International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 12 Issue 5 ǁ May 2024 ǁ PP. 135-143

Enhancing Human Activity Recognition Through A Fusion Of Hybrid Features And Transformer Models

Prof. N GAYATHRI
Darshan M[1], Pavan Yadav A[2], Tribhuvan N[3], Varun Nagraj[4], Dr. Nijaguna G.S[5], Prof. N Gayathri [6], Prof. Sumitha[7]
[1,2,3,4] Students of Dept of ISE, S.E.A College of Engineering and Technology, Bangalore, India
[5,6,7] Professors, of Dept of ISE, S.E.A College of Engineering and Technology, Bangalore, India

ABSTRACT - *Human activity recognition is a System Recognizing human behavior which is demanding and ongoing research area in computer science, particularly in video surveillance.for surveillance and healthcare monitoring both use Human Activity Recognition (HAR), which is an important task. The automation of this process has shown encouraging results in recent years when advances in Machine Learning (ML) and Computer Vision approaches are combined. The several machine learning methods and computer vision techniques used in HAR systems are reviewed and analyzed in detail in this work. We examine the difficulties encountered in this field, including shifts in viewpoint, obstruction, and the variety of human activity, and we talk about the various ways these difficulties are addressed. In addition, we highlight the most important datasets used in assessment and training. We analyze the efficacy of various machine learning models in precisely categorizing human actions from sensor data or video feeds through a thorough empirical assessment. What our results show is how effective. Even though automatic feature extraction solves the issues with manual features, current methods still require improvement in terms of accuracy and efficiency. The goal of this research is to raise the accuracy and efficiency of HAR systems. The HAR system shown in this study uses data enhancement techniques to capture a set of robust and discriminative features from each activity occurrence across 45 activities.*

Date of Submission: 01-05-2024 Date of acceptance: 12-05-2024

I. INTRODUCTION:

The practice of automatically recognizing actions and behaviors based on sensor data is known as human activity recognition (HAR) [1]. The field of HAR research is expanding quickly because of its applicability in several fields, including as healthcare [2], sports [3], robotics [4], and security [5], as illustrated in Figure 1. The goal of the HAR system is to design and create reliable, accurate algorithms that, in any situation or setting, can precisely identify and recognize human activity. The ability of HAR to detect monitor human activity in real-time is crucial since it can yield insightful information and enable a variety of applications. For example, in the healthcare industry, the HAR system can be used to track the movements. Following are the contributions of this study.

- Captured robust and relevant features for HAR
- The data enhancement methods are employed to enhance the size of training data.
- A transformer model was employed to increase the identification rate of human activities and lessen the training time of model under limited computational resources

II. LITERATURE REVIEW:

In this research, we apply four hybrid deep learning models to the HAR problem. Each hybrid model combines a CNN and a variation of RNNs. A well-known and publicly available dataset (PAMAP2) is utilized to assess the performance of the suggested hybrid models. The analysis findings show that each model has a high level of accuracy, which is higher than that produced by employing CNNs or RNNs alone. Other variables used to evaluate categorization outcomes include accuracy, precision, recall, F-score, sensitivity, and specificity. Overall, the results show that models with bidirectional RNNs outperform those with unidirectional RNNs. This result is reasonable because in the former, the data are processed both[1].

The main methods for recognising human activity from 3D data are compiled in this study, with an emphasis on methods that make use of depth data. The use of various features leads to the identification of broad categories for algorithms. The advantages and disadvantages of the algorithms in each category are examined, and a potential course for further research is suggested. We can now easily access 3D data to supplement typical RGB pictures, thanks to the recent advancements in range sensing technologies. [2].

In this research, we look at how time series data augmentation can improve the accuracy of different deep learning models using human activity data collected from mobile phone accelerometers. We evaluate the performance of the Vanilla, Long-Short Term Memory, and Gated Recurrent Units neural network models using three open-source datasets. We investigate the impact of two time series data augmentation approaches on target model accuracy. [3]

This research demonstrates that deep activity recognition models (a) improve recognition accuracy of human actions, (b) avoid the costly production of handcrafted features in previous systems, and (c) use huge unlabeled acceleration data for unsupervised feature extraction. Furthermore, a hybrid strategy combining deep learning and hidden Markov models (DL-HMM) is given for sequential activity identification. This hybrid method combines the hierarchical representations of deep activity recognition models with the stochastic modeling of temporal sequences in hidden Markov models..[4]

III. EXISTING SYSTEM & DRAWBACKS:

Existing systems for human activity recognition (HAR) using machine learning often employ traditional machine learning algorithms rather than deep learning techniques.

These systems have been successful in recognizing human activities based on various types of sensor data. Here are some key aspects of existing systems for HAR using machine learning:

1.Feature Extraction:

Extract relevant features from sensor data, such as accelerometer or gyroscope readings.

- Commonly used features include statistical measures (e.g., mean, standard deviation), frequency domain features (e.g., Fourier Transform), or time-domain features (e.g., autocorrelation).
- Domain knowledge and expert input are often used to select the most informative features.

2. Algorithm Selection:

- Choose a suitable machine learning algorithm to train a classifier on the extracted features.
- Popular algorithms include Support Vector Machines (SVM), k-Nearest Neighbors(k-NN), Random Forests, or Naive Bayes.
- The choice of algorithm depends on factors such as the nature of the data, the number of classes, and the desired computational efficiency.

<u>DRAWBACK OF EXITING SYSTEM:</u>

• The approach with Machine Learning takes longer time in computations.
• Efficiency of the algorithms are low compared to other transfer learning approaches.
• Images are encoded, decoding is less efficient.

IV. PROPOSED METHODOLOGY:

The architecture of proposed methodology isdepicted in Fig I:

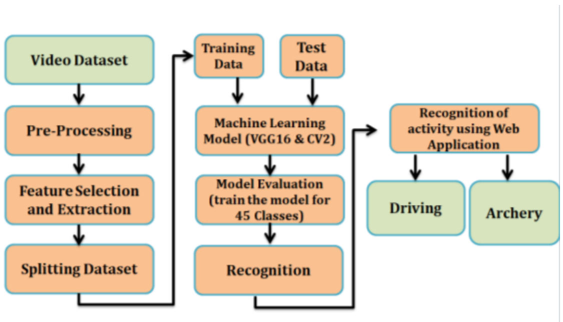

Fig 1: Architecture of proposed methodology

To perform human activity recognition using VGG16, ImageNet & OpenCV, you can follow a general methodology that involves the following steps:

1. Data collection:

Gather a dataset of videos or live video streams containing examples of different human activities that you want to recognize. These activities could include walking, running, sitting, standing, etc.

2. Pre-processing:

Pre-process the video frames to improve the quality of the input data. This may involve resizing the frames, converting The practice of automatically recognizing and categorizing human behaviors or activities using sensor data or video inputs is known as human activity recognition, or HAR. It is essential to many fields, including as surveillance systems, smart settings, fitness tracking, healthcare monitoring, and human-computer interaction. HAR systems use data analysis from sensors, including magnetometers, accelerometers, gyroscopes, and cameras, to infer human activities.them to grayscale, applying filters for noise reduction, or performing other image processing techniques.

3. Feature extraction:

Extract meaningful features from each video frame or a sequence of frames that can capture the characteristics of different activities.

4. Activity recognition model:

Train a machine learning or deep learning model using the extracted features and corresponding activity labels. Depending on the complexity of your task, here we are using VGG16, ImageNet model and OpenCV. Consider splitting your dataset into training and testing sets for model evaluation.

5. Model training:

Feed the extracted features and corresponding activity labels into your chosen model. Adjust the model parameters and optimize the training process to achieve good performance.

6. Activity prediction:

Once the model is trained, use it to predict the activities in new video frames or sequences. Apply the same pre-processing and feature extraction steps as before to the input frames. Pass the extracted features through the trained model and obtain the predicted activity labels.

VGG16:

*A convolutional neural network is also known as a ConvNet, which is a kind of artificial neural network. A convolutional neural network has an input layer, an output layer, and various hidden layers. VGG16 is a type of C*NN (Convolutional Neural Network) that is considered to be one of the best computer vision models to date. The creators of this model evaluated the networks and increased the depth using an architecture with very small (3 × 3) convolution filters, which showed a significant improvement on the prior-art configurations. They pushed the depth to 16–19 weight layers making it approx — 138 trainable parameters.

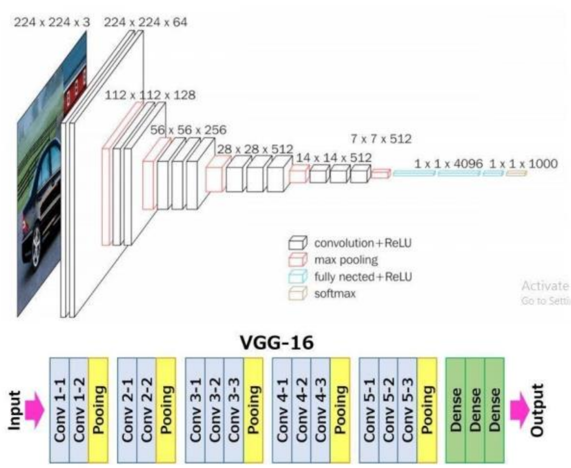

Figure 2: VGG16 Architecture

Open-CV:

OpenCV was built for maximum efficiency and performance of computing-intensive vision tasks. Therefore, it has a strong focus on real-time applications of AI vision. The software is written in optimized C and is able to take advantage of multicore processors (multi-threading). The goal of OpenCV is to provide an easy-to-use computer vision infrastructure that helps people build sophisticated vision applications quickly by providing over 500 functions that span many areas in vision. OpenCV is often used in factory product inspection, medical imaging, security analysis, human-machine interface, camera calibration, stereo vision (3D vision), and robotic vision. The comprehensive image processing capabilities support video stream processing, image stitching (combining multiple cameras), camera calibration, and diverse image pre-processing tasks. Because machine learning is essential in computer vision, OpenCV contains a complete, general-purpose ML Library focused on statistical pattern recognition.

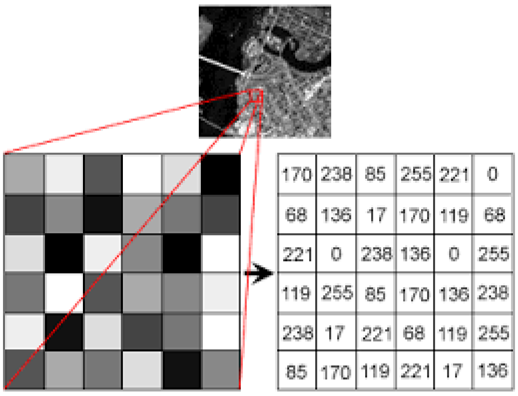

Figure 3: Working of Open CV

<u>FLASK:</u>

Flask is a popular web framework for building web applications and APIs using the Python programming language. It is a lightweight and flexible framework that provides the necessary tools and libraries to develop web applications quickly and efficiently.

V. PROJECT OVERVIEW & PROJECT LIMITATIONS:

<u>PROJECT OVERVIEW:</u>

The practice of automatically recognizing and categorizing human behaviors or activities using sensor data or video inputs is known as human activity recognition, or HAR. It is essential to many fields, including as surveillance systems, smart settings, fitness tracking, healthcare monitoring, and human-computer interaction. HAR systems use data analysis from gyroscopes, and cameras, to infer human activities.

Human Activity Recognition, or HAR, is a challenging time series classification task. It involves predicting the movement of a person based on sensor data and traditionally involves deep domain expertise and methods from signal processing to correctly engineer features from the raw data in order to fit a machine learning model. Recently, deep learning methods such as convolutional neural networks and recurrent neural networks have shown capable and even achieve state-of-the-art results by automatically learning features from the raw sensor data. Activity recognition is the problem of predicting the movement of a person, often indoors, based on sensor data, such as an accelerometer in a smartphone. Streams of sensor data are often split into subs-sequences called windows, and each window is associated with a broader activity, called a sliding window approach. Convolutional neural networks and long short-term memory networks, and perhaps both together, are best suited to learning features from raw sensor data and predicting the associated movement.

<u>PROJECT LIMITATIONS:</u>

- **Variability in Human Movements:**

Human actions can vary greatly in terms of speed, amplitude, duration, and style. HAR systems may struggle to generalize across different individuals, demographics, and cultural contexts, leading to reduced accuracy in recognizing diverse activities.

- **Limited Environmental Adaptability:** HAR systems may perform well in controlled environments but struggle to adapt to real-world settings with complex backgrounds, varying lighting conditions, and environmental clutter. Changes in ambient conditions such as noise, weather, or occlusion can degrade recognition performance.

- **Data Imbalance and Bias:**

HAR datasets often suffer from class imbalance, where certain activities are overrepresented while others are underrepresented. This imbalance can lead to biased models that perform well on dominant activities but poorly on minority classes. Additionally, biases in data collection, annotation, or labeling may affect the generalization ability of HAR systems.

- **Sensor Placement and Wear ability:**

The placement of sensors or wearable devices can impact the quality and reliability of data captured for activity recognition. Inconvenient or uncomfortable sensor placement may discourage user adoption, leading to incomplete or unreliable data.

- **Real-time Processing Constraints:**

HAR systems deployed in real-time applications require efficient algorithms capable of processing sensor data or video streams in near real-time. Computational limitations, especially on resource-constrained devices such as smartphones or embedded systems, may restrict the complexity and accuracy of HAR models.

- **Privacy and Ethical Considerations:**

HAR systems raise concerns related to user privacy, data security, and ethical implications. Collecting and analyzing sensitive personal data such as movement patterns, routines, or health-related activities requires careful handling to ensure user consent, data anonymization, and protection against misuse or unauthorized access.

VI. FUTURE SCOPE:

Implementing human activity recognition using VGG16 and OpenCV can offer several business benefits across various industries:

1.Healthcare and Wellness:
In healthcare, this technology can monitor patient activities, aiding in eldercare or patient rehabilitation. It can detect falls or irregular movements, ensuring timely assistance.

2.Fitness and Sports:
Fitness tracking applications can use this technology to analyze workout routines, provide personalized feedback, and track activities accurately. Sports coaching can also benefit by analyzing players' movements.

3.Security and Surveillance:
Enhanced surveillance systems can detect suspicious behaviors or unauthorized activities, improving security measures in public spaces, airports, or high-security areas.

4.Retail and Customer Insights: Analyzing customer movements in stores can help retailers understand buying patterns, optimize store layouts, and offer personalized shopping experiences.

5.Manufacturing and Quality Control: Monitoring worker activities on the factory floor can enhance safety protocols and ensure adherence to standard operating procedures. It can also identify inefficiencies in workflows.

6.Human-Computer Interaction:
Implementing gesture recognition through this technology can improve human-computer interaction in gaming, virtual reality, and other interactive systems.

7.Automotive and Transportation:
In driver assistance systems, recognizing driver activities can contribute to safer driving experiences by alerting or adapting to driver behaviors.

8. Education and Training:
Training scenarios, such as simulations or e-learning platforms, can benefit from recognizing and assessing learner activities, providing personalized feedback.

9.Marketing and Advertisement:
Analyzing consumer activities in response to marketing campaigns or advertisements can refine marketing strategies and optimize ad placements.

10. Efficiency and Process Optimization: Understanding human activities within various processes can optimize workflows, reduce inefficiencies, and enhance overall operational efficiency.

These applications demonstrate the wide range of potential business benefits that human activity recognition using VGG16 and OpenCV can offer. The ability to accurately detect and analyze human activities opens doors to improved safety, efficiency, and tailored experiences across multiple industries.

VI. SYSTEM DESIGN:

FLOW CHART:

Figure 4: Flowchart

DATA FLOW DIAGRAM:

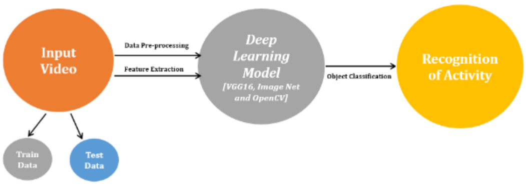

Figure 5: Data Flow Diagram

VII. FINDINGS / RESULTS OF ANALYSIS:

Figure 6: HAR Home Page

This is the home page for HAR developed using Flask.

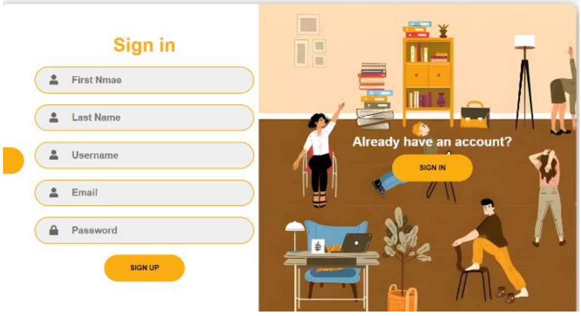

Figure 7: Registration Page

The User must register their details in order to create an account for themselves.

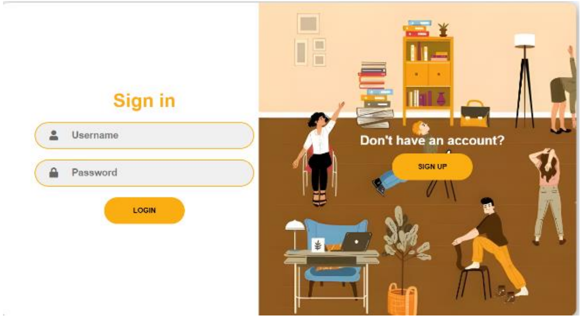

Figure 8: Login Page

The user must then proceed forward and login to their account by entering the username and password of their account.

Figure 9: Upload Video

User should enter the path/URL for prediction.

Figure 10: Video Processed

Once the video is processed it will randomly generate the result with a unique avi number.

Figure 11: Prediction Result 1

Figure 12: Prediction Result 2

Figure 13: Prediction Result 3

VII. CONCLUSION:

The project focused on developing a system for human activity recognition using OpenCV and leveraging the VGG16 & ImageNet16 model. The application aimed to recognize and analyze human activities in real-time, particularly in 45 Classes performance scenarios are created. The system utilized OpenCV for video processing, feature extraction, and activity recognition, while the ImageNet16 model provided pre-trained capabilities for classifying activities.

By combining OpenCV and the ImageNet16 model, the project aimed to harness the power of computer vision and machine learning for accurate and efficient activity recognition. Overall, the project sought to demonstrate the practical application of human activity recognition using OpenCV and the integration of the ImageNet model, highlighting the potential benefits in sports performance analysis and coaching. The combination of these technologies allowed for real-time analysis, personalized feedback, and data visualization, enabling different activity and to make informed decisions and track progress over time.

REFERENCES:

[1]. Abbaspour S, Fotouhi F, Sedaghatbaf A, Fotouhi H, Vahabi M, Linden M (2020) A comparative analysis of hybrid deeplearning models for human activity recognition. Sensors20(19):5707

[2]. Aggarwal JK, Xia L (2014) Human activity recognition from 3Ddata: a review. Pattern Recogn Lett 48:70– 80

[3]. Alakwaa W, Nassef M, Badr A (2017) Lung cancer detectionand classification with 3D convolutional neural network (3D-CNN). Lung Cancer 8(8):409

[4]. Alawneh L, Alsarhan T, Al-Zinati M, Al-Ayyoub M, JararwehY, Hongtao L (2021) Enhancing human activity recognitionusing deep learning and time series augmented data. J AmbientIntell Humaniz Comput 12(12):10565–10580

[5]. Almaslukh B, AlMuhtadi J, Artoli A (2017) An effective deep autoencoder approach for online smartphone-based human activity recognition. Int J Comput Sci Netw Secur 17(4):160–165 Alom MZ, Taha TM, Yakopcic C, Westberg S, Sidike P, Nasrin MS,

Hasan M, Van Essen BC, Awwal AAS, Asari VK (2019) A state-of-the-art survey on deep learning theory and architectures. Electronics 8(3):292

[6]. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[7]. J.Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[8]. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[9]. K. Elissa, “Title of paper if known,” unpublished.

[10]. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[11]. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].