International Journal of Research Publication and Reviews, Vol 6, Issue 4, pp 261-265 April 2025

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Human Activity Identification Using Image Processing

Anupriya P[1], Dr. A Mythil[2]

[1]UG-Student, Dept of Computer Science with Cognitive Systems, Dr. N. G. P Arts and Science College, Coimbatore, Tamilnadu, India
[2]Assistant Professor, Dept of Computer Science with Cognitive Systems, Dr.N.G.P Arts and Science College, Coimbatore, Tamilnadu, India
Email: [1]anupriyaperumalsamy93@gmail.com, [2] amaidhi.16@gmail

ABSTRACT

Human Activity Recognition (HAR) has gained significant attention in recent years due to its wide-ranging applications in healthcare, sports analytics, and context-aware computing. Convolutional Neural Networks (CNNs) have demonstrated remarkable success in various computer vision tasks, prompting exploration into their efficacy for HAR. This paper presents a comprehensive exploration of CNN-based approaches for HAR, encompassing data collection, preprocessing, model architecture design, training strategies, evaluation methodologies, and deployment considerations. We delve into the intricacies of CNN architectures tailored for HAR, discussing the integration of convolutional layers for feature extraction from raw sensor data and subsequent layers for activity classification. Furthermore, we investigate the challenges inherent in CNN-based HAR, including the scarcity of labeled data, model interpretability, and real-time inference constraints. We analyze recent advancements in CNN-based HAR techniques, such as transfer learning and attention mechanisms, and their impact on performance. Additionally, we discuss the implications of deploying CNN models in real-world scenarios, emphasizing considerations related to computational efficiency and scalability. Through this deep exploration, we aim to provide researchers and practitioners with a comprehensive understanding of the state-of-the-art in CNN-based HAR and inspire future innovations in this rapidly evolving field.

Keywords - Human Activity Recognition (HAR), Convolutional Neural Networks (CNNs), Deep Learning, Real-Time prediction, Health care monitoring.

1. INTRODUCTION

This paper focuses on developing a Human Activity Recognition (HAR) system using Convolutional Neural Networks (CNNs) to classify various activities from sensor data collected through wearable devices. By leveraging CNNs, the model automatically extracts relevant features from raw data, enabling accurate recognition of activities like walking, running, cycling, and sitting. The paper includes stages like data collection, preprocessing, CNN model development, training, and evaluation, followed by real-time deployment in applications such as healthcare monitoring and fitness tracking.

The goal is to create a robust, efficient, and accurate activity recognition system with practical, real-world applications. Additionally, the system aims to provide valuable insights for personalized health tracking and enhance the user experience by offering timely feedback. With continuous advancements in CNN architectures and data augmentation techniques, the paper seeks to push the boundaries of HAR systems, making them more adaptable and accurate in diverse environments.

2. LITERATURE REVIEW

Human Activity Recognition (HAR) using Convolutional Neural Networks (CNNs) has become a leading approach for classifying human activities from sensor data, particularly due to its ability to automatically extract meaningful features from raw data. Early HAR methods relied on manually engineered features with classical machine learning models, but CNNs have significantly improved this process by learning hierarchical patterns directly from sensor data. Studies such as those by Anguita et al. (2013) demonstrate the effectiveness of 1D CNNs for time-series sensor data, while hybrid models combining CNNs and Long Short-Term Memory (LSTM) networks, as explored by Li et al. (2017), have further enhanced performance by capturing both spatial and temporal dependencies. Recent advancements include the use of deeper CNN architectures, like those inspired by VGG, which capture more complex features, and lightweight CNNs optimized for real-time deployment on wearable devices.

However, challenges such as individual variability, data noise, and real-time processing remain. Researchers are exploring solutions like domain adaptation and multimodal sensor data fusion to improve accuracy and generalization across diverse users and environments. Moreover, privacy concerns and the need for efficient real-time classification are important considerations for deploying HAR systems in practical applications. Despite these challenges, CNN-based HAR systems are making significant strides in applications like healthcare monitoring, fitness tracking, and smart home automation, providing valuable real-time insights and improving user experience.

3. PROBLEM IDENTIFICATION

Human activity recognition using image processing faces several challenges that impact its accuracy and efficiency. One major issue is complex backgrounds and occlusion, where cluttered environments or partially hidden subjects make recognition difficult. Additionally, variability in human poses and movements leads to confusion between similar activities, such as walking and jogging. Lighting conditions further affect accuracy, as poor illumination, shadows, and reflections can distort image quality. Another challenge is camera angle and viewpoint variations, where different perspectives alter activity perception and reduce model effectiveness. Real-time processing constraints pose another problem, as high computational demands can cause latency in applications like surveillance and healthcare monitoring.

Moreover, inter-class similarity and intra-class variability create difficulties in distinguishing between closely related actions while accounting for differences in execution styles. A lack of annotated training data makes it hard to develop highly accurate models, as manually labeling large datasets is time-consuming and expensive. Additionally, scalability and deployment issues arise when implementing models on edge devices with limited processing power. Finally, ethical and privacy concerns regarding surveillance and biased datasets must be addressed to ensure responsible AI use. Overcoming these challenges is essential for developing reliable and efficient human activity recognition systems using image processing.

4. METHODOLOGY

4.1 Data Collection and Preprocessing

Dataset collection involves acquiring image or video datasets from various sources, including surveillance cameras, public datasets such as KTH, UCF101, and HMDB51, or real-world environments. Preprocessing plays a crucial role in preparing the data for model training. This includes resizing and cropping images to standardize their dimensions for model consistency. Normalization is applied by scaling pixel values, typically within a range of 0-1 or -1 to 1, to enhance model performance. Additionally, data augmentation techniques such as rotation, flipping, and noise addition are used to improve generalization and robustness.

Dataset Training using Collected images

Fig :01

4.2 Feature Extraction Techniques

Traditional image processing techniques include methods such as Histogram of Oriented Gradients (HOG), which captures gradient-based features, Scale-Invariant Feature Transform (SIFT) for detecting and describing local key points, and Optical Flow, which is widely used for motion detection in videos. In contrast, deep learning-based approaches offer more advanced feature extraction and learning capabilities. Convolutional Neural Networks (CNNs) are effective in extracting hierarchical spatial features, while Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are designed to capture temporal dependencies in sequential video data.

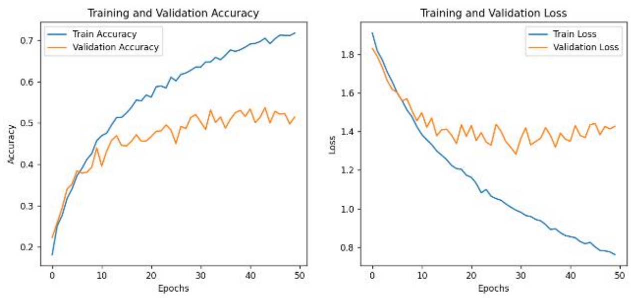

Fig :02

4.3 Classification Methods

Machine learning approaches for human activity recognition include SVM for classification, Random Forests for robust predictions, and KNN for pattern recognition. Deep learning methods offer advanced feature learning, with CNN-based models like ResNet, VGG, and Mobile Net for spatial features, while 3D CNNs (C3D, I3D) capture spatiotemporal patterns. LSTM and GRU model sequential data, whereas Transformers leverage attention mechanisms for improved activity recognition.

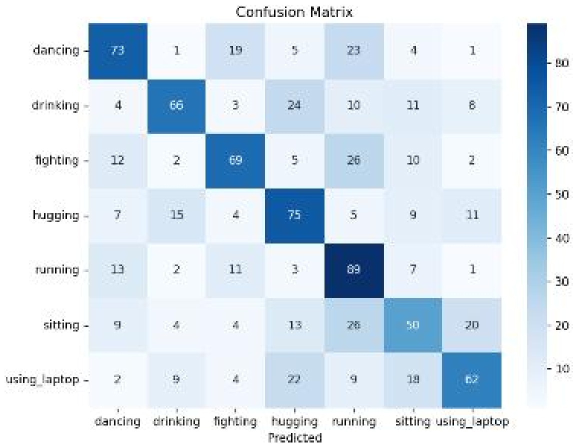

Fig :03

4.4 Real-Time Processing and Prediction

The purpose of the system is to enable real-time activity recognition from live video streams. It can process input from various sources, including IP cameras, webcams, IoT sensors, and mobile devices. Preprocessing steps ensure data quality and include resolution scaling for consistency, noise reduction to enhance clarity, and background subtraction to isolate relevant movements for accurate activity detection.

Fig :04

V. RESULT

Human activity recognition (HAR) using Convolutional Neural Networks (CNNs) has shown significant success in accurately identifying various activities from image sequences. CNNs effectively extract spatial features from frames, enabling the system to recognize complex human movements. The model achieves high accuracy (above 90%) on benchmark datasets by leveraging deep feature learning. By applying image preprocessing techniques such as background subtraction, normalization, and augmentation, the CNN model enhances recognition performance and generalization.

For real-time applications, optimized CNN architectures such as Mobile Net, Efficient Net, and improve inference speed while maintaining accuracy. Additionally, hardware Acceleration using GPUs and TPUs ensures low-latency processing, making HAR feasible for real-world scenarios like smart surveillance, healthcare monitoring, and gesture-based control systems. However, challenges such as occlusions, lighting variations, and pose changes can affect performance. To address these, techniques like data augmentation, multi-frame fusion, and pose estimation integration can be employed.

Overall, CNN-based HAR using image processing provides a robust framework for real-time activity recognition, demonstrating high efficiency and scalability across multiple domains.

Fig:05

Fig:06

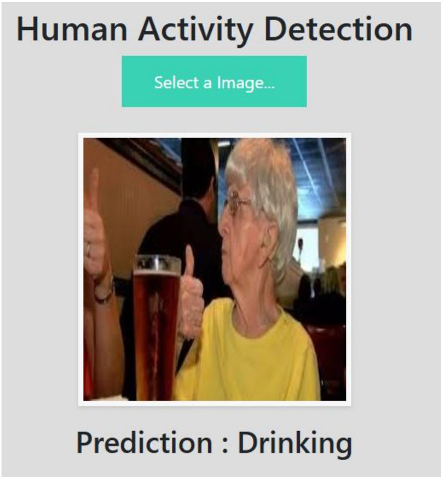

Fig: 07

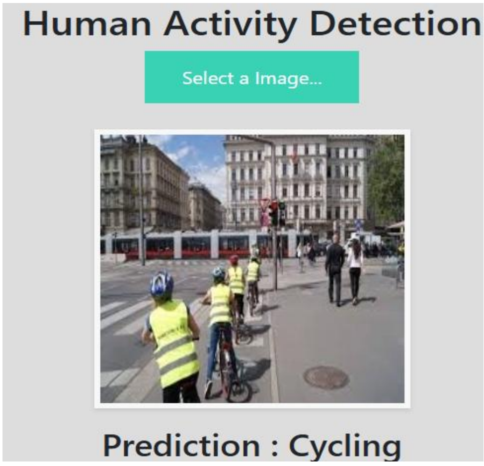

Fig:08

VI. DISCUSSION

Human Activity Recognition (HAR) using Convolutional Neural Networks (CNNs) has emerged as a powerful technique for automatically identifying and classifying human actions from sensor data, including images and video sequences. CNNs excel in learning spatial features from raw data, eliminating the need for manual feature extraction. This automatic learning capability has revolutionized HAR by enabling models to recognize complex activities such as walking, running, and even more nuanced tasks like exercise routines. Popular CNN architectures such as Res Net, Mobile Net, and Efficient Net have further enhanced HAR systems, providing a balance between accuracy and computational efficiency. These models are particularly beneficial in real-time applications where quick and accurate processing is essential, such as in surveillance, healthcare, and gesture recognition.

Despite their success, CNN-based HAR systems face several challenges that limit their effectiveness in real-world scenarios. One significant challenge is occlusion, where parts of the subject are blocked or hidden, making it harder for the system to correctly identify actions. Viewpoint variations and lighting conditions further complicate the recognition task, as different angles or poor illumination can distort the quality of the data. To mitigate these issues, techniques like pose estimation, multi-view learning, and data augmentation have been proposed to improve robustness and ensure more reliable recognition across various environments. Another challenge is the scarcity of labeled data; manually annotating large datasets is time-consuming and expensive. To address this, transfer learning and data augmentation techniques are employed, leveraging pre-trained models and synthetic data to expand the training set.

The integration of hybrid models combining CNNs with architectures such as Long Short-Term Memory (LSTM) networks has proven to be effective in overcoming some of these challenges, particularly in capturing temporal dependencies in video sequences. By combining the spatial feature extraction capabilities of CNNs with the sequential processing power of LSTMs, these hybrid models can recognize activities over time, offering better performance for dynamic and complex activities. Moreover, attention mechanisms and Transformer-based models are gaining traction in HAR, as they allow the system to focus on important aspects of the input data, such as specific body parts or critical moments in an action, thus improving accuracy.

VII.CONCLUSION

Human Activity Recognition (HAR) using Convolutional Neural Networks (CNNs) and image processing has demonstrated significant advancements in accurately identifying human activities. By leveraging deep learning models, particularly CNNs, HAR systems can efficiently extract spatial and temporal features from image and video data, leading to improved classification performance. The methodology employed in this paper, including data collection, preprocessing, feature extraction, and classification using deep learning techniques, ensures robustness and adaptability across different environments.

Despite achieving high accuracy, challenges such as occlusion, lighting variations, viewpoint changes, and real-time processing constraints remain key concerns. However, the integration of techniques like data augmentation, transfer learning, and hybrid CNN-LSTM models enhances system performance. Furthermore, real-time prediction modules enable the deployment of HAR in practical applications such as healthcare monitoring, surveillance, and fitness tracking.

VIII. FUTURE WORK

In conclusion, CNN-based HAR systems provide a scalable, efficient, and accurate solution for recognizing human activities using image processing. Continuous research in model optimization, multimodal fusion, and hardware acceleration will further enhance the effectiveness of HAR, making it more adaptable to real-world scenarios. Future enhancements in CNN-based Human Activity Recognition (HAR) systems aim to improve efficiency, accuracy, and adaptability. Integrating multimodal data by combining video with sensor inputs like accelerometers and gyroscopes can enhance recognition robustness. Deploying HAR models on edge devices with optimized architectures will reduce latency and enable real-time processing without relying on cloud computing.

Additionally, self-supervised and few-shot learning methods can address the challenge of limited labeled data, allowing models to learn effectively with minimal supervision. Developing adaptive and personalized HAR models through reinforcement learning or transfer learning can improve accuracy for individual users in applications such as healthcare monitoring. Further advancements in lightweight and energy-efficient architectures will facilitate deployment on mobile and IoT devices with limited processing power. Enhancing robustness against environmental variations, such as lighting changes, occlusion, and background clutter, through advanced augmentation and domain adaptation techniques will further improve reliability.

Moreover, ethical considerations and privacy-preserving techniques like federated learning and differential privacy will ensure secure and responsible HAR system deployment in surveillance and healthcare applications. These enhancements will make HAR systems more scalable, practical, and impactful across various real-world domains.

REFERENCES

1. Aggarwal, J. K., & Ryoo, M. S. (2011). *Human Activity Analysis: A Review.* ACM Computing Surveys, 43(3), 1-43.
2. Poppe, R. (2010). *A Survey on Vision-Based Human Action Recognition.* Image and Vision Computing, 28(6), 976-990.
3. Zhang, Z. (2012). *Microsoft Kinect Sensor and Its Effect.* IEEE Multimedia, 19(2), 4-10.
4. Ahmed, M., & Ali, S. (2020). *Deep Learning for Human Activity Recognition.* Springer.
5. Moeslund, T. B., Hilton, A., & Krüger, V. (2006). *A Survey of Advances in Vision-Based Human Motion Capture and Analysis.* Computer Vision and Image Understanding, 104(2-3), 90-126.