments and Loss moments – from the input video.

2. Event Recognition

Event recognition is conventionally formulated as a classification problem. The input is a video of football. We have to extract individual frames and classify according to various expressions and reactions of players, the referee and crowd that contain a

specific action (Goal, Happy, and Loss) of interest, and the goal is to trim the video accordingly.

The event detection model to detect events contained in football videos, is divided into two stages.

(1) Classification of football events - The classification model employs the 3D CNN network and the Softmax classifier for feature extraction and predictive classification for event segments, respectively, (2) Football event detection - the event detection model is based on the classification model by adding the BLSTM structure to better obtain dynamic information between multiple frame sequences.

3. Football Event Detection

Event detection is the process of locating the time boundary of an event in a complete football video and then classifying it. In the model, the video is divided into frame sequences of a specific length (segment), the entire video is scanned by a sliding window and the starting position of the event is predicted by extracting and integrating the features of multiple frame sequences.

4. Football Events Classification

The event classification model combines a 3D CNN and a Softmax classifier, the input is an event segment, and the sequence features are extracted from the original frame image through 3D CNN. The probability value is calculated using the Softmax classifier to get the predicted classification of the event. After detecting the exact type of event, depending on the type of event we can add on +/- 15s for an emotional /happy/loss moment and +/- 30s for a goal moment.

Fig 1: Detection and Classification Sequence

In this proposed solution, the output of the model/container can be customized based on requirements. In this model, we can find the goal moments using the goal detection solution and the happy and loss moments using the sentiment/emotion

analysis of the players, referee and the crowd. Depending upon the request, we can further pick and choose the appropriate output from the container/model

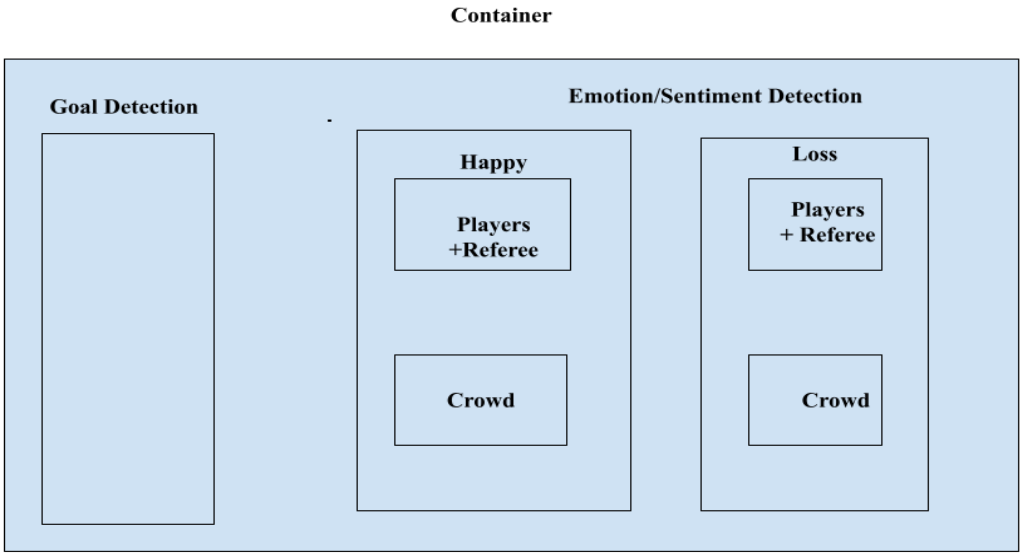

Fig:2 Proposed System Design

References

1. https://www.researchgate.net/publication/232252059_Event_Detection_Based_Approach_for_Soccer_Video_Summarization_Using_Machine_learning
2. https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-021-00295-6
3. https://viso.ai/deep-learning/yolov7-guide/
4. https://www.hindawi.com/journals/cin/2022/3284156/