IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DEEP LEARNING FOR RECOGNIZING HUMAN ACTIVITIES THROUGH MOTION OF SKELETAL JOINTS

Aman Nijhawan[1], Tanish Checker[2], Vaishali Kedare[3], Dr. Archana Gulati[4]

[1-3]Student, Computer Department, Datta Meghe College of Engineering, Navi Mumbai, Maharashtra, India

[4]Faculty, Computer Department, Datta Meghe College of Engineering, Navi Mumbai, Maharashtra, India

Abstract - *With advances in consumer electronics, demands have increased for greater granularity in differentiating and analyzing human daily activities. Moreover, the potential of machine learning, and especially deep learning, has become apparent as research proceeds in applications such as monitoring the elderly, and surveillance for detection of suspicious people and objects left in public places. Although some techniques have been developed for Human Action Recognition (HAR) using wearable sensors, these devices can place unnecessary mental and physical discomfort on people, especially children and the elderly. Therefore, research has focused on image based HAR, placing it on the front line of developments in consumer electronics. This project proposes an intelligent human action recognition system which can recognize human activities specific to sports, from an image using a skeletonization model that analyzes specific joint angles, combining the techniques of image processing and deep learning. Moreover, due to low computational cost and high accuracy outcomes, an approach using skeleton information has proven very promising[1]. Therefore, this project discusses the development of an effective skeleton information based system which can enhance sport-specific activities.*

Key Words: **Deep Learning, Human Action Recognition, Skeletonization model, Joint Angles.**

1.INTRODUCTION

The area of athletics and other various sports concerned with it has been something that many people have been intimidated by for a long time. Truth be told, some of the most common reasons for this is that most people have not been guided correctly, or they have been discouraged from doing that while growing up. Another reason for this is that most people just cannot afford to hire a coach to learn how to become a better athlete. Through our research, we have attempted to solve this problem with the help of a personal virtual trainer. The program will be helpful in assessing and diagnosing specific faulty mechanics of the user's input motion, and suggest specific enhancements to it. This research does not intend to suggest that a virtual assistant can be better than a hands-on coach. No technology, in our opinion, can match the experience, and the instinct of a real coach who has spent years in the trenches, training athletes. Rather, our work can be said to be a much-needed tool, which is available anywhere and at any time, as convenient to the user. Furthermore, this can take away the fear and intimidation that many people have of approaching an expert, granting them complete privacy.

2. METHODOLOGY

2.1 Project Design

The basic architecture of the proposed approach is shown in the figure and the flow chart. The first step involves the recording of a live video of a person's activity, for example, running or jumping. Each frame in the video is processed through the human pose estimation algorithm to obtain a skeleton image comprising 25 body points.

The obtained image is given as an input to a CNN trained to classify multiple classes, which will include the specific mechanics for each activity.

Once the trained CNN receives the input skeletal image of a person, the output is given as the image with a label of predicted class.

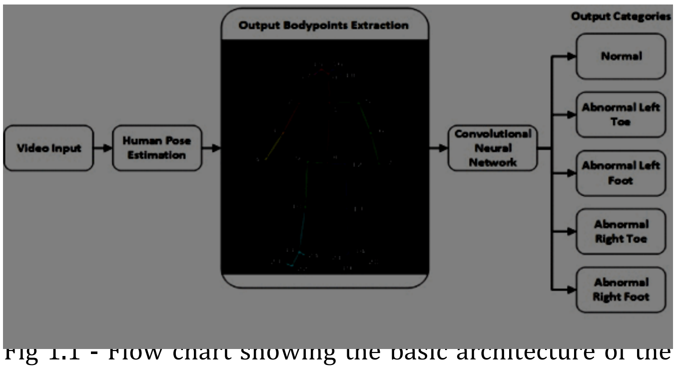


Fig 1.1 - Flow chart showing the basic architecture of the system.

The outputs will be given in real time, such that the user will be able to identify faulty mechanics and correct them right away. The estimation of proper mechanics for each activity would be done by analyzing the joint angles, as they are the best way to estimate the proper execution of an activity.

2.2 Technologies Used

Convolutional Neural Network

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics.

The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlap to cover the entire visual area.

TensorFlow

TensorFlow is a free and open-source software library for machine learning. It can be used across a range of tasks but has a particular focus on training and inference of deep neural networks.[4][5]

Tensorflow is a symbolic math library based on dataflow and differentiable programming. It is used for both research and production at Google.[6][7][8]

TensorFlow was developed by the Google Brain team for internal Google use. It was released under the Apache License 2.0 in 2015.

OpenPose

OpenPose is the first real-time multi-person system to jointly detect human body, hand, facial, and foot key-points (in total 135 key-points) on single images. It was proposed by researchers at Carnegie Mellon University. They have released in the form of Python code, C++ implementation and Unity Plugin.

Hardware Requirements

The training part of the CNN will be performed on high performance NVIDIA GPUs. OpenPose requires an NVIDIA GPU, with CUDA (Compute Unified Device Architecture) installed. CUDA is NVIDIA's proprietary parallel computing platform and application programming interface model. (NVIDIA GTX 1050 used for the project).

Ideally, a machine having a minimum of an Intel Core i5, or an equivalent AMD Ryzen chip will be required. (Intel Core i5 – 8250U used for the project).

Functional Requirements

Should be able to capture a proper image for analysis of specific joint angles. Should apply the pose estimation algorithm instantly after capturing the image, and transfer it to the CNN. An analysis page for the faults found in the user's action.

Non Functional Requirements

Usability – The most important requirement of our project. The program should be easy enough to use and understand by everyone.

Performance – Quick results and feedback are the cornerstone of this software. The user should not have to wait for specific functions to execute.

Extensive capacity – The system will be highly capable of accepting new modules for integration. Constant efforts will be made to ensure that the storage requirements stay within an acceptable and compact range.

3. PROPOSED METHOD

Working

Several technologies embedded in specialized devices are being used for real-time human activity detection and recognition, muscle activation measurement, and a lot of other measurements of the functionality of the human body. With the recent improvements in the field of Artificial Intelligence (AI), especially in deep learning, it is possible now to create a mechanism where the translation of the data can be performed by a deep learning tool such as Convolutional Neural Network (CNN). Therefore, this work presents an approach where human pose-estimation is combined with a CNN for classification between normal and abnormal mechanics of a human's athletic ability with an ability to provide information about the detected abnormalities form an extracted skeletal image in real-time. Recently, Convolutional Neural Networks (CNNs) have achieved great results in different fields of pattern recognition, detection, and classification, especially in computer vision. CNN is a class of deep neural networks and is mostly applied to analyze visual imagery. Human pose estimation is combined with a deep neural network such as CNN to classify the normal and abnormal mechanics of a specific activity. The reason to use human pose estimation for the project is that in pose estimation, deep learning-based CNN is used to detect the body points of a person. This gives us skeletal images where body points are joined to form a skeleton of a person. Only these obtained skeletal images are further used for classification and analysis of the mechanics. All of this is being done by the OpenPose algorithm.

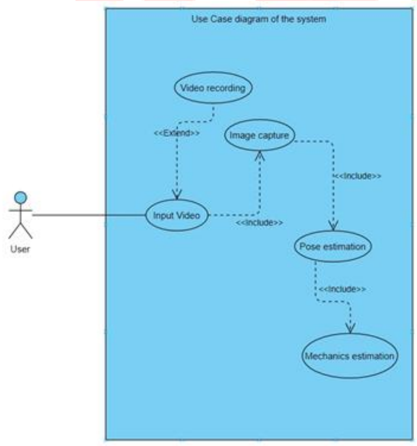
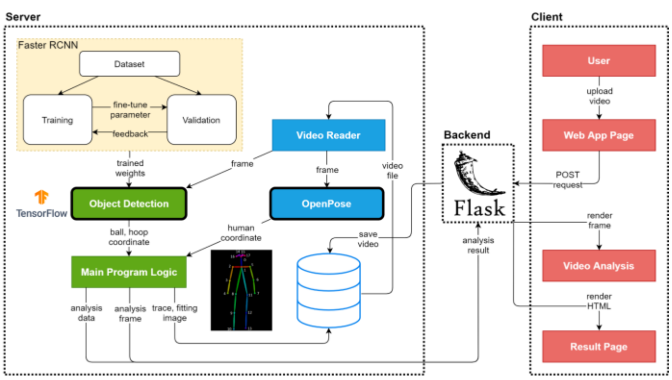


Fig 3.1 – Use Case Diagram

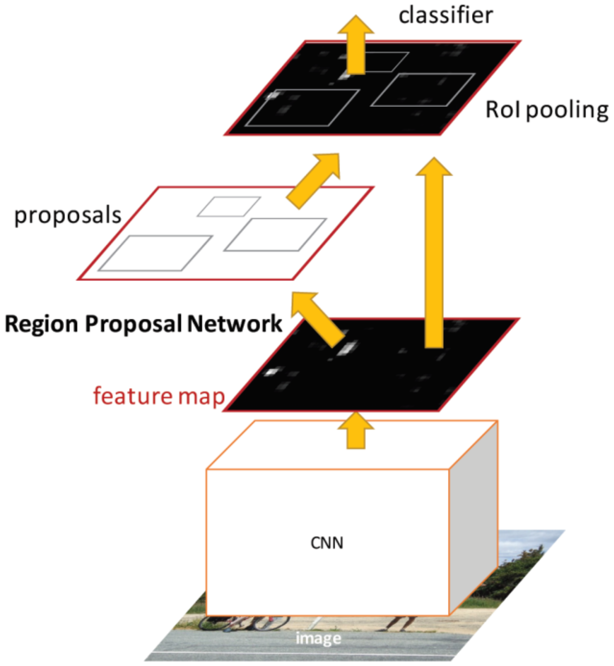


Ours is an artificial intelligence application built on the concept of object detection. We analyze basketball shots by digging into the data collected from the object detection API by Tensorflow, whuch is pretrained on the faster RCNN. We can get the result by simply uploading files to the web app. All the data for the shooting pose analysis is calculated by implementing OpenPose. The OpenPose API and the Tensorflow object detection API are managed by Flask, which communicates their results with the front end, hich is constructed with HTML5 and CSS3. We used Flask because the integration of APIs with Flask is much simpler than with other models, and also because Flask could help us host the entire program online in the future, if we wanted to.

This project has three main features, **shot analysis, shot detection, detection API**. Counting shooting attempts and missing shots, and scoring shots from the input video. Implementing **OpenPose** to calculate the angle of elbow and knee during shooting.

Release angle and release time are calculated by all the data collected from shot analysis and pose analysis. Please note that there will be a relatively big error for the release time since it was calculated as the total time when the ball is in hand.

Detection Model



The object detection model is trained with the Faster R-CNN model architecture, which includes pretrained weight on COCO dataset.

Testing

For testing, we are coming up with a self-defined data set. It will be made later and trained accordingly with the required parameters of the sport. The testing process of our program will be the most extensive part, as compared to the others. Since the entire efficiency of the application depends upon accurate test results, which will in turn train the program itself, this part will consume a lot of time. Testing in itself will serve as a form of training for the program since the CNN will be able to associate the correct parameters with the results of respective accuracy.

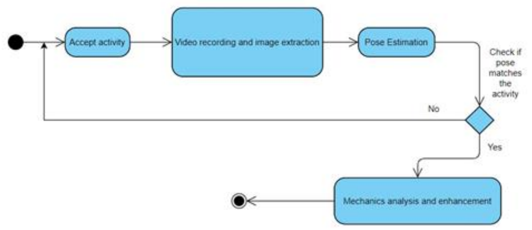


Fig 3.2 – Activity Diagram

4. CONCLUSION

In this work, an approach to develop an efficient activity analysis mechanism using deep learning tools such as Convolutional Neural Network (CNN)-based classifier is presented. The accuracy and the effectiveness of the proposed system remains to be seen. However it is estimated to be more than that of the methods previously undertaken. CNN will undoubtedly boost the accuracy of the pose estimation and motion recognition part, compared to methods like regression analysis. Future work can include the addition of a wide variety of activity recognition and enhancement features for different sports, seeing as each sport has its own unique mechanics, demanding a different set of abilities from each player. A project will also be undertaken to port the system to a mobile-based application, seeing as smartphones are the most popular choice of devices among people these days.

REFERENCES

1. A. Rohan, M. Rabah, T. Hosny and S. -H. Kim, "Human Pose Estimation-based Real-time Gait Analysis using Convolutional Neural Network," in IEEE Access, doi: 10.1109/ACCESS.2020.3030086

2. E. Corona, A. Pumarola, G. Alenyà and F. Moreno-Noguer, "ContextAware Human Motion Prediction," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 6990-6999

3. S. Win and T. L. L. Thein, "Real-Time Human Motion Detection, Tracking and Activity Recognition with Skeletal Model," 2020 IEEE Conference on Computer Applications (ICCA), Yangon, Myanmar, 2020, pp. 1-5, doi: 10.1109/ICCA49400.2020.9022822.

4. Oudah M, Al-Naji A, Chahl J. Hand Gesture Recognition Based on Computer Vision: A Review of Techniques. Journal of Imaging. 2020; 6(8):73.

5. Golestani, N., Moghaddam, M. Human activity recognition using magnetic induction-based motion signals and deep recurrent neural networks. Nat Commun 11, 1551 (2020).24

6. . İnce, Ö.F., Ince, I.F., Yıldırım, M.E., Park, J.S., Song, J.K. and Yoon, B.W. (2020), Human activity recognition with analysis of angles between skeletal joints using a RGB-depth sensor. ETRI Journal, 42: 78-89.

7. Zhang H-B, Zhang Y-X, Zhong B, Lei Q, Yang L, Du J-X, Chen D-S. A Comprehensive Survey of Vision-Based Human Action Recognition Methods. Sensors. 2019; 19(5):1005.

8. R. Ding, Q. He, H. Liu and M. Liu, "Combining Adaptive Hierarchical Depth Motion Maps With Skeletal Joints for Human Action Recognition," in IEEE Access, vol. 7, pp. 5597-5608, 2019, doi: 10.1109/ACCESS.2018.2886362.

9. Pichao Wang, Wanqing Li, Philip Ogunbona, Jun Wan, Sergio Escalera, RGB-D-based human motion recognition with deep learning: A survey, Computer Vision and Image Understanding, Volume 171, 2018, Pages 118-139, ISSN 1077-3142

10. Xu, Honghua & Li, Li & Fang, Ming & Zhang, Fengrong. (2018). Movement Human Actions Recognition Based on Machine Learning. International Journal of Online Engineering (iJOE). 14. 193. 10.3991/ijoe.v14i04.8513