Journal of Autonomous Intelligence (2023) Volume 6 Issue 1
doi:10.32629/jai.v6i1.577

ORIGINAL RESEARCH ARTICLE

Transfer learning model for the motion detection of sports players

Wael Y. Alghamdi

Department of Computer Science, College of Computers and Information Technology, Taif University, P. O. Box 11099, Taif 21944, Saudi Arabia. E-mail: W.Alghamdi@tu.edu.sa

ABSTRACT

Recognizing and analyzing moving targets is an important research subject since computer vision is employed in so many facets of our daily lives, including intelligent robotics, video surveillance, medical education, sporting events, and the maintenance of our national defense. This is because it may be difficult to properly analyse and keep up with moving materials. The various training postures of an athlete are explored in this study through the examination of a weightlifting video. This article was written to assist coaches in their efforts to improve the performance of their athletes in their respective sports. A technique for extracting essential poses from sports films has been proposed. The classification of different subjects of interest serves as the foundation for this technique. Because of its inadequate edge detection method, the current motion identification system does a bad job of detecting athletes, which is one of the reasons why it does a poor job of identifying motion in general. This flaw is one of the reasons why the system isn't very strong at detecting athletes. The following was one of the factors that contributed to this outcome: in truth, the situation is currently in this state. The result of the newly developed system outperforms the prior system in terms of tracking recognition accuracy and convergence speed. The system was put to the test. The findings of the system's study served as the foundation for this decision. Finally, the findings of the categorization reveal that the selection approach tries to separate fundamental postures.

Keywords: Deep Learning; Motion Detection; Sportsman; Convolutional Neural Network; Segmentation

ARTICLE INFO

Received: 30 March 2023
Accepted: 24 April 2023
Available online: 13 June 2023

1. Introduction

In conventional text-based query systems, the query phrases may be used to express the question's intended purpose[1]. The reason for this is that query intent is not always accurately represented by or matched to the underlying features[2]. Traditional text-based query methods cannot always fully match the intent of a question, resulting in irrelevant or incomplete responses. Among the many such constraints are: users' textual inquiries frequently include several possible interpretations of words, typos, and a lack of context, all of which can lead to useless results in search engines. These constraints, however, can be circumvented by utilizing natural language processing (NLP) and machine learning algorithms to acquire a better understanding of user intent. Search engines can increase the performance of text-based query systems through semantic search, autocorrect and recommendation technologies, customization tools, and voice search technology. In contrast to autocorrect and suggestion systems, semantic search assists search engines in understanding the intent of a question rather than merely the words entered into it. The greater accuracy and relevance of search results made available by voice search technology are truly astounding when combined with customization. Personalization allows you to customise outcomes for

specific users. However, human actions captured on camera in sports are exceedingly complex and demand a high degree of competence, making the assessment of sports footage substantially more difficult and time-consuming than that of conventional sports[3]. Because of this, the assessment of sports records may not only improve the viewing experience of athletic events but also help coaches and teammates evaluate contests and aid athletes in their preparation. The issues of analyzing complex sports videos, deep learning and computer vision-based approaches have been created. They have become an important tool in target identification and may be employed in a variety of settings. Picking a region, extracting characteristics from that region, and naming those features are examples of these procedures. Color and depth maps, as well as still photos, can be analysed to show concealed motion. Pattern recognition techniques combined with image-based recognition technologies may help identify a person's posture, which is also very important. Signal capture solutions may also be able to detect motion. In conclusion, DL-based techniques offer some viable solutions for evaluating sports videos. Unlike human monitors, computers never get tired, and they never overlook anything crucial while keeping tabs on anything. Thus, the computer may improve productivity while simultaneously cutting costs by reducing the need for human labour and other resources. This technique is very crucial in the field of medicine. Using ward surveillance as an example, unusual ward behaviour might be flagged to doctors instantly. This not only reduces the cost of medical treatment but also decreases the likelihood that a patient may experience an unanticipated event. Deep learning-based technologies that employ computer vision to differentiate moving objects may be used to enhance patient outcomes and save costs in medicine through the use of computer surveillance. Computerised surveillance may do this by detecting moving objects. These approaches may identify complex motion patterns, which aid in posture and movement analysis and enable continuous patient monitoring. This technology allows doctors to track their patients' movements and health in real time, which can help with diagnosis and treatment. Furthermore, computer vision might be used to determine if patients are taking their drugs as recommended by reviewing video recordings. As a result, patients may be more likely to adhere to their treatment programmes and experience fewer adverse medication responses. Furthermore, computerised monitoring can assist in identifying people who are at risk for undesirable outcomes such as falls, pressure ulcers, and other accidents. The ability to take such safeguards enables doctors to minimise total healthcare costs. In terms of patient monitoring and improved healthcare results, computer surveillance may be a cost-effective and efficient strategy. In entertainment video retrieval, the query condition may be a series of images, or a textual description of the information being searched and then the results are compared to see whether they are similar. Having a high degree of adaptability and transparency, it foreshadows future directions in intelligent image processing[4]. The benefits of artificial intelligence-based computer vision problems[5] have become more apparent as new technologies such as clever license plate recognition and genuine 3D effects replaying of sports games have emerged. The degree to which the athletes' movements are standardised throughout the weightlifting process has an immediate influence on their performance. An athlete's effectiveness in the weightlifting process depends on maintaining several crucial postures throughout. The importance of these starting positions in weightlifting cannot be overstated. There are a lot of tricky and inevitable interference variables to consider while capturing surveillance footage. Occlusion between objects, noise, the presence of other influences, and rapid shifts in lighting are all examples of such phenomena, which requires the algorithm to have an ever-increasing amount of processing power. This is because conventional algorithms were created at a time when there was far less video footage. Previous weightlifting films relied heavily on instructors' subjective judgments of athletes' form during training. Not only did this waste time and resources, but it also led to inaccurate key posture extraction that was very susceptible to bias. In this research, we use video analytic technologies to study weightlifting training and its effects on the body. This research takes on weightlifting head-on by analyzing weightlifting movement through the lens of video analysis technologies. In addition, the data given examines the potential of deep learning-based computer vision algorithms for motion analysis and posture iden-

tification. Though it does not directly answer the question of how video analysis might be used to investigate the effects of weightlifting training on the body, it does provide some insights into how video analysis can be used for motion analysis, which could be applied to weightlifting training as well. The findings imply that deep learning approaches might be used to analyse video footage in order to determine human motion patterns and posture. To accomplish this, regions must be identified, features extracted, and attributes classified. These operations may be carried out using 2D representations of the human body that are not affected by the observer's perspective. Furthermore, the evidence shows that integrating geographical and temporal aspects may improve one's ability to account for one's own behaviour. Video analysis might be used to evaluate the physiological effects of weightlifting based on these findings. Observing people lifting weights and examining their posture and movement patterns may aid in this objective. This may shed light on the effects of weightlifting on various regions of the body as well as the effectiveness of specialised weightlifting tactics for targeting specific muscle groups. Video analysis might also be utilized to track any changes in motion or posture caused by weightlifting training programmes, offering further information for the programmes' efficacy.

2. Related work

To extract the most fundamental movements from an input video stream, Lamas *et al*.[6] first identify the stream's edges, apply the edge attributes to the method of creating movement descriptions, and finally cluster the attributes calculated by the incoming live feed. The loop incremental model can thus be used to find values that are close to the true values of the parameter estimation. Support vector machines (SVM), which were talked about in the study of Nandyal and Kattimani[7], were used to solve the problem of recognising human motion. As discussed by Zhao *et al*.[8], spatial characteristics are extracted from fixed places of interest in video. Simultaneously, it uses moving nodes of interest to derive spatial and temporal characteristics. At last, a combined set of spatial characteristics and spatio-temporal features is obtained. When using a classifier machine for identification, this configuration is one option. Guo[9] proposed that: after the film was edge-detected, the resulting edge characteristics were used in the training process to extract the gesture characteristics. A vote was then held, and most of the group opted to approve the motion. According to Bruno *et al*.[10], a comprehensive and quarter model should be used, where the body model employs intensively collected contour environment descriptions and the preceding version is trained extensively in the databases utilizing number of co and multi-pose. In addition, the original model included shape context descriptors with a high density of samples. The movement, space modeling, and quality are utilized to determine which portions of the human body are visible. To overcome CNN's limitations on input picture size and improve accuracy, Xu[11] included an SVM classification pool layer in the network. Based on this, Liu *et al*.[12] presented a new, more efficient detection approach called Rapid as well as an innovative new technique called identifying the regularisation of regions of interest. This would bring the technological progression from RCNN to Faster-RCNN to a close. Recent scientific findings form the basis of the new approach, which explains the enhancements. The NN (nearest neighbour) classifier is utilized to evaluate the efficiency of the tracking procedure by Kalakoti and Prabakaran[13]. The neural network classifier can calculate the degree to which the most recent target image resembles the previously collected correct image. The top-down approach has certain limitations due to its characteristics, such as the inaccurate identification of objects in noisy environments and the misalignment of important spots in congested settings. The predicted two-dimensional posture will be unstable if the subject in the video is engaging in rapid, complicated movements, which in turn will distort the image. According to the study by Li-quan *et al*.[14], directing all pelvic joints is an effective and efficient way to organise nodes in a two-dimensional posture space.

It is critical for successful target extraction from an image under any circumstances to select a robust feature set and then apply it in such a way that objectives of multiple categories have high unequal treatment and can adapt to shifts among both benchmarks in the same class. Important factors in identifying targets include the target's shape, texture, and color. On the other hand, the target's shape is not affected by environmental variables like light,

so it is often regarded as an excellent frontrunner for target selection. When describing an object, designers must take its shape into account if they want to design an accurate feature set.

There are five distinct movements involved in lifting weights: straightening the knees to elevate the bell, guiding the knees to elevate the bell, applying force, bending, and maintaining, and standing up. There are several biomechanical factors that coaches and players alike are interested in observing. This is because estimating the poses of individual body parts is not possible using standard pose estimation methods. The synchronisation here between frames is likewise a major issue of concern. With constant motion, the delay between video frames is minuscule. RoI-KP is a fundamental posture extraction approach that we detail in this section for use with sports video. The goal of this tactic is to teach users how to organise their interests into distinct buckets shown in **Figure 1**. An approach that involves extracting major frames from the movie is one of the methods that can be employed to achieve the needed area. In addition, each video frame is retrieved and analysed using the CNN network to provide potentially significant frames; this is done to achieve the best results possible[15]. Users will go to this step after finishing the previous one. The technique for selecting important frames then makes use of this value as a point of reference. The focal region that is the subject of this inquiry has been partitioned to decrease the impact of context on the choice of important frames. This is one of our investigation's goals. Weightlifting movies, which frequently incorporate distracting background music or other visual features, inspired the development of this discipline. These films provided the impetus for the growth of this profession. If the land was first separated more precisely, the procedures outlined further down on this page could be used to subdivide it. Begin with any of the image's four corners and move toward the center. Begin by selecting the label value with the lowest value among the four closest places. Move clockwise from one corner to the next to do this. Repeat this procedure until all label values are the same. Until now, the area of interest has been marked on a new map, and every area that is neither zero nor discontinuous has been assigned a number. As a result, selecting the zone with the highest score as the focal point is all that is required. Using three dimensions to analyse human posture is more difficult than using two dimensions. This is mostly because 2D pose estimation has more training data than 3D pose estimation, allowing it to handle accuracy and occlusion concerns more effectively[16]. It is hard to predict the connected nodes in a three-dimensional space from images or videos because of regression, but this step is still important no matter what the problem is. It's widely used in game development, behavioural research, animation, and motion capture systems. It could be used in a variety of human-body tasks, such as whole-body analysis, as well as a supporting component in algorithms like the one used to identify pedestrians.

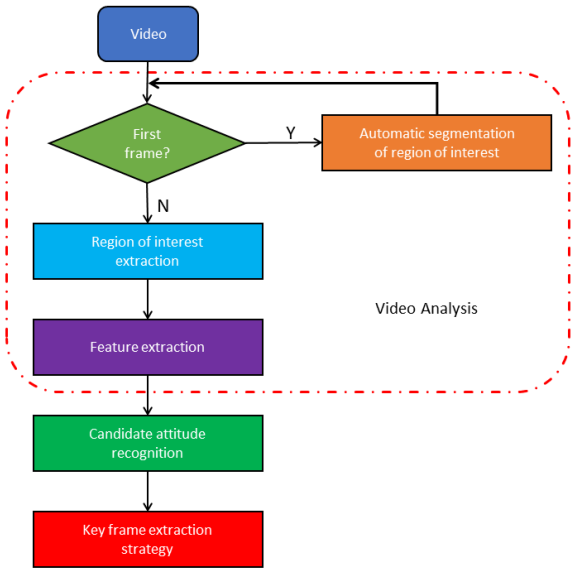

Figure 1. An approach for key frame extraction from video.

A simple batch normalising multi-layer deep neural network is used to construct each building component. If starting with the 2D vertices as the starting point, 3DPoseNet may be used to estimate the 3D coordinates of the vertices. We then give a visual depiction of each of these major characteristics. The first coordinate pinpoints the actual site of the 3D skeleton that is now visible, while the following coordinates are 3DPoseNet estimates for many critical skeleton places. One of them is deciding where to place the three-dimensional skeleton. The formula for the loss function established in this work can be found in the previously cited paper. The formula for the loss function is accessible here. In the previous study, the loss function with symmetric constraint was also utilized. In this case, the set consists of the beginning point, every point in the symmetrical segment, and any points

that are relatively close to one another. One of the limitations imposed by the human skeleton on the human body is the possible range of motion at each joint. The skeleton limits the human body in several ways, including this one. The problem of occlusion is investigated in this work, and a possible generic solution is proposed. The torso, together with the arms, legs, brain, and other anatomical components, is critical to the resolution of this problem. Because precise predictions of three-dimensional posture may be made from two-dimensional posture, recovered three-dimensional posture can be classified into four states. Taborri *et al*.[17] gave a thorough introduction to the idea of time series data, considering the latest developments in the fields that are relevant. The investigation may be much simplified by utilizing the previous instant's joint points to determine the joint location of the unexpected (occluded) segment, which is a considerable difference. For the sake of this discussion, we will take the coordinates of the projected joint point (occlusion node) at that moment as the location, a vector that depicts the goal that should be the focal point of all actions and events.

3. Research method

3.1 Taking steps towards DL-based motion recognition

Deep learning-based techniques that employ computer vision to recognise moving objects are gaining popularity as a viable alternative. This system was designed to detect complex movement patterns. In addition to its obvious applications in the military, police work, and other security-related professions, computer vision offers applications in the transportation, healthcare, and security industries. Because of the extensive usage of DL in computer vision and the ever-increasing processing power at our disposal, DL-based target identification has become an important tool. To apply AI-enhanced computer vision to improve entertainment video retrieval, solutions for signal collection, target identification, posture recognition, and object recognition are among them. These approaches have the potential to increase the precision and efficiency with which data from sensors and cameras is collected, as well as the analysis of moving objects and people. This has the ability to create more fascinating and immersive shows for audiences in the entertainment business. More precise and thorough information on the motion of the film can improve the viewing experience. This strategy was created in response to these two factors. Methods like region selection, feature extraction, and classic classifiers like the SVM model can be used to give an overview of the main steps in the process of identifying a target. To accomplish this, first, choose significant areas of photographs, then extract visual qualities, and then classify the images. By performing these three methods in the correct order, one can construct a summary. DL-based algorithms for object identification are employed in various industries, including transportation, healthcare, and security. DL has increased the performance of traditional approaches while simplifying them. There are two basic ways of identifying the properties of a target: the one-stage methodology and the two-stage strategy. Because of DL, it is now feasible to classify human motions and extract 3D coordinates from video using 2D alterations in posture. Both of these functions were previously inaccessible. Viewpoint-invariant feature extraction, viewpoint normalisation, and multi-perspective traversal are prominent ways to capture human motion that is independent of the viewpoint from which it is observed. The use of deep learning for the identification of human posture has made major contributions to the progress of pattern recognition. Sensors are positioned in a certain way to learn about human posture from the way people walk, and DL-based algorithms are also used to gather data for human posture identification.

The detection performance of traditional approaches has increased because of the development and use of DL, which has also reduced their complexity. In this kind of target identification, there are two primary schools of thought, both of which rely on region extraction. Both target identification and region extraction employ both one-stage and two-stage procedures. For feature extraction and picture labelling, computer vision and DL-based algorithms are frequently employed. CNNs trained on depth maps can distinguish between the foreground and background of a movie, whereas RGB maps properly portray the colors and textures of the human body and its surroundings. Because of its learning capabilities, RNN is effective for recognising time-dependent features of human motion

in videos. Pattern recognition in combination with image-based identification methods can be used to determine a human's position. Finally, there are signal-gathering devices that take limb movement into consideration when determining a person's posture. These are the one-stage approach and the two-stage strategy, respectively. Extraction of 3D coordinates from video is a difficult task. 2D changes might be used in posture to directly represent human motions. This is the alternative in this case. Every two-dimensional depiction of the human body may take on several different forms depending on the viewer's perspective. To achieve the objective of capturing human motion in a way that is invariant to the viewpoint from which it is viewed, viewpoint invariant feature extraction, viewpoint normalisation, and multi-perspective traversal are often used. Without the camera moving in tandem with the subject, the latter will quickly get confused. For example, while photographing a diver, the camera will descend with them. We provide a data-driven approach for identifying activities across many modalities as shown in **Figure 2**.

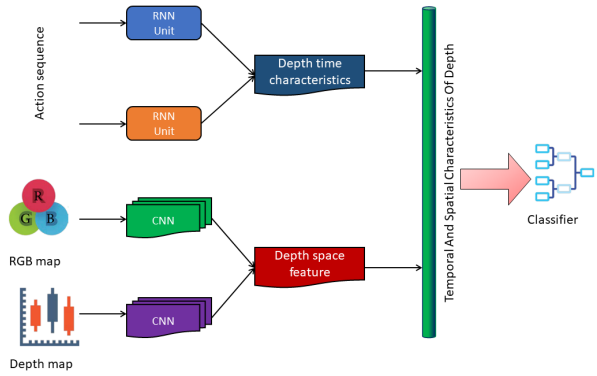

Figure 2. Extraction of motion using deep learning.

This research makes use of static data, such as color and depth maps. Remove unwanted elements from the backdrop, for example, using the information provided by the RGB map, which contains data about the environment and the body. Using the DNN learning paradigm, we design several network designs for use in different settings. RGB maps are superior at accurately depicting the colors and textures of the human body and background in a video; CNNs trained on depth maps are superior at accurately differentiating between the foreground and background scenes in a video, avoiding confusion over feature interpretation due to background interference. Additionally, the film accurately reflects the human body's and the environment's natural colours and textures, thanks to CNN's RGB map. Here, we use the Softmax loss function at the first node of the network. It is possible to individually get the characteristics of a depth map and an RGB map. CNNs trained on depth maps outperform CNNs trained on RGB maps in distinguishing foreground from background scenes in movies, according to the data. Using RGB maps, colors and textures in a movie, such as the human body and the backdrop, are more properly depicted. To filter away distracting components in the backdrop, the study uses static data such as color and depth maps. In a film, the depth map determines which items and places are in the foreground and which are in the background. This helps to remove any uncertainty that may occur in the interpretation of features as a consequence of background interference. RGB maps, on the other hand, may correctly depict the intrinsic colors and textures of the human body and its environment. Data-driven activity recognition across several media channels makes use of complementary spatial and temporal information. When used with image-based identification approaches, pattern recognition is most effective for assessing posture. One can learn a lot about human posture just by watching how people walk, and sensor placement can help with that.

The representational impact of this function may be said to be outstanding. Time-dependent aspects of human motion in video make it difficult to depict using just spatial depth information retrieved from movies. The networks used by standard DL methods for temporal data analysis all use a recurrent neural network (RNN) like approach to their architecture. When studying human movement, the amount of mobility of important bone sites tells us a lot about how much time has passed. The RNN network performs a series of studies on sequence data on a recurring basis, each time adjusting its analysis to a finer temporal resolution. RNN excels in modelling and extracting characteristics of sequence data when compared to other processing techniques. This method makes use of the cross-entropy loss function, which is written as "where is the current label and is the expected label". While the dynamic model can comprehend the action video's temporal aspects, the static model can comprehend its spatial ones. There are more differences and complementarities in multi-modal information features than

in single-modal ones. Therefore, it's possible that combining spatial and temporal features will give the data a stronger capacity to account for its own behavior.

3.2 Identifying posture

In recent years, the study of human posture has made remarkable gains, and one area that has made major strides is the science of pattern recognition[17]. Several researchers have also modified pattern recognition for the use of image-based recognition technologies in the field of wearable device-based human posture identification[18]. These researchers discovered that using pattern recognition in conjunction with image-based recognition technologies is particularly advantageous. This challenge was completed using pattern recognition. **Figure 3** displays the approach used to locate a person using inertial sensors. This methodology is divided into five major components: data collection, pre-processing, segmenting, extracting features, and training a classifier, in that order. The data is then processed for use in the system during the data preprocessing phase, for example, by drying and normalising the information[19]. Unknown samples were successfully classified, and using the collected samples, classification models based on several classification principles were created[20]. This happens at the final stage of the classification process, the classifier phase. The unit action analysis is performed during the feature extraction process, and the key attribute features are computed and extracted as sample data.

3.3 Signal acquisition solutions

Identification of human posture necessitates the ability to recognise how a person's arms and legs move in reaction to their surroundings. According to the findings, walking straight creates the greatest constant angular velocity in the legs. In this study, we focus on how people use their arms and legs to learn about human posture from how people move, see **Figure 4** for sensor arrangement this is because human movement is so intricate. After successfully collecting data, it is transferred to the base station through a wireless protocol[21,22]. After the data-gathering process is complete, the base station will use the serial connection to communicate the data to the host computer for further processing, which will allow for more efficient data utilization.

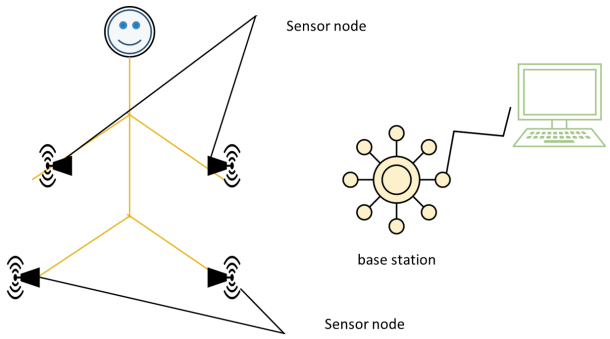

Figure 4. Sensors and base station model.

3.4 Basketball stance definition

Figure 5 depicts a breakdown and investigation of the construction of basketball stances, including how they are disassembled and assembled. Basketball is a physically demanding sport that always requires a wide variety of activities from its participants. The positioning of each of the basketball players' limbs determines whether it is in motion or static mode[23]. When a limb is "in motion", it is actively participating in basketball;

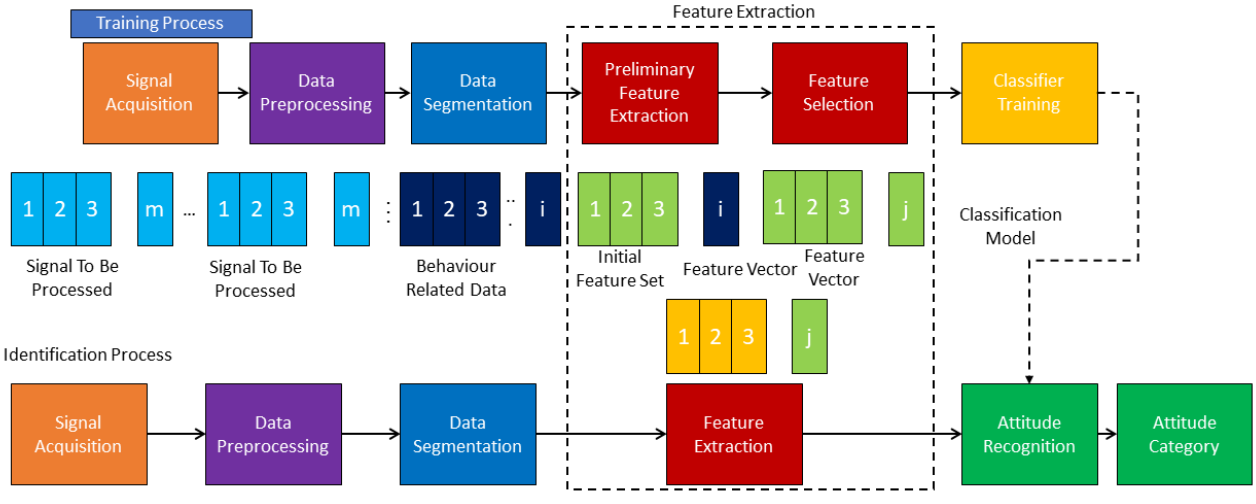

Figure 3. Flowchart of feature extraction and action recognition.

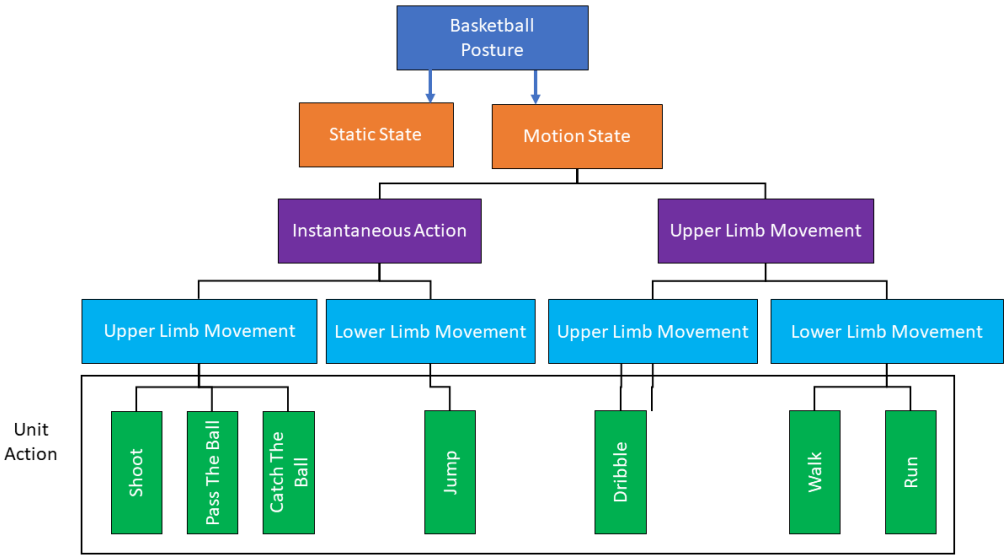

Figure 5. Sportsman movement details.

when it is "in stasis", its posture has not changed. Because the position of the athlete's leg does not change when the athlete catches the ball. However, when attempting to catch a ball, the arm used to do so enters a state of motion that could be described as dynamic. While continuous actions include shooting, catching, passing, and dribbling, transitory activities include jumping, strolling, and sprinting[24,25]. Upper-limb activities are considered transient, but lower-limb activities are considered continuous. For the purposes of this classification, it is critical to determine whether there are frequent transitions in the state of movement. Temporary activities, such as shooting and catching a basketball, are rare; instead, they consist of a single instance of the activity being performed. Over time, continuous exercise will repeat unit activities such as walking and dribbling in the same order. This will be the case due to the sequence of unit activities. As a result, to be recognised in the sport of basketball, it is critical to distinguish between upper-limb motions and lower-limb activities. This article will provide a way for job separation based on the division of certain processes, as well as an example of how it might be applied.

The angles made by the small arm and leg are shown along the vertical axes of subplots (1) and (2). The horizontal axis of this chart, which goes across the page from left to right, denotes the time range it covers (3). Looking at the first and third subplots, it is evident that the signal expressing angular velocity is not as clear as it could be. The angular signal curves in subplots (1) and (3), on the other hand, are smoother than those in the other subplots, implying that the angle-based unit action split may be easier to implement shown in **Figure 6**.

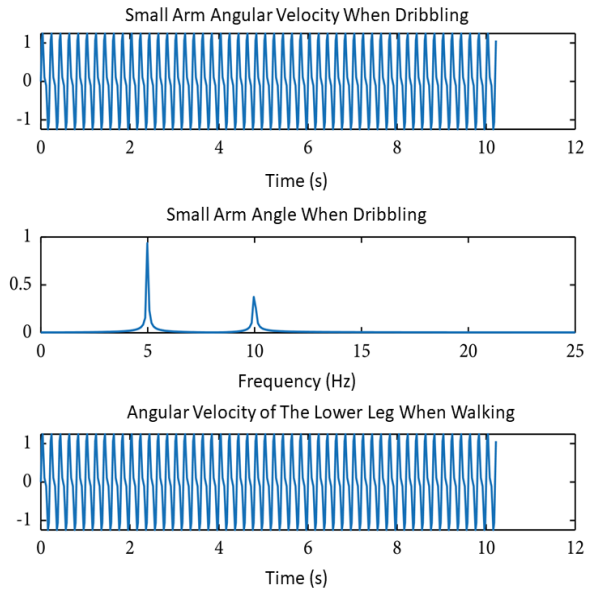

Figure 6. Arm and leg velocity graph.

Using computer vision, different athlete positions can be recognized; the major goal of the system is to track and monitor the numerous positions that athletes maintain throughout arena-based sporting activities[25]. This is the system's primary objective. The current mobility state of the gymnast must be determined before being analysed and assessed on a computer utilizing machine vision algorithm. As a result, it is an essential component of the system's architecture. Gymnast analysis necessitates the use of machine vision algorithms since the physiological motions of gymnasts, as well as their rate and direction

of movement, are unpredictable when competing. Machine vision algorithms are critical to the process of analyzing gymnast routines in competition and ranking them. This is due to the fact that these algorithms can identify and evaluate gymnasts' motions in real time with exceptional precision. In order to provide an appropriate judgement of the gymnast's performance, both the coaches and the judges must be familiar with the gymnast's movements. Picture capturing, image processing, feature extraction, and classification are all common components of machine vision algorithms. Furthermore, there might be the image acquisition subsystem, which may or may not contain cameras or other sensors, photographs the gymnast in action. The photos are then processed, which removes any distracting noise or undesired artefacts that might jeopardise the study's credibility. A "feature extraction" module will search through images of the subject's body to extract information about the subject's posture, kinematics, and location. Finally, the classification step evaluates the gymnast's performance by categorising her motions based on the elements retrieved (through machine learning). Coaches and judges may use machine vision algorithms to make reliable, unbiased decisions regarding gymnasts' performances. This results in better gymnast training and more fair competition judgement. This is the condition due to the nature of the sport. The wireless serial connection of the camera allows the user to regulate the focus and point it toward a moving object. The way how the system works is illustrated in **Figure 7**.

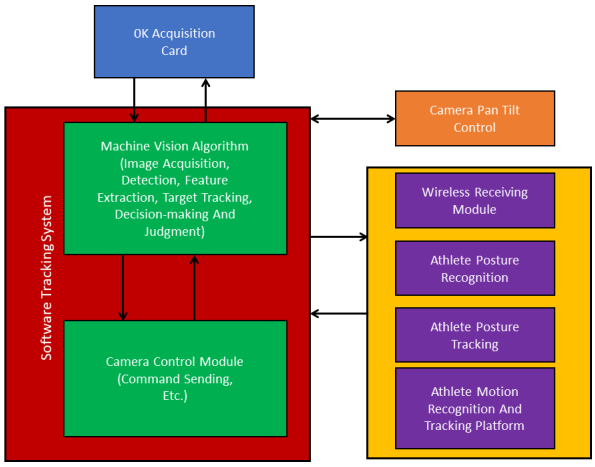

Figure 7. The major components of the posture recognition.

These components comprise hardware for collecting picture data and software for feature extraction, and segmentation[26–28]. These two components account for a significant amount of the system's total operation due to their respective contributions.

3.5 A better approach to object recognition and monitoring

Edge extraction may have an impact on the precision with which object detection and monitoring are performed. One of the most important stages in image processing and computer vision is "edge extraction", which entails recognising the boundaries between the numerous objects that occupy a picture. If the edges are extracted incorrectly, the precision of object detection and monitoring may degrade. Edge extraction, and hence object recognition and monitoring, may benefit from the application of migration learning and dataset accumulation. Migrating learning necessitates adjusting models that have previously been trained in one domain to be employed in another. By improving the precision of edge extraction and object detection, this can help decrease the requirement for vast volumes of labelled data. Massive datasets must be acquired and tagged in order to train algorithms for edge extraction and object identification. The performance of these models may be influenced by the quality and variety of the dataset. Having a significant amount of data, such as photographs taken in various lighting, camera angles, and object orientations, can enhance the precision of object recognition and monitoring. The more information we know about an object, the better we will be able to differentiate it and keep track of it. Overall, combining migration learning with dataset collection may increase the efficacy of object recognition and monitoring by improving edge extraction accuracy. To attain this purpose, the efficiency of object monitoring can be enhanced. Edge extraction loses some of its effectiveness when utilizing this technique, which affects how well it performs in terms of tracking recognition[29]. The histogram of oriented gradients (HOG) and the scale-invariant feature transform (SIFT) are two feature extraction methodologies used in pose-tracking recognition systems[30]. Both HOG and SIFT are acronyms for "Histogram of Oriented Gradients", but HOG is used to recover information about the geometry of things in a picture,

while SIFT is used to recover information about the texture and edges of objects. HOG and SIFT are two algorithms used in pose-tracking recognition systems. By collecting data from an image or video, these algorithms can recognise and follow the motions of a person's posture or body. These characteristics are used to create a model that can track a person's motions even if the camera's viewpoint changes. With the development and widespread acceptance of deep learning (DL) technology, the complexity of HOG and SIFT, two conventional approaches for pose-tracking recognition, has been considerably decreased. As a result, DL-based approaches are gaining popularity as viable substitutes. This was accomplished using migration learning and the acquisition and calibration of a dataset of the athlete's critical game-relevant variables[31,32]. To attain this purpose, preliminary training in the migration technique was required. The algorithmic strategy covered in this article is illustrated in **Figure 8**.

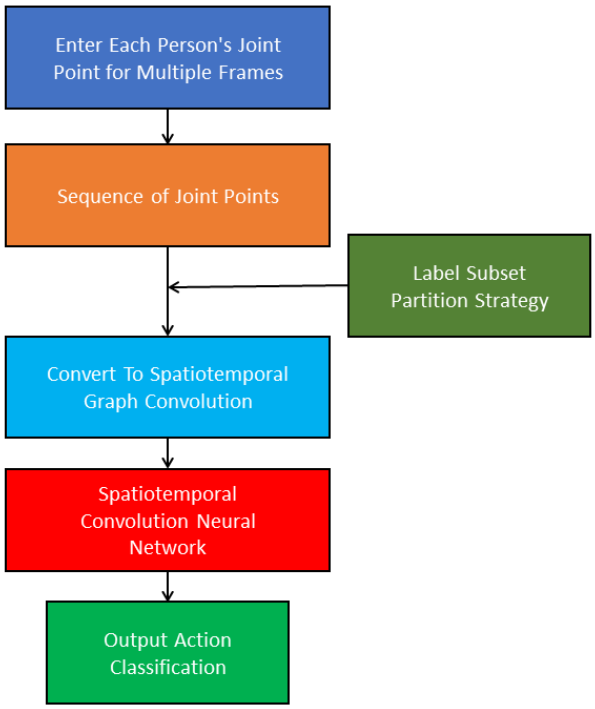

Figure 8. Sportsman action classification steps.

The previously stated spatiotemporal graph convolution technique is used to segment the athlete's body motion sequences acquired by the hardware. These sequences were identified by the hardware. Furthermore, the label subsets are used to categorise the joint points and the link relationships that exist between each of them. This is accomplished using the information contained in the labels. It can be used as a starting point to model the athlete's joints and limbs, add a temporal component to these models, and describe the movement as a series of postures[33]. The data structure provides a solid foundation for the joint matching algorithm to operate and produce the desired results. There are numerous methods for mapping labels to a specific convolution of a spatiotemporal network, such as utilizing a uniform division, a distance division, or a spatial division. Labels can also be mapped in relation to other convolutions in the network. These are only a few of the numerous techniques that could be used. To complete the classification of athlete posture, the information related to the matching features must be mapped. The output of the stacking module is one alternative source for acquiring this data. Posttracking will continue until the information dimension changes, after which it will be terminated. It has been determined that the design stage of the system that will recognise gymnastic stances may be completed effectively using machine vision.

4. Results analysis

In both image space and time series, human action behavior can be distinguished, mining spatial attributes are essential for picture identification and detection, while time is given more weight in videos. Because of this, a video-recognition algorithm must carefully examine not just the video's geographical and temporal details but also the human subjects themselves. Multiple still images are combined to create a moving image, or "frame", in a video clip. A large amount of time will be needed to process the video if all the recording tilts are used for it. Concurrently, the power of recognition will be lessened because not all pictures are linked to one another. To make these algorithms work better, it is important to find the most distinct spatial and temporal parts of a video. It's also important to remember that if the video's background color matches the woman's skin tone, it may be difficult to distinguish between her actions and attitude. There is an issue here that must be addressed. This topic must be carefully reviewed before taking any further action. The situation is unlikely to remain the same throughout the film, so spectators should brace themselves for an unpleasant encounter. **Figure 9** depicts four distinct types of probability evo-

lution curves.

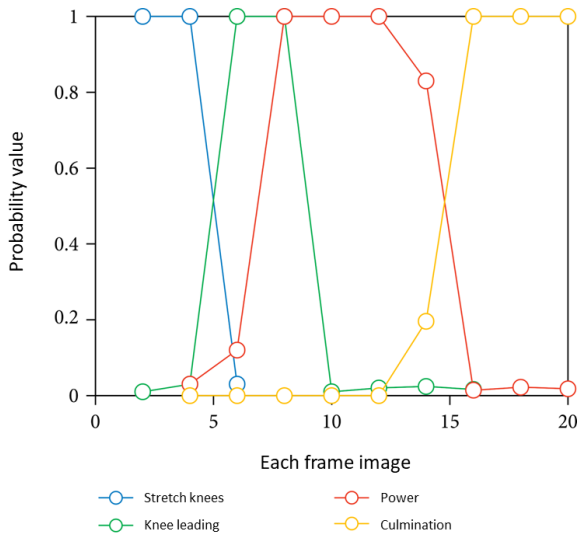

Figure 9. Probability value of each frame image.

The recommended ROI-KP technique in this research has led to a noticeable increase in efficiency. It can't be denied. When standard CNN classification algorithms are compared to cutting-edge ones, the results are less consistent and more unexpected. Because modern filmmaking methods have risen in popularity, it may be difficult for video data to maintain the original camera's position in each shot. Some ways include photographing the scene from above, photographing it from the perspective of someone who is in the location, and creating a completely invented persona. This means that capturing the same video fight sequence from different perspectives may result in vastly different feature representations. This means that deciding which camera angles to use is another critical problem that must be settled. The description of the qualities includes all the information about the location, as well as some information about the time because the research input is intended to be a continuously and meticulously recorded video meal. Because there is a limit to how much depth map data can be used for pre-training, the network will need to be totally retrained. **Figure 10** shows a side-by-side comparison of data from two independent video sets.

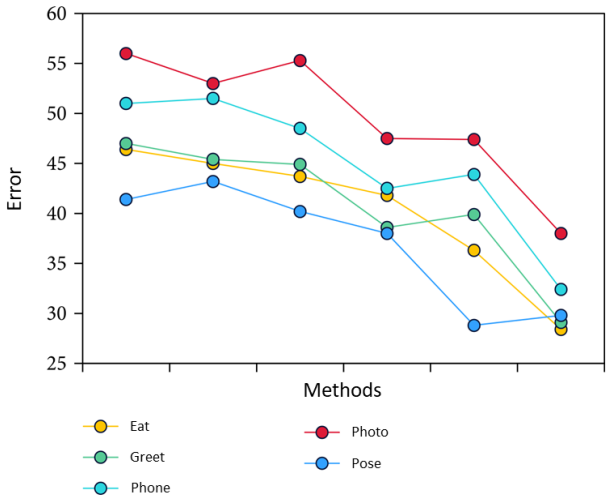

Figure 10. Error analysis on various parameters.

Because of the interaction of these two variables, the probability of any result other than first place is relatively large. The fourth keyframe performs exceptionally well under the normal CNN algorithm due to its high probability value and proximity to the key frame preceding it. As a result of these two variables, something has transpired because the features provided by the processing step are adequate for the training phase, RNNs can skip the entire pre-training phase. As a result, it is only used to store the most critical information. Researchers may still be able to obtain useful training results by utilizing knowledge transfer. Incorrect labelling or distribution that differs from the training set are only two examples of potential issues with the new data set. Even though the new data set is somewhat modest, this is because learning transfer allows for the diffusion of information across a wide range of data types. The rationale behind this is given below. The performance of the 3DPoseNet network is tested using the publicly accessible Human3.6M dataset, while the other two datasets serve as test subjects. As a result, the comparisons of the three datasets made by the authors are more reliable. Human3.6M is now the most popular public 3D posture data set because it has a lot of information about how people stand. It includes 11 different topics with 3.6 million video frames, seven of which employ 3D annotation postures.

It shows that the average error in the results generated by the 3DPoseNet model is lower than that produced by any other methodology and that these results are independent of the data and the operational procedures used to obtain them. The strategy's efficacy is shown by a reduction of at

least 6 mm in the average MPJPE error throughout the 3DPoseNet network. If the human posture differs considerably between the two time periods under discussion, there is a much lower connection between the observer's viewpoint position and the separation of the two descriptor variables under consideration. This is because the variable separation and the observer's position are linked. While a human is engaged in an activity, any two descriptive features that describe a human posture are close together. This is true even if the position is held for two distinct periods of time. Mostly because it takes a lot of effort to maintain proper posture. Someone's look can be used to glean information about their posture and demeanour by observing key traits. When data is analyzed using matrices that share properties, similar patterns emerge in the data, as demonstrated in **Figure 10**. These patterns can be discovered. This appears to show that self-similar matrices may endure changes in athletic performance. A comparison of the self-similar matrices created by various qualities demonstrates that the self-similar matrix is sensitive to the underlying quality of the image. This is shown through a top-down study of the matrices. This may be seen by looking at the matrices that each of the distinct attributes generates. If it is significant enough, we might be able to detect it by analyzing the matrices generated by the various qualities. A more detailed examination of the unique matrices that display instances of self-similarity demonstrates this abundantly. Evidently, the precision of this method has been greatly improved, as seen in **Figure 10**.

All the movies showing people moving about were filmed in controlled conditions, that is whenever a mobile lens is employed to fire, the target is kept in the field of vision so that it does not vanish entirely, but this makes it difficult to distinguish human movement from the visual flow field. Both the preliminary training phase and the final tuning phase of a 3D deep neural network are shown in **Figure 10**. During training, the network is more volatile and produces more loss; after fine-tuning, the network produces less loss, which improves the model's final classification impact and shortens the time required to learn the network. As a bonus, the paper explores the drawbacks and limits of single-mode video motion detection and gives a high-level overview of the multi-mode motion identification technique. High-quality DL results are achieved by making use of as much data as possible during training. Different varieties of depth networks have distinctive characteristics. Because it places a greater focus on the connections between seemingly unrelated data points, the CNN network excels at picture identification and detection. For the filter to be able to predict the stability of the attitude more accurately, it must be used after the 3D posture and attitude have been made. Successful motion capture allows us to automatically evaluate and interpret a wide variety of human actions and behaviors by analyzing and extracting human motion feature properties. For the purposes of this article, basketball will be used as an example of a sport that is continually pushing the boundaries. **Table 1** summarises the findings as well as a discussion of their recall and precision. Throughout the project, the 10-fold cross-validation approach and the Weka platform were frequently used.

As indicated in **Table 1**, where I is accuracy and II is recall value, the effectiveness of LSTM approach and artificial neural network leads to an improvement in the classification of diverse limb actions. Activity A is jump, B walk, C run, D average, E catch, F pass, G dribble, and H shoot. It has an average accuracy of 88.3% when it comes to upper-body motions and a recall of 88.3% when it comes to lower-body actions. However, it only has an average accuracy of 88.3% when it comes to general movements. **Table 2** shows that the average recognition and memory rates for the states of walking, running, and stationary dribbling with the

Table 1. Action recognition performance comparison on jumping, running, and walking

Activity	C5.6		Support Vector Machine		LSTM		The proposed method	
	I	II	I	II	I	II	I	II
A	88.2	88.0	87.4	87.8	87.8	87.7	88.6	88.8
B	88.1	88.2	87.8	87.8	86.6	87.6	88.7	88.6
C	87.2	87.2	87.8	87.7	88.2	86.8	88.4	88.5
D	87.8	87.8	87.7	87.7	87.3	87.3	88.5	88.6

Table 2. Action recognition performance comparison on catching the ball, passing and shooting

Behavior	C6.7		Support Vector Machine		LSTM		The proposed method	
	I	II	I	II	I	II	I	II
E	83.4	82.2	83.8	82.4	85.7	85.6	88.5	88.5
F	82.7	84.8	87.7	88	85.4	86.8	88	88
G	88.5	88.6	88.7	87.6	88.7	88.6	88.7	88.4
H	85.5	83.5	88.6	88.5	86.6	82.8	87.4	88.6

upper limbs can reach up to 99%. These rates are highest when dribbling while stationary, followed by walking, and finally sprinting. Dribbling while standing still causes the most injuries, followed by dribbling while walking, and finally dribbling while running. Both rates are far faster than the average person's personal, experience-based memory of the motion.

According to **Figure 11**, the overall accuracy percentage for detecting basketball-related activities was 98.85%. This is demonstrated by the percentage of correct identifications. Furthermore, the recognition accuracy of every basketball move was greater than 95%. The different kinds of activities are shown along the horizontal axis, and the degree of recognition accuracy is given along the vertical.

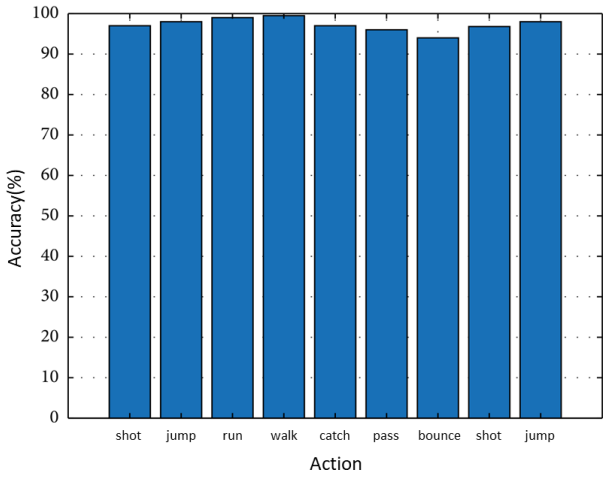

Figure 11. Proposed method action analysis.

5. Conclusion

Humans are the most important feature of their surroundings, and the information they communicate through their numerous and diverse behaviors is critical to human social interaction. As a result, the study of human migrations has significant theoretical and practical implications for a wide range of modern societal and economic elements. Since then, the network model's accuracy has significantly improved. We eliminate the human body from the weightlifting video using bone data to improve the expression of characteristics and accuracy. Self-occlusion is the norm in occlusion, and it can occur at any level of estimation, from the 2D position estimate to the 3D posture prediction. Any occlusion is, to a considerable extent, the product of the user's actions. Our present research is focused on improving the framework's processing speed and providing a more efficient approach to the occlusion problem. As a result, a new posture monitoring and identification system for callisthenics athletes is being developed with the use of machine vision technology.

Conflict of interest

The authors declare no conflict of interest.

References

1. Miller J, Nair U, Ramachandran R, Maskey M. Detection of transverse cirrus bands in satellite imagery using deep learning. Computers & Geosciences 2018; 118: 79–85. doi: 10.1016/j.cageo.2018.05.012.
2. Liu H, Jin J, Xu Z, *et al*. Deep learning based code smell detection. IEEE Transactions on Software Engineering 2021; 47(9): 1811–1837. doi: 10.1109/TSE.2019.2936376.
3. Gao M, Cai W, Liu R. AGTH-Net: Attention-based graph convolution-guided third-order hourglass network for sports video classification. Engineering 2021; 2021: 8517161. doi: 10.1155/2021/8517161.
4. Pan S. A method of key posture detection and motion recognition in sports based on Deep Learning. Mobile Information Systems 2022; 2022: 5168898. doi: 10.1155/2022/5168898.
5. Zhou J, Wang Y, Zhang W. Underwater image restoration via information distribution and light scattering prior. Computers and Electrical Engineering 2022; 100: 107908. doi: 10.1016/j.compeleceng.2022.107908.
6. Lamas A, Tabik S, Cruz P, *et al*. MonuMAI: Dataset, deep learning pipeline and citizen science based app for monumental heritage taxonomy and

classification. Neurocomputing 2021; 420: 266–280. doi: 10.1016/j.neucom.2020.09.041.
7. Nandyal S, Kattimani SL. Bird swarm optimization-based stacked autoencoder deep learning for umpire detection and classification. Scalable Computing 2020; 21(2): 173–188. doi: 10.12694/scpe.v21i2.1655.
8. Zhao X, Zuo T, Hu X. OFM-SLAM: A visual semantic SLAM for dynamic indoor environments. Mathematical Problems in Engineering 2021; 2021: 5538840. doi: 10.1155/2021/5538840.
9. Guo Q. Detection of head raising rate of students in classroom based on head posture recognition. Traitement du Signal 2020; 37(5): 823–830.
10. Bruno A, Gugliuzza F, Pirrone R, Ardizzone E. A multi-scale colour and keypoint density-based approach for visual saliency detection. Access 2020; 8: 121330–121343. doi: 10.1109/ACCESS.2020.3006700.
11. Xu Y. A sports training video classification model based on deep learning. Scientific Programming 2021; 2021: 7252896. doi: 10.1155/2021/7252896.
12. Liu J, Chen D, Wu Y, *et al*. Image edge recognition of virtual reality scene based on multi-operator dynamic weight detection. Access 2020; 8: 111289–111302. doi: 10.1109/ACCESS.2020.3001386.
13. Kalakoti G, Prabakaran G. Key-frame detection and video retrieval based on DC coefficient-based cosine orthogonality and multivariate statistical tests. Traitement du Signal 2020; 37(5): 773–784.
14. Li-quan C, You L, Shen F, *et al*. Pose recognition in sports scenes based on deep learning skeleton sequence model. Journal of Intelligent & Fuzzy Systems 2021; Preprint: 1–10.
15. Pu B, Zhu N, Li K, Li S. Fetal cardiac cycle detection in multi-resource echocardiograms using hybrid classification framework. Future Generation Computer Systems 2021; 115: 825–836. doi: 10.1016/j.future.2020.09.014.
16. Cui J, Wang M, Luo Y, Zhong H. DDoS detection and defense mechanism based on cognitive-inspired computing in SDN. Future Generation Computer Systems 2019; 97: 275–283. doi: 10.1016/j.future.2019.02.037.
17. Taborri J, Molinaro L, Santospagnuolo A, *et al*. A machine-learning approach to measure the anterior cruciate ligament injury risk in female basketball players. Sensors 2021; 21(9): 3141. doi: 10.3390/s21093141.
18. Wilkens S. Sports prediction and betting models in the machine learning age: The case of tennis. Journal of Sports Analytics 2021; 7(2): 99–117. doi: 10.3233/JSA-200463.
19. Liu G, Zhu W, Schulte O. Interpreting deep sports analytics: Valuing actions and players in the NHL. In: Brefeld U, Davis J, Van Haaren J, Zimmermann A (editors). Machine learning and data mining for sports analytics. MLSA 2018. Lecture notes in computer science, vol 11330. Cham: Springer Cham; 2019. p. 69–81.
20. Wang P, Liu ZQ. Research on evaluation of vehicle dangerous traveling state based on information fusion method. Advanced Materials Research 2013; 791–793: 1018–1022. doi: 10.4028/www.scientific.net/AMR.791-793.1018.
21. Xie T, Zhang C, Zhang Z. Utilizing active sensor nodes in smart environments for optimal communication coverage. IEEE Access 2018; 7: 11338–11348. doi: 10.1109/ACCESS.2018.2889717.
22. Li T, Sun J, Wang L. An intelligent optimization method of motion management system based on BP neural network. Neural Computing & Applications 2021; 33(2): 707–722. doi: 10.1007/s00521-020-05093-1.
23. Barshan B, Yüksek MC. Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units. The Computer Journal 2014; 57(11): 1649–1667. doi: 10.1093/comjnl/bxt075.
24. Zhang C, Xie T, Yang K, *et al*. Positioning optimisation based on particle quality prediction in wireless sensor networks. IET Networks 2019; 8(2): 107–113. doi: 10.1049/iet-net.2018.5072.
25. Chunfeng G. Research on pre-competition emotion recognition of student athletes based on improved machine learning. Journal of Intelligent and Fuzzy Systems 2020; 39(4): 5687–5698. doi: 10.3233/JIFS-219218.
26. Ding Q, Ding Z. Machine learning model for feature recognition of sports competition based on improved TLD algorithm. Journal of Intelligent and Fuzzy 2021; 40(2): 2697–2708. doi: 10.3233/JIFS-189312.
27. Li P, Cai J. Research on image recognition from wireless sensor data based on deep learning. Journal of Critical Care 2017; 42(6): 2607–2612.
28. Colyer SL, Evans M, Cosker DP, Salo AIT. A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system. Sports Medicine-Open 2018; 4: 24. doi: 10.1186/s40798-018-0139-y.
29. Pan D, Liu H, Qu D, Zhang Z. Human falling detection algorithm based on multisensor data fusion with SVM. Mobile Information Systems

2020; 2020: 8826088. doi: 10.1155/2020/8826088.
30. Nagalakshmi Vallabhaneni DPP. The analysis of the impact of yoga on healthcare and conventional strategies for human pose recognition. Turkish Journal of Computer and Mathematics Education (TURCOMAT) 2021; 12(6): 1772–1783. doi: 10.17762/turcomat.v12i6.4032.
31. Yan L, Shen M, Yao W, *et al*. Recognition method of lactating sows' posture based on sensor MPU6050. Transactions of the Chinese Society for Agricultural Machinery 2015; 46(5): 279–285. doi: 10.6041/j.issn.1000-1298.2015.05.040.
32. Hu X, Zhang Y. Study of face detection algorithm based on complexion. Journal of Hefei University of Technology 2012; 35(7): 908–912. doi: 10.3969/j.issn.1003-5060.2012.07.011.
33. Kashyap R. Evolution of histopathological breast cancer images classification using stochastic dilated residual ghost model. Turkish Journal of Electrical Engineering & Computer Science 2021; 29: 2758–2779. doi: 10.3906/elk-2104-40.