International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307

Available Online at : www.ijsrcseit.com

doi : https://doi.org/10.32628/IJSRCSEIT

Unravelling Human Actions with Deep Learning Techniques

K. Bharani[1], Dr. M. Saravanamuthu[2]

PG Student[1], Assistant Professor[2]

Department of Computer Applications, Madanapalle Institute of Technology &Science, Madanapalle, Andhra Pradesh, India

ARTICLEINFO

Article History:

Accepted: 01 July 2023

Published: 20 July 2023

Publication Issue

Volume 9, Issue 4

July-August-2023

Page Number

131-139

ABSTRACT

Human activity recognition plays a crucial role in interpersonal communication and relationships as it provides insights into a person's identity, personality, and psychological state. Extracting this information is challenging due to its complex nature. The scientific fields of computer vision and Deep learning extensively study the human ability to recognize activities, leading to the development of various applications such as video surveillance systems, human-computer interaction, and characterization of human behaviour in robotics. Recognizing multiple activities simultaneously is a requirement for many of these applications. Numerous publications have focused on the important field of human activity recognition in images and videos. An OpenCV-based deep learning algorithm, specifically a Convolutional Neural Network (CNN), is proposed in this paper. This calculation can successfully prepare datasets and precisely perceive human activities and exercises.

Keywords : Recognition of human actions/activities, deep learning techniques, convolutional neural networks (CNN), MobileNet, Nasnet, Convnext, OpenCV.

I. INTRODUCTION

Human action recognition is a complex task in computer vision that involves the identification and comprehension of human activities from visual data. Deep learning techniques, specifically convolutional neural networks (CNNs), have made significant advancements in this field. When combined with OpenCV, a popular computer vision library, deep learning models have demonstrated promising outcomes in accurately recognizing and categorizing various human actions.

Deep learning models designed for human action recognition aim to autonomously learn and extract significant features from raw visual data like images or videos. These models utilize convolutional layers to capture spatial information and temporal dependencies, enabling them to comprehend the motion and dynamics associated with different actions.

The Convolutional Neural Network (CNN) is a widely employed deep learning architecture for action

recognition. CNNs possess the ability to automatically acquire hierarchical representations by applying convolutional filters to input data. This characteristic makes them well-suited for capturing relevant features in human action recognition tasks.

In recent years, several CNN-based architectures such as MobileNet, Nasnet, and Convnext have gained popularity due to their effectiveness in various computer vision tasks, including action recognition. These architectures are designed to strike a balance between accuracy and computational efficiency, making them suitable for real-time or resource-constrained applications.

OpenCV, a powerful and widely-used computer vision library, offers a variety of tools and functions for image and video processing. It provides functionalities for preprocessing, feature extraction, and post-processing, which are critical steps in human action recognition pipelines. OpenCV seamlessly integrates with deep learning frameworks like TensorFlow and PyTorch, allowing the incorporation of pre-trained CNN models for action recognition tasks.

The combination of deep learning models, such as CNNs, and the extensive toolset provided by OpenCV empowers researchers and developers to create robust and accurate human action recognition systems. These systems have diverse applications in domains such as video surveillance, human-computer interaction, sports analytics, and healthcare.

human action recognition using deep learning with OpenCV has emerged as a promising approach for automatic identification and classification of human activities from visual data. The integration of deep learning models, especially CNN architectures like MobileNet, Nasnet, and Convnext, with the powerful features offered by OpenCV establishes a solid foundation for developing efficient and accurate action recognition systems.

II. Related Works

[1]: The increasing sophistication of mobile devices has led to the integration of various powerful sensors, such as GPS, cameras, microphones, light sensors, temperature sensors, magnetic compasses, and accelerometers. These sensors enable mobile devices, particularly the latest generation smartphones with different operating systems, to collect and communicate textual and voice signals effectively. As a result, modern mobile devices have become equipped with highly sensitive sensors.This paper focuses on a specific approach that addresses the problem of human activity classification using mobile devices carried by users. The proposed method utilizes the K-Nearest Neighbor algorithm (K-NN) for this purpose. By leveraging the capabilities of the built-in sensors in mobile devices, the algorithm aims to classify and recognize different human activities accurately.

The approach described in this paper takes advantage of the diverse range of sensors available in mobile devices, including GPS, cameras, microphones, and accelerometers, among others. These sensors provide valuable data that can be used to infer and classify the activities performed by users.The utilization of the K-Nearest Neighbor algorithm allows the mobile device to compare the sensor data collected during various activities with the pre-defined patterns and labels. By finding the closest matches, the algorithm can accurately classify the current human activity.The proposed approach offers a promising solution to the challenge of human activity classification using mobile devices. By leveraging the multitude of sensors present in modern smartphones, this method enables the recognition and classification of diverse activities based on the collected sensor data.

This paper presents a dedicated method for addressing the classification of human activities through the use of a mobile device carried by the individual. The current approach relies on the application of the K-Nearest Neighbor algorithm (K-NN) and utilizes the magnitude of the accelerometer data. By leveraging this algorithm, it becomes feasible to accurately recognize and categorize the general activities performed by the user.

[2]: Activity recognition plays a crucial role in various applications, including human surveillance systems, medical research, and the emerging fields of smart homes and smart health. Mobile devices, equipped with built-in sensors like gyroscopes, accelerometers, GPS, and compass sensors, provide an opportunity to capture user behaviour and activity. By leveraging data mining and machine learning techniques, an activity recognition (AR) system can process the raw sensor data from these mobile sensors and accurately estimate human activities. This study focuses on evaluating the performance of two algorithms, namely Random Forest (RF) and Modified Random Forest (MRF), within an online activity recognition framework on Android platforms. The proposed method enables real-time training and efficient classification of accelerometer data for activity recognition.

Activity Recognition (AR) frameworks analyze raw sensor data obtained from small-sized sensors to assess human movements using data mining and machine learning techniques. This paper investigates the performance of two classification algorithms, namely Random Forest (RF) and Modified Random Forest (MRF), within an online Activity Recognition framework implemented on Android platforms. The proposed technique enables real-time training and accurate classification of accelerometer data, making it highly effective.

[3]: Mobile devices have advanced significantly, and the latest generation of smartphones incorporates a wide range of powerful sensors. These sensors include GPS, vision (camera), audio (microphone), light, temperature, direction (magnetic compass), and acceleration (accelerometer) sensors. The availability of these sensors in widely used communication devices opens up exciting opportunities for data mining and its applications. This paper presents and evaluates a system that utilizes the accelerometer sensor in smartphones to perform activity recognition. The task of activity recognition involves identifying the physical activity being performed by the user. To develop our system, we gathered labelled accelerometer data from twenty-nine users engaged in everyday activities such as walking, jogging, climbing stairs, sitting, and standing. We then processed this time series data and created examples that summarize the user's activity over 10-second intervals.In this research paper, we present and assess a system that leverages accelerometer sensors in smartphones to perform activity recognition, which entails identifying the specific physical activity being performed by a user. To develop our system, we gathered accelerometer data from twenty-nine participants, who engaged in various everyday activities such as walking, jogging, climbing stairs, sitting, and standing. We then processed and consolidated this time-series data into concise examples that represent the user's activity within 10-second intervals.

[4]: In this study, the researchers developed and assessed algorithms for detecting physical activities based on data collected from five small biaxial accelerometers worn simultaneously on different body parts. The acceleration data was obtained from 20 subjects in an unsupervised manner, without direct researcher supervision. The subjects were instructed to perform a series of typical daily tasks without specific guidelines on where or how to execute them. From the acceleration data, various features like mean, energy, frequency-domain entropy, and correlation were calculated. These features were put to the test in a number of classifiers. Among them, choice tree classifiers displayed the best execution in perceiving ordinary exercises, accomplishing a general exactness pace of 84%. The discoveries show that while specific exercises can be precisely perceived utilizing subject-autonomous preparation information, others might require subject-explicit preparation information for ideal execution.The main objective of the study was to create and evaluate algorithms for detecting physical activities using data collected from five small biaxial accelerometers positioned on different parts of the body. The acceleration data was gathered from 20 participants who carried out everyday tasks without any supervision or observation from researchers. To

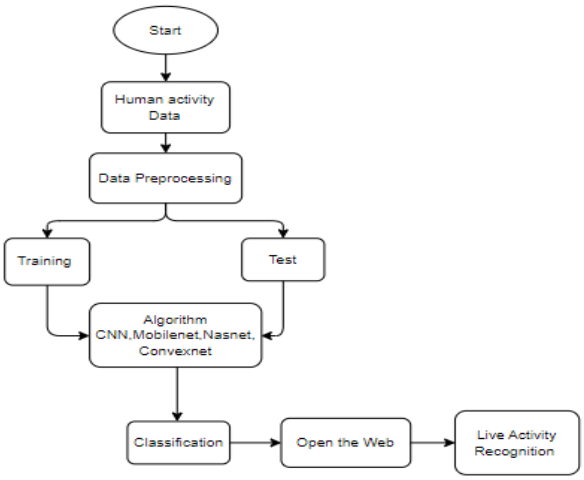
assess the effectiveness of activity detection, several classifiers were tested using various features derived from the acceleration data, including mean, energy, frequency-domain entropy, and correlation. The study aimed to determine the performance of these classifiers in accurately detecting different activities.

[5]: We propose a framework that uses an Android cell phone to gather and show sensor information on the gadget's screen while at the same time streaming it to a focal server. To ensure seamless data transfer between devices, the system makes use of both Bluetooth and wireless Internet connections. Also, we have carried out Close to Handle Correspondence (NFC) innovation to lay out programmed Bluetooth associations and start application execution, improving proficiency and comfort.This system holds particular value in the domain of body sensor networks (BSN) developed for medical healthcare applications. To exhibit its capacities, we use an accelerometer, a temperature sensor, and electrocardiography (ECG) signal information for exploratory purposes. Through graphical or textual representations, the raw sensor data is processed and presented on both the smartphone and the central server. Additionally, a Java-based central server application is used to communicate with the Android system and make it possible to store and analyze data.

III.Proposed Method

We are developing a system for recognizing human actions or activities using Convolutional Neural Networks (CNN), specifically the CNN models MobileNet, Nasnet, and Convexnet. Our approach involves utilizing a dataset of images extracted from videos, with each image representing a frame. We train the dataset using CNN models. Once the training is completed, we employ OpenCV to recognize actions in real-time video footage. By capturing the video, our system can accurately identify and classify human actions. The proposed method's block diagram is illustrated in the accompanying figure.

Block Diagram:



Advantages:

- High feature compatibility
- Time Saving
- Low complexities

Applications:

- Surveillance.
- Gaming.
- Animation and Active and Assisted Living (AAL).

IV. Methodology

Convolutional Neural Network:

An advanced form of deep learning known as a Convolutional Neural Network (CNN) was made to process structured grid-like data like images. By achieving exceptional performance in a variety of image-related tasks like image classification, object detection, and semantic segmentation, its emergence has revolutionized the field of computer vision.

CNNs are made up of multiple layers, such as convolutional layers, pooling layers, and fully connected layers, which are modeled after the visual processing mechanisms that are found in the human brain. Local operations on receptive fields are carried out by the convolutional layers, which use learnable filters to extract features from the input data. At various scales, these filters capture various patterns like edges, textures, and shapes.

Pooling layers play a crucial role in reducing the spatial dimensions of the feature maps, which reduces computational complexity and provides translational invariance. Common pooling operations employed in CNNs include max pooling and average pooling.

Following the convolutional and pooling layers, the output is fed into fully connected layers. These layers facilitate high-level feature learning and decision-making. By connecting every neuron to each neuron in the subsequent layer, the network can learn intricate relationships and make predictions based on the acquired features.

To train a CNN, a large labeled dataset is utilized, enabling the network to optimize its parameters through backpropagation. During training, the network adjusts its weights and biases to minimize a defined loss function, thus enabling accurate predictions on unseen data.

CNNs have demonstrated exceptional performance across various computer vision tasks, particularly image classification. Prominent models such as AlexNet, VGGNet, ResNet, and InceptionNet have achieved remarkable results on benchmark datasets like ImageNet. Transfer learning, where pre-trained CNN models are fine-tuned for specific tasks, has become prevalent due to the effectiveness of CNNs.

In summary, Convolutional Neural Networks are a powerful class of deep learning algorithms designed specifically for processing grid-like data, such as images. Their ability to capture meaningful features has significantly advanced the field of computer vision, enabling accurate image recognition and analysis.

MobileNet:

MobileNet is an efficient convolutional neural network (CNN) architecture specifically designed for deep learning tasks, particularly targeting mobile and resource-constrained devices. Its primary purpose is to address the challenges associated with deploying deep learning models on devices with limited computational power and memory.

The key objective of MobileNet is to strike a balance between model size and accuracy, as traditional deep learning models tend to be large and computationally demanding, making them impractical for deployment on resource-limited devices. MobileNet aims to provide a lightweight alternative that can still achieve competitive performance.

To achieve this, the architecture of MobileNet introduces a technique called depthwise separable convolution, which splits the standard convolution operation into two parts: depthwise convolution and pointwise convolution. This division significantly reduces the computational cost by decreasing the number of parameters and operations involved.

The depthwise convolution applies a single filter to each input channel independently, producing a set of intermediate feature maps. This step captures spatial information within each channel separately. Subsequently, the pointwise convolution, also known as 1x1 convolution, combines information from different channels to generate new features through a linear combination.

The use of depthwise separable convolutions in MobileNet enables it to reduce the model size and computational complexity while still capturing

essential features necessary for various computer vision tasks. This architectural design allows MobileNet to achieve a favorable trade-off between efficiency and accuracy.

Furthermore, MobileNet models often incorporate additional techniques such as depthwise separable bottlenecks, which further enhance efficiency by reducing the number of parameters and computations. These bottlenecks employ 1x1 convolutions to shrink the channel dimensions and subsequently expand them to capture complex features.

MobileNet has proven successful in a wide range of applications, including image classification, object detection, and semantic segmentation. Its lightweight architecture makes it particularly well-suited for mobile and embedded devices where computational resources are limited.

In summary, MobileNet stands as an important contribution to the field of deep learning by providing an efficient CNN architecture for deploying models on mobile and resource-constrained platforms without compromising performance. Its design principles have opened doors for further advancements in developing lightweight deep learning models applicable to various real-world scenarios.

NasNet:

Nasnet, abbreviated for Neural Architecture Search Network, is a deep learning architecture that aims to automate the design of neural networks. It focuses on discovering optimal network architectures for specific tasks using a technique known as neural architecture search (NAS).

In traditional approaches, neural network designs are manually crafted by experts based on their domain knowledge. However, this process can be time-consuming and limited in exploring the vast design possibilities.

Nasnet addresses this challenge by automating the design process. It employs reinforcement learning or evolutionary algorithms to search and optimize network architectures for a given task. The goal is to find network configurations that achieve high performance while minimizing computational requirements.

The innovation of Nasnet lies in its ability to discover architectural building blocks, such as cells, which can be stacked and repeated to create complex network structures. These building blocks are learned through the NAS process and capture important features and connectivity patterns relevant to the task at hand.

Nasnet explores a wide range of potential architectures by iteratively sampling and evaluating different candidates. It employs a search strategy to guide the exploration towards promising architectures that deliver high performance. This process often involves training and evaluating thousands or even millions of architectures to identify the optimal one.

The resulting Nasnet architectures are typically efficient and effective, achieving state-of-the-art performance in computer vision tasks like image classification, object detection, and semantic segmentation. They are designed to strike a balance between accuracy and computational efficiency, making them suitable for deployment on devices with limited resources.

Nasnet has significantly advanced the field of neural architecture design by automating the process and enabling researchers and practitioners to explore and discover novel network architectures. By leveraging Nasnet, the development of neural networks can be expedited and customized for specific tasks, leading to enhanced performance and efficiency in deep learning applications.

Convnext:

ConvNeXt is a deep learning architecture that tackles the challenge of capturing local and global dependencies in visual data. Its main objective is to enhance the performance of convolutional neural networks (CNNs) by effectively modeling complex relationships within the data.

ConvNeXt introduces two key innovations: grouped convolutions and cross-channel interactions. In traditional CNN architectures, standard convolutions are used, where each filter operates on the entire input volume. However, ConvNeXt implements grouped convolutions, dividing the input channels into groups and applying separate convolutions to each group. This approach enables the network to capture more localized information while maintaining the independence of different channel groups.

In addition to grouped convolutions, ConvNeXt incorporates cross-channel interactions through a technique called "cross-channel parametric pooling." This pooling operation allows the network to capture global dependencies across different channels, facilitating the extraction of higher-level features.

By combining grouped convolutions and cross-channel interactions, ConvNeXt aims to strike a balance between capturing local and global information. This architectural design enhances the expressiveness and effectiveness of the network in modeling intricate patterns and relationships within the data.

ConvNeXt has demonstrated impressive performance on image classification, object detection, and semantic segmentation, among other computer vision tasks. It has accomplished serious or cutting edge results on benchmark datasets, exhibiting its viability in demonstrating complex visual examples and accomplishing high exactness.

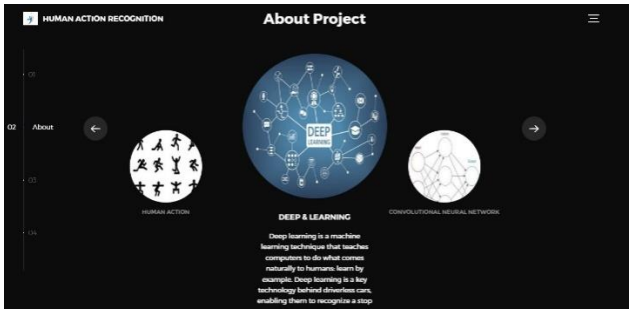
ConvNeXt's architectural design can also be used with different backbone CNNs, like ResNet or Inception, to better capture local and global dependencies for researchers and practitioners. ConvNeXt can be adapted to specific datasets and tasks by incorporating it into existing deep learning frameworks thanks to its adaptability.

To summarize, ConvNeXt is a deep learning architecture that addresses the modeling of local and global dependencies in visual data. Through grouped convolutions and cross-channel interactions, ConvNeXt enhances the expressive power of CNNs, leading to improved performance in various computer vision tasks.

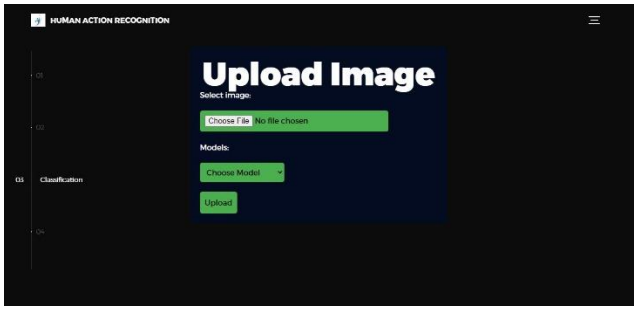
V. Result

Home Page:

About Page:



Upload Page:



Prediction Page:

Output Page:

VI. Conclusion

Our proposed model focuses on the recognition of human actions or activities by employing deep learning techniques, specifically Convolutional Neural Networks (CNN), MobileNet, Nasnet, and Convnext, in conjunction with the OpenCV library. To accomplish this, we utilized a dataset comprising images that depict various actions and activities. These images were then used to train the deep learning algorithms, including CNN, MobileNet, Nasnet, and Convnext. After completing the training process, we leveraged the capabilities of OpenCV to capture video and perform real-time recognition of the actions or activities exhibited in the video.

VII. REFERENCES

[1]. Activity Recognition with Smartphone Sensors, Xing Su, Hanghang Tong, and Ping Ji, tsinghua science and technology ISSN ll1007-0214ll02/11pp235-249, June 2014, Volume 19.

[2]. Nicholas D.Lane , Emiliano Miluzzo, Hong Lu, Daniel Peebles,‖A Review of Cell Phone Sensing‖, IEEE Interchanges Magazine September 2010

[3]. Activity recognition using the knearest neighbor algorithm on a smartphone with a tri-axial accelerometer, SahakKaghyan and HakobSarukhanyan, International Journal "Information Models and Analyses" Vol. 1 / 2012J.

[4]. Mustafa Kose, OzlemDurmazIncel,CemErsoy ‖Online Human Movement Acknowledgment on Shrewd Phones‖2nd Global Studio on Versatile Detecting, April 16, 2012.

[5]. "Recognizing Human Activities User Independently on Smartphones Based on Accelerometer Data," by PekkaSirrtola and JuhaRöning, is published in the International Journal of Artificial Intelligence and Interactive Multimedia, Vol. 1, No 5.

[6]. Activity Recognition Using Cell Phone Accelerometers, sensor KDD '10, July 25, 2010, Washington, DC, USA, Jennifer R. Kwapisz, Gary M. Weiss, and Samuel A. Moore. Right of use: 2010ACM.

[7]. Activity Recognition from User-Annotated Acceleration Data, by Ling Bao and Stephen S. Intille. Ferscha and F. Mattern (Eds) Inescapable 2004, LNCS 3001, pp. 1–17, 2004. 2004. Springer Berlin Heidelberg

[8]. "Mobile Sensor Data Collector using Android smartphone," by Won-Jae Yi, WeidiJia, and JafarSaniie, Department of Electrical and Computer Engineering, Illinois Institute of Technology, 3301 S. Dearborn St.103SH, Chicago, IL, USA.

[9]. B. Brown, H. Muller, I. Anderson, J. Maitland, S. Sherwood, L. Barkhuus, M. Chalmers, M. Hall, and

B. Brown Shakra: following and sharing everyday action levels with unaugmented cell phones. In 2007, Mobile Networks and Applications, 12(2-3), pp. 185-199

[10]. Android. A look at the sensors. 2014. [http://developer.android.com/guide/topics/sensors/sensors_overview.html] Online; [accessed March 1, 2014]

[11]. D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz. A support vector machine with multiple classes that is friendly to hardware is used to identify human activity on smartphones. In Encompassing Helped Residing and Home Consideration, pages 216-223. Springer, 2012.

[12]. Apple. 2014 Nike + iPod Application [Online; gotten to 01-Walk 2014]

[13]. Apple. Reference to the UIAcceleration Class. https://developer.apple.com/library/ios/documentation/uikit/reference/UIAcceleration_Class/Reference/UIAcceleration.h tml, 2014. [Online; gotten to 17-Walk 2014]

[14]. W.- Y. Deng, Q.- H. Zheng, and Z.- M. Wang. Recognizing cross-person activities with an extreme learning machine with a reduced kernel Neural System

[15]. Chen Z, Zhu Q, Soh YC, and Zhang L. (2017). CT-PCA and online SVM-based robust human activity recognition with smartphone sensors. 13(6):3070–3080, IEEE Trans IndustrInf

[16]. Cho H, Yoon SM (2018) Separation and vanquish based 1D CNN human action acknowledgment utilizing test information honing. 18 (4): 1055, Sensors

[17]. Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y (2014) Learning phrase portrayals utilizing RNN encoder-decoder for measurable machine interpretation. In: Doha, Qatar, Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1724–1734

[18]. Chung J, Gulcehre C, Cho K, Bengio Y (2014) Experimental assessment of gated intermittent brain networks on succession demonstrating. In: Deep learning workshop at NIPS 2014

[19]. Arcface: Deng J, Guo J, Xue N, and Zafeiriou S (2019) for deep face recognition, additive angular margin loss In: Procedures of the IEEE Meeting on PC Vision and Example Acknowledgment, pp 4690-4699

[20]. Dobbin KK, Simon RM (2011) Ideally dividing cases for preparing and testing high layered classifiers. BMC Drug Genom 4(1):31

[21]. Personalization of classification models for the recognition of human activity by Ferrari A, Micucci D, Mobilio M, and Napoletano P (2020). IEEE Access 8, pages 32066-32079

[22]. Accelerometry-based detection of posture and motion, by Foerster F, Smeja M, and Fahrenberg J a monitoring ambulatory validation study. Comput Murmur Behav 15(5):571-583

[23]. Gao W, Zhang L, Teng Q, Wu H, Min F, He J (2020) DanHAR: double consideration network for multimodal human action acknowledgment utilizing wearable sensors.

[24]. Johnson RA, Miller I, and Freund JE (2000) provided engineers with probability and statistics. Pearson Training, London, p 642

[25]. Jordao A, Kloss R, Schwartz WR (2018) Inert HyperNet: investigating the convolutional neural network's layers. In: IEEE, Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), 1–7

Cite this article as :

K. Bharani, Dr. M. Saravanamuthu, "Unravelling Human Actions with Deep Learning Techniques", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 4, pp.131-139, July-August-2023.