Hindawi

Research Article

A Deep Learning Algorithm for Special Action Recognition of Football

Sheng Wang ⓘ

College of Arts and Sciences, Yangtze University, Jingzhou 434020, China

Correspondence should be addressed to Sheng Wang; shengwang83@yangtzeu.edu.cn

Soccer (football) is a popular form of exercise on the planet. There are a lot of individuals who tune into football matches in real time on television or the Internet. A game of American football lasts 90 minutes, but to save time, spectators may simply want to see a few highlights. As far as we know, no such tool exists that can be used to extract intelligent highlights from a football match. In this research, we present a technique for clever editing of live football matches. Our technology allows for the automatic extraction of key players' goals, shots, corner kicks, red and yellow cards, and the presence of key players from a football match's live stream. During the 2018 FIFA World Cup, our solution was integrated into live streaming platforms and it functioned admirably.

1. Introduction

Extensive research has been conducted on video analytics technology in order to provide customers with more rapid and accessible access to engaging or critical segments of films [1–3]. The demand for high-performance image and video indexing and retrieval solutions has skyrocketed as the quantity of multimedia videos has increased tremendously. From their perspective, the video summary is very important to them [4]. Users just need to view a few crucial areas to save time. It requires considerable time and effort to manually evaluate and summarize video recordings. Given the number of distinct sequences and the length of time necessary, an automated motion video sequence highlighting strategy is quite valuable.

Sports video games are one of the most studied video genres due to their large audience and more consistent characteristics than other video genres [5–8]. The high-lighting strategy, which focuses on how to compose a match recap that incorporates all of its important aspects, is an effective abstraction technique for sports film [9, 10]. Previous researchers have provided several highlighting strategies for a wide range of sports contests, from the most basic to the most specialized ones. To conclude, in sports video

footage, Ekin and Tekalp discovered play and break events [11]. Other research studies utilized slow-motion playback to describe sports videos [12–14]. However, evaluating general sports games remains difficult due to the variety and complexity of the games. Basketball, diving, and football are just a few of the sports in which academicians have chosen to concentrate. This research focuses on methods for extracting highlights from football videos. SVM classifiers were suggested by Ancona et al. as a tool for recognizing soccer targets [15]. When a video sequence is divided into distinct shot types, it is then organized into smaller video shots using the technique described by Zawbaa and colleagues [16]. Using SVM and an artificial neural network approach, the system then picks segments with outstanding performance. The device subsequently detects the vertical goalpost and goal net. Finally, in the football video recap, the game's most significant moments will be highlighted. Focusing on low-level and text-based processing, for soccer films, Fendri et al. developed a segmentation and indexing system based on scientific principles [17]. Automated methods for assessing and summarizing football films based on cinematic- and object-based criteria have been suggested by Ekin et al. [18]. In order to summarize the film, Lofti and Pourreza came up with a way of removing unnecessary footage [19]. Tabii and

Thami proposed a new approach for automatically creating summaries of soccer videos based on lens recognition, categorization of lenses, and finite state machine technology [20]. All of the following methods process the video using artificially created qualities that lack generalizability and are challenging to implement in reality. An innovative "One to Key" notion developed by Yan et al. aims to enhance the detection of group activity by aggregating the temporal dynamics between key players with varying levels of involvement across time [21]. It is quick and knowledgeable, but it cannot recognize the difference between the game's best moments. Automated and intelligent lives are becoming more commonplace, and future multimedia processing technologies will be more sophisticated and efficient.

In recent years, deep learning has been widely used in image processing [22–25] and pattern recognition [26–29]. Using deep learning, we have developed a technology that allows for intelligent editing that creates highlights from watching a football game on TV or watching it live online while also resolving the concerns that an in-depth study could cause. Feature extraction techniques do not need to be created by hand, and algorithms do not need to be updated for different scenarios, as was the case with previous methods. After the live broadcast is over, the technology may produce a wide variety of entertaining short videos, such as goal celebrations, penalty kicks, red and yellow cards, corner kicks, and so on. The method achieved will not be forgotten during the FIFA World Cup 2018!

2. System Overview

Deep learning plays a positive role in football teaching. The training objective of deep learning is consistent and relevant to the teaching characteristics of football. Deep learning can stimulate students to actively participate in classroom teaching and actively observe, think, summarize, internalize, and practice what they have learned, which is in line with the teaching characteristics of football projects. Teachers should actively design teaching situations in football teaching, let students practice football skills in themed, targeted, and practical teaching situations, and develop the students' high-level sports ability by completing challenging learning tasks. In the process of developing students' football skills, we should promote the all-round development of students through continuous evaluation, realize the multidimensional value of sports, and cultivate students' lifelong sports awareness. In order to capture the best moments from live football games, we created an automated editing system. Live footage of a football match is used to create shot boundary frames, special action frames, and star player frames by using shot segmentation, red and yellow card recognition, corner kick detection, penalty kick detection, shoot and celebration detection, score detection, and face identification. Afterward, the data are integrated to provide highlights in an integration module. Figure 1 depicts our system's flow diagram. To do this, we annotate the objects we need from football match films, such as the 2018 World Cup in Brazil, as well as footballs, players, and red and yellow cards.

3. Methods

3.1. Segmentation of Photographs. Since highlights are created by synthesizing continuous photos in order to avoid discontinuities, shot segmentation is at the heart of our technique. Figure 2's shot boundary frame is what we are looking for. For football match recordings, we employ a histogram technique to determine shot boundaries, which may reduce oscillations caused by the movement of objects in the frame. The most basic histogram approach compares the histograms of two sequential frames in grayscale or color, respectively. When two histograms are compared, the bin-wise difference between them is used to identify a shot boundary. To determine shot boundaries, Ueda et al. turned to the color histogram change rate [30]. In contrast, they just employ the absolute value ratio between two frames in the color histogram; hence, the outcome is insensitive.

A shot's average DAV is believed to be bigger than the difference D_{cur} between the shot's boundary frame and the frame immediately preceding it. D_{cur} and D_{avg} are made up of the following:

$$D_{\text{cur}} = \sum \frac{(h(c) - h(c-1))^2}{\max(h(c), h(c-1)) + \epsilon}, \quad c \geq 2,$$

$$D_{\text{avg}} = \frac{\sum_{i=2}^{c-1} D_{\text{cur}}(i)}{c-2}, \quad c \geq 3, \tag{1}$$

where $h(\cdot)$ is a histogram calculation function and c is the frame index. ϵ is an arbitrarily small positive quantity to prevent the denominator being 0. It was set to 0.0001. When the ratio of D_{cur} to D_{avg} exceeds a threshold, we use the last frame as a shot boundary frame. Cross-validation on a set of football match videos yielded the threshold.

3.2. Detection of Red/Yellow Cards. In several sports, infractions are punished by the presentation of a red or yellow card. The yellow card acts as a cautionary note, while the red card serves as an escape sign. If a player receives two yellow cards during a game, they will be issued with a red card.

Object recognition techniques must be used to identify red and yellow cards in football match recordings. Multistage, two-stage, and one-stage object identification techniques are available. Early examples of cross-strategic networks are R–CNN and SPPNet [31–33]. Each of the stages of a search may be taught on its own or in conjunction with the others. It is possible to train R–CNN simultaneously for feature extraction, location regression, and classification. For this reason, the two-stage process is referred to as a "two-step" process. The removed region-proposal network, one-stage network, and SSD of the Yolo series were able to significantly increase object recognition speed. SSD was used in the development of our object identification algorithm in order to achieve a good balance between speed and precision. For each frame of the football game video, as shown in Figure 3, we apply SSD and output the frames having red or yellow cards for analysis.
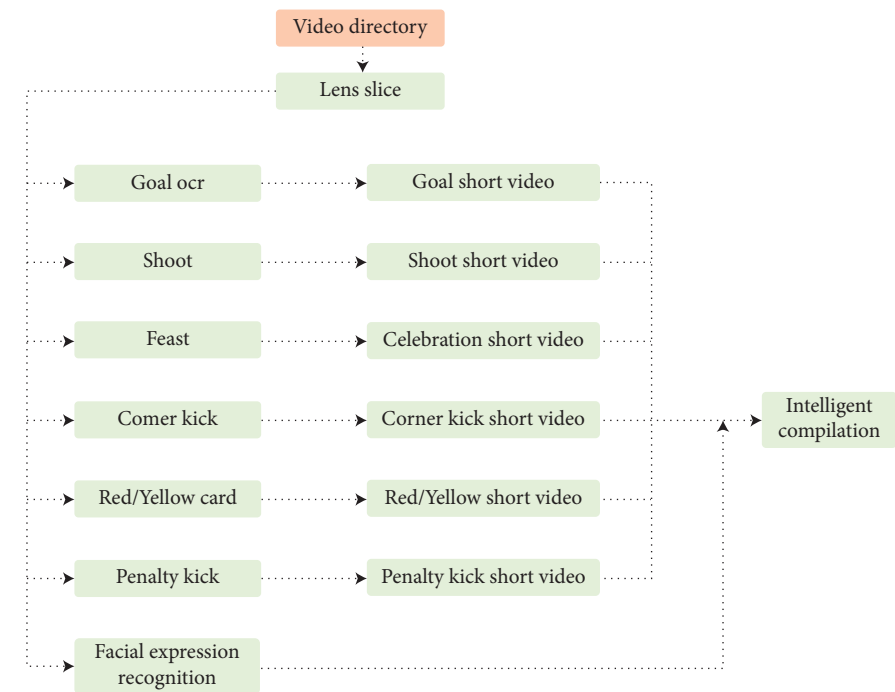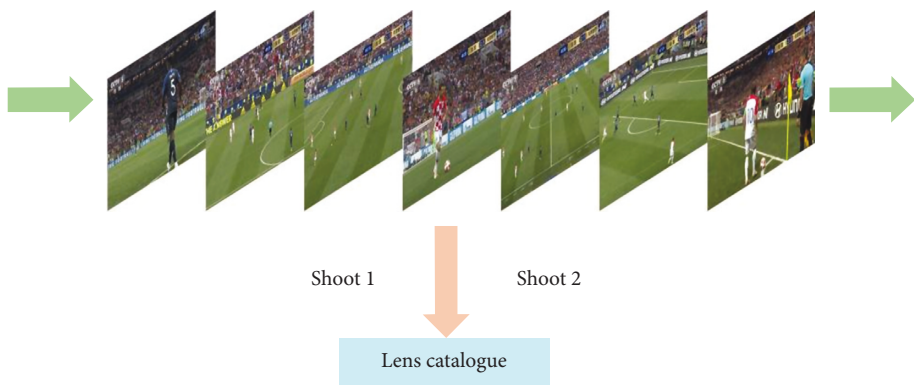
Figure 1: The overall system.

Figure 2: Shot segmentation. On each side of the red arrow, there are various shots. Boundary frames are used to describe the first frame of each shot.

Figure 3: Detection of a yellow or red card. Using the red and yellow boxes in the graphic, this module is able to identify the red and yellow cards in the frame.

3.3. Corner Kick. When the ball crosses the goal line without being scored on and is touched by one player from the opposing team, the game is restarted with a corner kick. Many fans enjoy watching highlights of corner kicks because they are considered excellent goal-scoring opportunities for the attacking team.

The SSD is also responsible for detecting corner kicks. A corner kick frame should show the player, ball, and corner flag all at the same time as seen in Figure 4. An SSD that has been taught to recognize what is in a frame is what we use to process frames. We output a corner kick keyframe when a player, the ball, and the corner flag can all be found in the same frame.

3.4. Penalty Kick Detection. In competitive football, the goal of the game is to win. Choosing a suitable tactic is very important for the preparation before the game. When analyzing tactical behavior, the concept of tactics is defined by some researchers as the actions taken by players to adapt to the dynamic changes of the game situation. Other researchers try to distinguish tactics from strategy and believe that strategy can be described as a preplanned competition element after considering the influence of available information. From this perspective, the difference between tactics and strategy lies in the relationship between them and time. The implementation of the strategy allows for longer planning and careful consideration, while the tactical behavior is done under strong time pressure. Tactics determine how a team manages space, time, and personal actions to win games. In this context, space refers to the specific actions that take place on the court or the area of the court that the team wants to cover during attack and defense. Time describes things such as the frequency and duration of the event (such as ball control) or the speed at which the action is initiated. Individual actions specify the type of action being performed, such as errors, passes, and passes. According to the number of players, it can be further divided into individual tactics, group tactics, team tactics, and game tactics. Tactical behavior analysis at the personal level can be used to study the kinematic relationship between a player and his opponents or teammates, such as the interaction between a full back marking a winger or two central defenders. At the group level, the team's tactical movements of the selected group of players can be analyzed, such as the offside "trap" created by the consistent movement of the guard line. At the team level, tactical behavior analysis can capture the variables of all players' team actions, such as analyzing the space occupied by the team. Tactical behavior at the game level can be studied by examining the interaction of team actions between two opposing teams, such as the distance between teams. In some cases, strategic success depends on tactical success at different organizational levels.

A penalty kick is a means of resuming a play in football in which a player is permitted to attempt a single shot to the goal while the other team's goalkeeper is the only one defending it. It is awarded when a player commits a foul in his or her own penalty area that results in a straight free kick. This module also makes use of the SSD. As indicated in
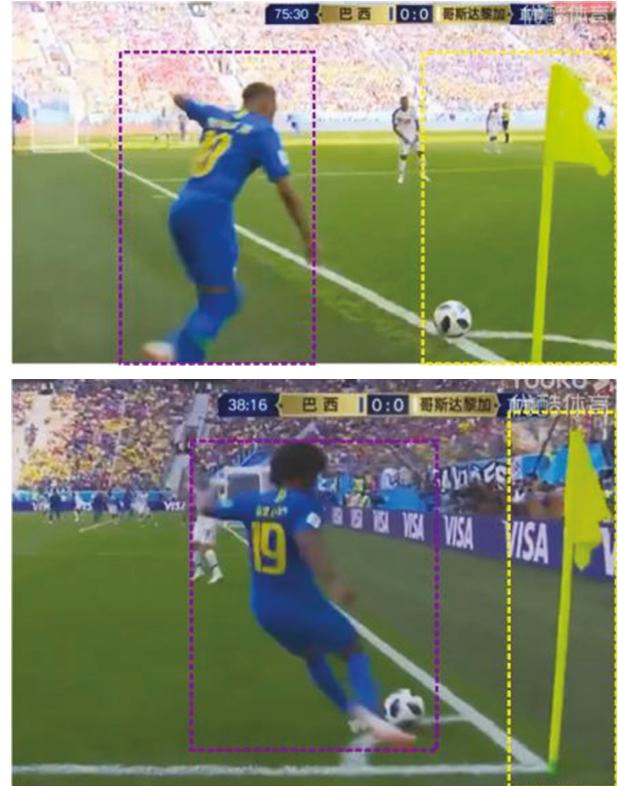

FIGURE 4: Slamming the ball into the corner. The players, the ball, and the corner flag are shown in the picture through the use of the red, blue, and yellow boxes in the frame.

Figure 5, we must determine the player's and goal's positions. As long as there are more than ten opposing players on one team, this period is used for the penalty kick shot and output.

3.5. Fire and Commemorate Detection. We are down to our last play of the football game. Fans are captivated by the game because they know that each shot they take has the potential to result in a goal. Celebrating also serves as a way to honor key occurrences in one's life. In order to complete this lesson, we must locate examples of goals and celebrations at live football games around the country. Despite the fact that filming and celebrating are both transient activities, we need a steady stream of frames to help us track them down. As an action recognition problem, we consider shooting and celebrating as a means of distinguishing the kind of human activity in a film. Prior to the widespread use of deep learning, IDT was the most stable and reliable technique for action detection, but it is also the slowest [34]. Matching the optical flow between two frames of video and SURF critical places reduces the impact of camera movement. As deep learning has developed, there have been several action recognition systems based on deep learning. For every two frames in a video sequence, a dense optical stream is generated and the CNN model is trained for both the video image and this dense optical stream. To arrive at the final classification result, two independent networks are fused together and the class scores of each network are
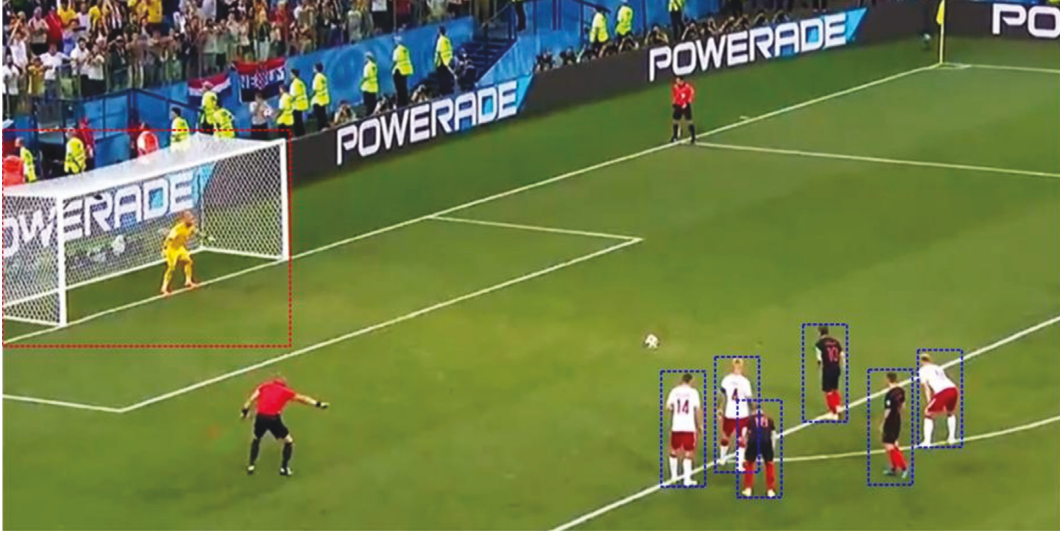
FIGURE 5: Free throw. The yellow and red boxes in the figure indicate how this module uses a frame to depict the players and the goal.

directly compared. In order to process video, C3D employs 3D convolution and 3D pooling. Pre-trained on the kinetic dataset, I3D has a better architecture and can be used for a wider range of tasks. P3D conducts a research study to highlight the relevance of time domain information in action recognition. Figure 6 shows the results of using the well-known action detection model I3D to identify shots and celebrations. Data augmentation is necessary due to the game's restricted number of shots and celebrations. We utilize clips with gunshots or rejoicing as examples of good examples. I3D will be trained using enough data from positive samples that were offset back and forth in time at random. After receiving a video feed of some kind, a module creates a point in the video that was identified as either a firing or celebratory event in the video.

3.6. Detection of Scores.

The most thrilling aspect of a game is the goal, which is one of the most significant components of highlights. By keeping an eye on how the score changes, we develop a scoring system that can identify goals. The three parts of this system are text area identification, text recognition, and postprocessing. CTPN is used to recognize text regions. To recognize text areas in each frame of football match footage, we train our model on a vast number of text pictures. However, in football match footage, the score occupies relatively little space, resulting in erratic outcomes. As a result, we have devised a straightforward method for reducing the size of the gap between the scoring area and the input frame when using CTPN. To get more precise findings, we re-enter the text sections containing the scores that we found into CTPN. Boxes $(B_1(c), B_2(c), \ldots B_N(c))$ are the text areas found in this frame, where N_c is the number of boxes. As long as the intersection over union (IoU) between them is greater than a threshold, we consider these regions in subsequent frames as possible score regions. According to several definitions, the IoU between two locations is as follows:
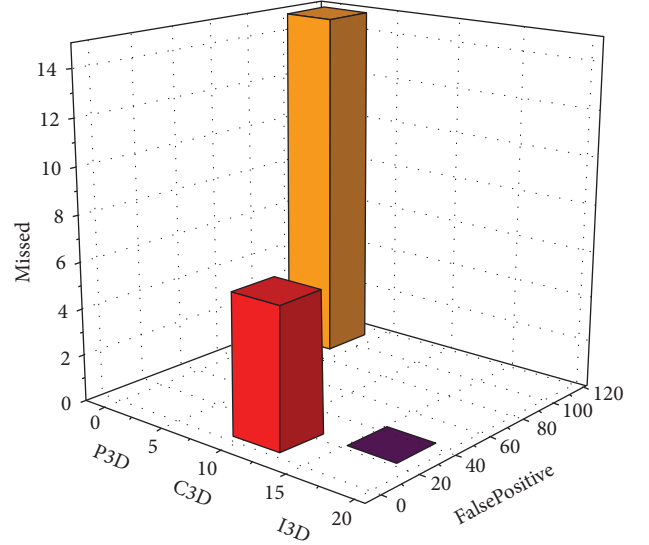

FIGURE 6: Shoot detection and celebration.

$$\text{Iou}_{i,j}(c) = \frac{B_i(c) \cap B_j(c-1)}{B_i(c) \cup B_j(c-1)}, \quad c > 1. \qquad (2)$$

If you look at the cth frame at the ith box, you will notice that there is an IoU between the ith and the jth boxes (c). Frame c will then be expanded to include more potential sites. The length and breadth of our system have both grown by 150 pixels. As a last step, we run the potential regions through CTPN to create scores for each of them. In order to recognize text, we use OCR technology to identify words inside text fields. We use Tesseract since it is an open source and mature. The result of this component is a list of all the words it has found. It is nice to have a long list of words to work with. Postprocessing is required to retrieve the scores from this sequence. Regular expressions are the perfect tool for this. It is capable of finding words in a string that conform to the score's structure. We note the score and

verify the time range when the score goes from 0 colon 0 to 1 colon 0.

3.7. Recognition of a Person's Face. For a long period, networks with a pure cascade design led the WIDER FACE challenge. The MTCNN detector, which conducts both face detection and face landmarking, was one of the most commonly utilized. The network is divided into three subnetworks: P-Net, R-Net, and O-Net. The first step does coarse face detection, resulting in proposal zones. The non-maximum suppression method next minimizes the amount of overlapping boxed areas, resulting in more specific regions that are supplied to R-Net. R-Net refines the chosen suggestions, while O-Net does face landmarking. MTCNN is still being recommended for use in the cutting-edge facial recognition system described in. Here, we want to use face-recognition technology to identify soccer stars in match recordings. Due to the obscurity of soccer match footage, this is a challenging task but it can be done. The bulk of facial recognition training datasets uses just photos of the front of people's faces. In the end, performance is harmed because of the data bias. It is not necessary to finetune the ArcFace face-recognition model since the current model performs well. We used the MTCNN face detection model, which is a quick and effective model. Prior to recognizing a face, a picture needs to be registered. Since registration pictures are often current front-facing shots of the subject to be recognized, we are unable to collect recent front-facing photos of soccer stars. In our opinion, soccer match films benefit more from the use of photos than web images do. As a result, we recommend a two-step approach to spot soccer stars in match films. The first step is to find soccer stars in recent match videos by utilizing web photographs as registration images. For registration photos, we search the Internet for front-facing photographs of these superstars. Every five frames in a match video, a face is detected and preprocessed in the same way as a picture is registered on a discovered face area. After preprocessing, we send the identified facial regions to ArcFace for face recognition. In order to get as many star face shots as possible in match records, we have lowered the bar. Second, we personally look for a clean front-face photograph of each celebrity. Each star has a total of around 15 official photographs. During the trial period, the previously prepared registration photographs are utilized. What the face-recognition module produces is the frame number and location of each star's face.

3.8. Integration Module. The positions of key frames of highlights and shot boundary frames are output by the modules mentioned above. We make a short film that lasts around a minute by combining a photo containing a critical frame with neighboring shots. The highlights of our system are these bundled short videos as shown in Figure 7.

4. System Details

Each module needs its own speed settings to ensure that the whole system is real time during the testing period. The shot
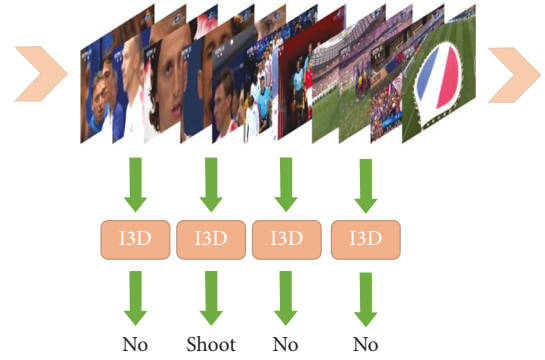

FIGURE 7: Detection of both the shooting and the celebration. The live feed is broken up into a series of short movies to forecast whether there will be a shooting or a celebration.
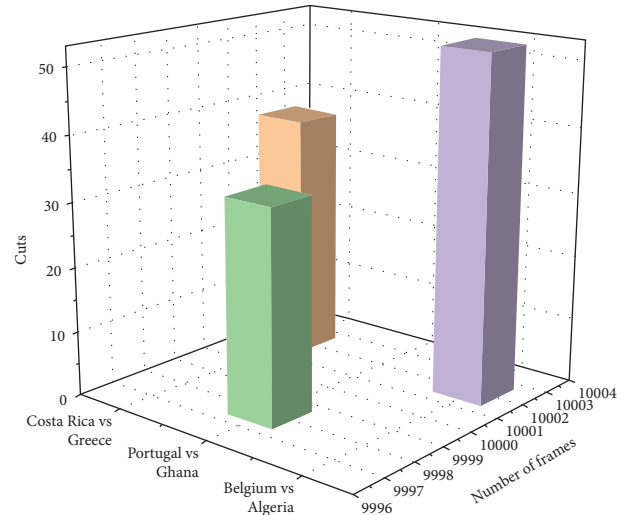

FIGURE 8: The data statistics used in the experiments.

segmentation module is used in every frame. The input stream is evaluated every eight frames for red and yellow cards, as well as corner kick detection. A 2.56-second movie is created by selecting one frame from the live stream for each eight that are shown throughout the detection process. In order to import these movies, I3D will be used. Detection of scores, faces, and identification are carried out every 25 frames.

5. Experiments

5.1. Aspect Ratio Calculation. Shot segmentation was tested in three 2014 World Cup games: Costa Rica vs. Greece, Portugal vs Ghana, and Belgium vs. Algeria. The first 10,000 frames of every 720p video are included in every movie. Figure 8 displays the statistical analysis of the test results. As demonstrated in Figure 9, our method outperforms the competition in terms of accuracy and stability. Figure 8 shows the number of applications in each class as well as the application lists that have been chosen.
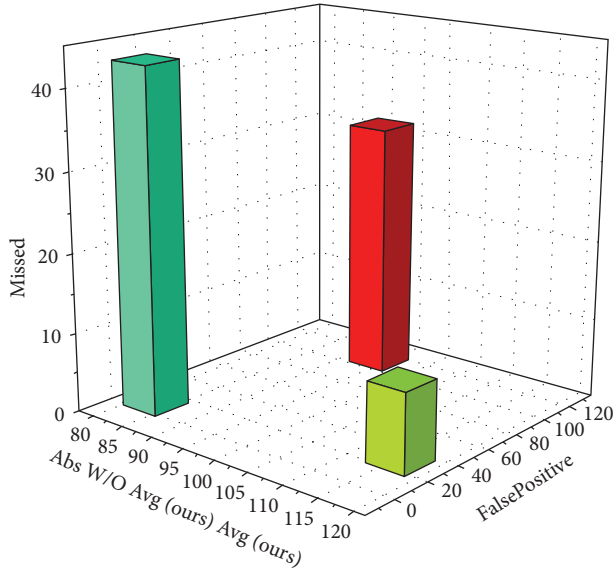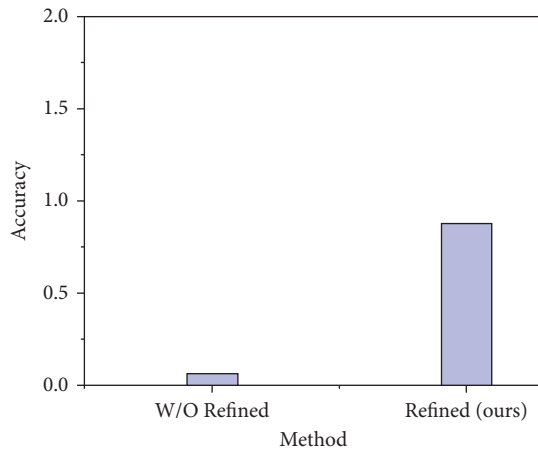
FIGURE 9: Shot segmentation results.
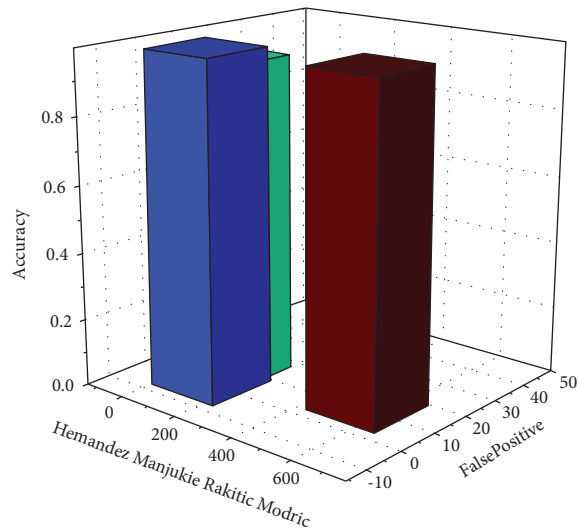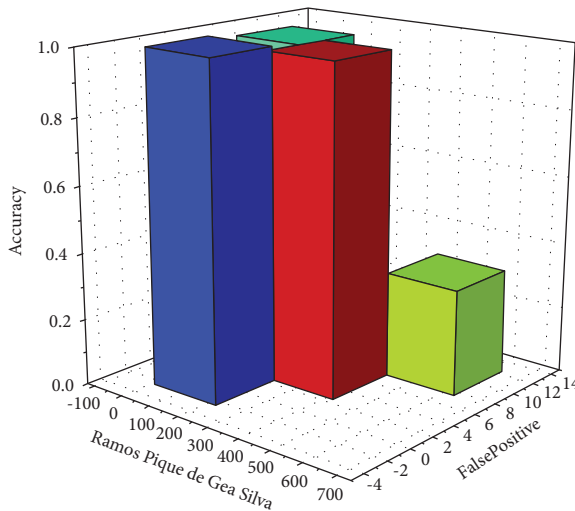

FIGURE 10: The score detection results.

5.2. Detection of Shooting and Celebration. In a football match video, we examined the C3D, P3D, or I3D renderings to see which one was the best. Based on a movie, we extract 1000 segments, each of which is 2.56 seconds in length and consists of 16 frames. To put it another way, we choose one frame from every four that are available. In addition to the 1,000 footage, there are 13 gunshots and five celebrations. The outcomes are shown in Figure 6. There will be no further attention given to outcomes with less than 0.5 confidences. A greater percentage of false positives than with C3D have been seen using I3D's test samples, but the experimental data show that it is capable of identifying all essential actions. We were more concerned with the model's ability to distinguish between the photography and the celebration, since we wanted to see whether the model was capable of doing so.

5.3. Finding Scores. A brief match film was used to compare the results of CTPN with refined CTPN. Figure 10 displays the results. It is the ratio of properly recognized scores to total occurrences that determines a system's accuracy. In comparison to the previous technique, our new approach is far more accurate.

5.4. Face Detection. Video footage from two World Cup matches in Brazil is used to assess the facial recognition module's effectiveness. Figure 11 displays the results. Our method has a recognition accuracy of more than 98 percent for most stars. Even in exceedingly intricate landscapes, as seen in Figure 12, these stars may be reliably identified.

5.5. System as a Whole. During the 2018 World Cup, our method was used. Highlights are brief movies that cover things like shooting, red and yellow cards, corner kicks, penalty kicks, goals, and celebrations. During the game, our system obtained 100% recall, but not with great precision. Before uploading small movies to the network, manual filtering is necessary. Therefore, recall is more critical than perfection.

FIGURE 11: Face-recognition research projects.

FIGURE 12: In the 2018 FIFA World Cup, face detection and recognition will be used. Here are some instances of how to correctly identify and recognize objects in difficult situations, such as those including side faces, occlusion, and blur.

6. Conclusion

In this research, we provide a unique technique for automatically extracting highlights from soccer films or match replays. This system is made up of deep learning-based action recognition, object identification, and facial recognition components. To extract shot boundary frames, we initially used color histogram characteristics. The keyframes of the highlights and star shots are then found using deep learning algorithms. Finally, by merging pictures using keyframes, stunning short films and star clips are created. The experimental findings reveal that the different components of this system function well. During the 2018 FIFA World Cup, our solution was also implemented onto live streaming platforms and it functioned admirably.

Data Availability

All data, models, and code generated or used during the study appear in the submitted article.

Conflicts of Interest

The authors declare that they have no conflicts of interest and they do not have any commercial or associative interest in connection with the work submitted.

Acknowledgments

References

[1] M. Stoeve, D. Schuldhaus, A. Gamp, C. Zwick, and B. M. Eskofier, "From the laboratory to the field: IMU-based shot and pass detection in football training and game scenarios using deep learning," *Sensors*, vol. 21, no. 9, p. 3071, 2021.

[2] C. Merhej, R. J. Beal, and T. Matthews, "What happened next? Using deep learning to value defensive actions in football event-data," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 3394–3403, New York, NY, USA, August 2021.

[3] B. Wang, W. Shen, and F. S. Chen, "Football match intelligent editing system based on deep learning," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 13, no. 10, pp. 5130–5143, 2019.

[4] M. R. Begum, M. R. NarasingaRao, and N. Polala, "Deep learning based Football player's Health Analysis," *Journal For Innovative Development in Pharmaceutical and Technical Science (JIDPTS)*, vol. 4, no. 8, 2021.

[5] D. Tang, "Hybridized hierarchical deep convolutional neural network for sports rehabilitation exercises," *IEEE Access*, vol. 8, pp. 118969–118977, 2020.

[6] H. Li, "Analysis on the construction of sports match prediction model using neural network," *Soft Computing*, vol. 24, no. 11, pp. 8343–8353, 2020.

[7] N. Abdelhade, T. H. A. Soliman, and H. M. Ibrahim, "Detecting twitter users' opinions of Arabic comments during various time episodes via deep neural network," in *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics*, pp. 232–246, Cairo, Egypt, October 2017.

[8] A. Tejero-de-Pablos, Y. Nakashima, T. Sato, N. Yokoya, M. Linna, and E Rahtu, "Summarization of user-generated sports video by using deep action recognition features," *IEEE Transactions on Multimedia*, vol. 20, no. 8, pp. 2000–2011, 2018.

[9] H. Ma and X. Pang, "Research and analysis of sport medical data processing algorithms based on deep learning and Internet of things," *IEEE Access*, vol. 7, pp. 118839–118849, 2019.

[10] X. Ning, W. Tian, W. Li et al., "BDARS_CapsNet: Bi-directional attention routing sausage capsule network," *IEEE Access*, vol. 8, pp. 59059–59068, 2020.

[11] A. Ekin and M. Tekalp, "Generic play-break event detection for summarization and hierarchical sports video analysis," in *Proceedings of the 2003 international conference on multimedia and Expo. ICME'03*, pp. 1–169, Baltimore, MD, USA, July 2003.

[12] K. Tirdad, A. Dela Cruz, A. Sadeghian, and M. Cusimano, "A deep neural network approach for sentiment analysis of medically related texts: an analysis of tweets related to concussions in sports," *Brain informatics*, vol. 8, no. 1, pp. 12–17, 2021.

[13] H. Pan, B. Li, and M. I. Sezan, "Automatic detection of replay segments in broadcast sports programs by detection of logos in scene transitions," in *Proceedings of the 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 3385–3388, Orlando, USA, 2002.

[14] X. Du and W. Cai, "Simulating a basketball game with hdp-based models and forecasting the outcome," in *Proceedings of the 2018 7th International Conference on Digital Home (ICDH)*, pp. 193–199, Guilin, China, December 2018.

[15] N. Ancona, G. Cicirelli, and A. Branca, "Goal detection in football by using support vector machines for classification," in *Proceedings of the International Joint Conference on Neural Networks*, pp. 611–616, Washington, DC, USA, July 2001.

[16] H. M. Zawbaa, N. El-Bendary, and A. E. Hassanien, "Event detection based approach for soccer video summarization using machine learning," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 7, no. 2, pp. 63–80, 2012.

[17] E. Fendri, H. Ben-Abdallah, and A. B. Hamadou, "A novel approach for soccer video summarization," in *Proceedings of the 2010 second international conference on multimedia and information technology*, pp. 138–141, IEEE, Kaifeng, China, April 2010.

[18] A. Ekin, A. M. Tekalp, and R. Mehrotra, "Automatic soccer video analysis and summarization," *IEEE Transactions on Image Processing*, vol. 12, no. 7, pp. 796–807, 2003.

[19] E. Lotfi and H. R. Pourreza, "Event detection and automatic summarization in soccer video," in *Proceedings of the 4th Iranian Conference on Machine Vision and Image Processing (MVIP07)*, Mashhad, Iran, 2007.

[20] Y. Tabii and R. O. Thami, "A new method for soccer video summarizing based on shot detection, classification and finite state machine," in *Proceedings of the 5th international conference SETIT*, March 2009.

[21] R. Yan, J. Tang, and X. Shu, "Participation-contributed temporal dynamic model for group activity recognition," in *Proceedings of the 26th ACM international conference on Multimedia*, pp. 1292–1300, New York, NY, USA, October 2018.

[22] P. Chen, Q. Xiao, and J. Xu, "Facial attribute editing using semantic segmentation," in *Proceedings of the 2019 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS)*, pp. 97–103, IEEE, Shenzhen, China, May 2019.

[23] P. Chen, Q. Xiao, and J. Xu, "Harnessing semantic segmentation masks for accurate facial attribute editing," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 12, Article ID e5798, 2020.

[24] X. Ning, D. Gou, and X. Dong, "Conditional generative adversarial networks based on the principle of homologycontinuity for face aging," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 12, Article ID e5792, 2020.

[25] S. Li, L. Sun, and X. Ning, "Head pose classification based on line portrait," in *Proceedings of the 2019 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS)*, pp. 186–189, Shenzhen, China, May 2019.

[26] X. Ning, W. Li, B. Tang, and H. He, "BULDP: biomimetic uncorrelated locality discriminant projection for feature extraction in face recognition," *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2575–2586, 2018.

[27] W. Cai, B. Zhai, Y. Liu, R. Liu, and X. Ning, "Quadratic polynomial guided fuzzy C-means and dual attention mechanism for medical image segmentation," *Displays*, vol. 70, Article ID 102106, 2021.

[28] W. Cai, D. Liu, X. Ning, C. Wang, and G. Xie, "Voxel-based three-view hybrid parallel network for 3D object classification," *Displays*, vol. 69, no. 1, Article ID 102076, 2021.

[29] Z. Chen, J. Huang, and H. Ahn, "Costly features classification using Monte Carlo tree search," in *Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, Shenzhen, China, June 2021.

[30] H. Ueda, T. Miyatake, and S. Yoshizawa, "IMPACT: an interactive natural-motion-picture dedicated multimedia authoring system," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 343–350, Massachusetts, USA, April 1991.

[31] R. Girshick, J. Donahue, and T. Darrell, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587, 2014.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.

[33] X. Du, S. Tang, Z. Lu, K. Gai, J. Wu, and P. C. K. Hung, "Scientific workflows in IoT environments: a data placement strategy based on heterogeneous edge-cloud computing," *ACM Transactions on Management Information Systems*, 2022.

[34] H. Wang and C. Schmid, "Action recognition with improved trajectories," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3551–3558, Sydney, NSW, Australia, December 2013.