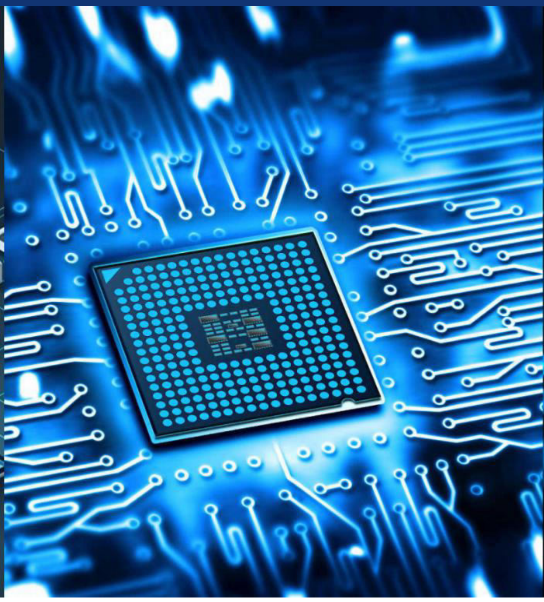
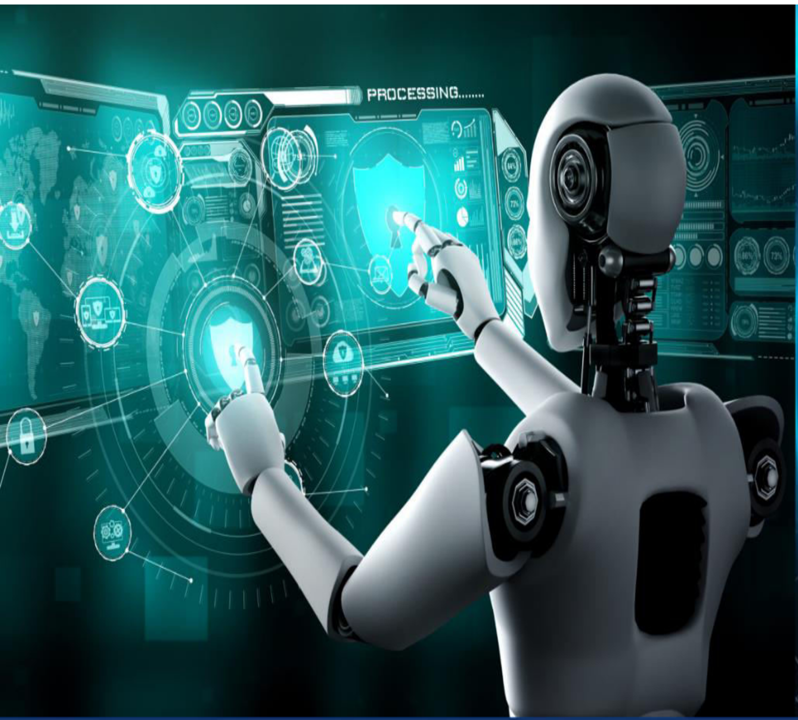
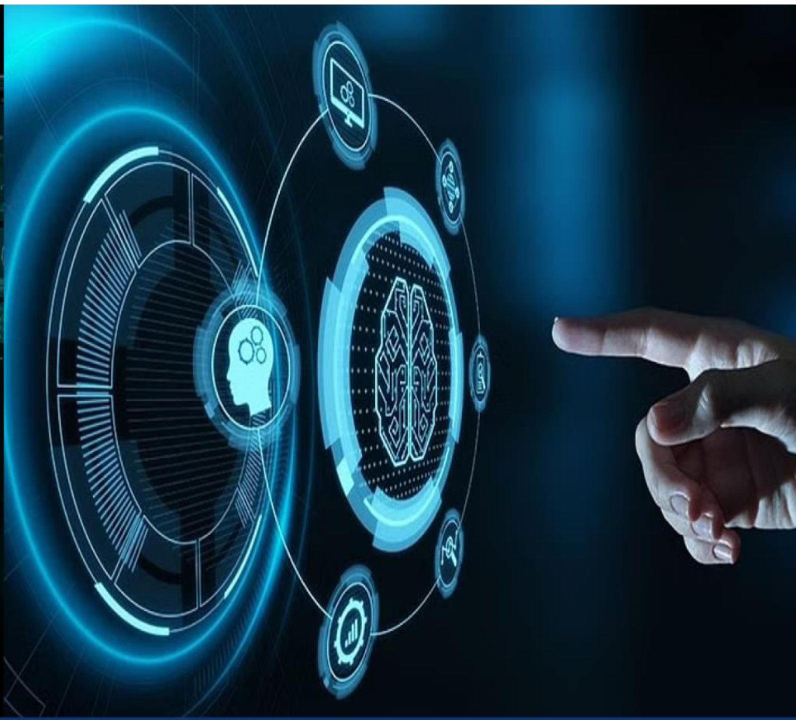
ISSN(O): 2320-9801
ISSN(P): 2320-9798


International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 8.625

Volume 13, Issue 1, January 2025

© 2025 IJIRCCE | Volume 13, Issue 1, January 2025| DOI: 10.15680/IJIRCCE.2025.1301145

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

A Review on Advancements in Deep Learning for Human Detection, Activity Recognition and Real-time Application

Swetha D, Vyshnavi N, Amaan Muhzin, M S Harinandan, Hima Vijayan V P

Department of ECE, ACE College of Engineering, Trivandrum, India

Department of ECE, ACE College of Engineering, Trivandrum, India

Department of ECE, ACE College of Engineering, Trivandrum, India

Department of ECE, ACE College of Engineering, Trivandrum, India

Department of ECE, ACE College of Engineering, Trivandrum, India

ABSTRACT: This review paper discusses the developments in Human Activity Recognition (HAR) and its contribution to smart city safety and efficiency with new applications, e.g., video surveillance. HAR systems enable real-time HMD detection, which is critical for public safety and resource control and thus enhances energy conservation in smart homes and industrial applications. The most important datasets for training HAR models are classified into sensor-based, video-based, image-based, and depth/crowd-based data types, with different degrees of accuracy.the most More precisely, sensor-derived datasets reach accuracy between 85% and 95%, and video-based datasets can surpass 95%. The paper relates how strong the role of deep learning techniques has played in HAR and also argues that future research should be concentrated on the fusion of multiple kinds of data sources to further improve HAR systems. A concise review of HAR and its importance in promoting sustainable resource exploitation and automation has been provided in this work across fields such as healthcare, education, and safety.

KEY WORDS: deep learning, human activity recognition, crowd density analysis, and real-time motion detection.

I. INTRODUCTION

Video surveillance is used mainly in smart cities to improve quality of life and security. Surveillance cameras are installed in schools, colleges, traffic signals, malls, public buildings, etc. Installing human activity recognition in these surveillance cameras will help provide security and monitor public services. The real-time human detection and surveillance system is used for crowd computing, irregularity detection, and improved monitoring of public places [1], [2].

Since HAR is a contactless motion detection healthcare system, it is used for patient monitoring and improving health-care solutions [3]. From an academic point of view, attention- monitoring systems are used to assess student engagement and behavior in the classroom [4]. Utilizing human presence detection for automation and improvement of safety [1], [4]

in smart home and industrial buildings will result in energy efficiency. For military purposes, aerial surveillance is done by unmanned aerial vehicles.

The use of conventional neural networks (CNNs) and en- semble learning has applications in entertainment and sports [5], [6]. The techniques and the development of models have improved real-motion recognition in video systems [7], [8], contributing to breakdowns in artificial intelligence research [2], [9].

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

II. MOTIVATION

Human activity recognition (HAR) has several advantages for the environment, promoting sustainability and safety and enhancing energy efficiency across various sectors. One of the principal advantages of implementing recognition of human activity in smart homes and buildings is energy efficiency. HAR systems can monitor human presence and activity and adjust the environment accordingly. For example, adjusting heating and cooling systems according to the number of people in the building, which accordingly saves energy when spaces are unoccupied [1], [10]. In smart cities, HAR monitors public services like waste management and traffic control based on real-time human data, which reduces resource waste [11] and congestion. In an industrial environment, HAR can ensure a safer work environment by detecting unsafe behaviors or conditions and reducing accidents. Automating tasks using HAR based on human activity leads to a more sustainable and environmentally conscious future.

III. STATISTICS

From the papers we have reviewed, the datasets used play a major role in the accuracy of the system. The publicly available datasets, like UCI HAR, Kinetics, NTU RGB+D, and HMDB51, dominate academic studies and research. The main data sets used here can be classified into sensor-based, video- based, image-based, and specialized types, including depth- and crowd-based data sets.

A. Sensor-Based Datasets

Sensor-based datasets such as UCI HAR, PAMAP2, and HAR benchmarks detect simple physical activities such as walking, running, and lying down. The UCI HAR dataset used in [2], [6], [12] collects data via accelerometers and gyroscopes. The PAMAP2 data set used in [2], [6] and the HAR reference data set in [12], [13] collect data from wearable sensors. These wearable sensors collect real-time data by monitoring physical activities to create datasets and perform accurate measurements. According to the papers [2], [6], [12], [13]. Using sensors to collect datasets has been reported to have an accuracy of 85% to 95%

B. Video-Based Datasets

Video-based datasets like Kinetics used in [9], [12], [14], Charades [9], [14], HMDB51 [9], [14], NTU RGB+D mentioned in [14], and ActivityNet used in [9], [12] detect both basic and complex human activities, including cooking, danc- ing, etc. These datasets collect data from RGB cameras for general analysis, depth cameras used for precise motion and safety applications, and video cameras used for surveillance. According to papers [9], [12], [14], using video-based datasets to collect datasets has shown an accuracy above 95%.

C. Image-Based Datasets

Image-based datasets include MSCOCO used in [7], [15], VOC2012 referenced in [7], OpenImages used in [15], and custom UAV datasets mentioned in [15], [16]. These datasets are commonly used for object detection and application-specific tasks. The main focus of using these datasets is to improve precision using feature and fusion anomaly. Accord- ing to papers [7], [15], [16], using image-based datasets to collect datasets has shown an accuracy above 90%.

D. Depth and Crowd-Based Datasets

Depth and crowded-based data set major examples are DPDNet referenced in [16] and ShanghaiTech [11]. These datasets focus on safety, crowd analysis, and motion detec- tion in real-time environments. In articles [11], [16], these datasets are used for presence detection and positioning. In the article [11], ShanghaiTech is used for crowd computing and distribution, which would help in preventing stampedes and fatalities due to overcrowding. Using these datasets is reported to have accuracy ranges between 88% and 93%, especially with specialized datasets tailored for real-world environments.

IV. RELATED WORKS

This series of studies investigates new applications of deep learning for human detection, activity recognition, and real-time tracking[[2], [7], [8], [10], [15], [17], [18]. Several studies have highlighted the need to improve robustness and accuracy in human detection in adverse conditions, namely in crowded areas or dynamic situations [14]. Methods including multi-scale feature fusion, ensemble learning, and CNN ar- chitecture modifications were presented to

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

optimize real-time detection, motion recognition, and overall system robustness. Further contributions consist of the use of depth cameras in applications like specific altitude tracking in industrial safety as well as the assessment of deep learning models in motion detection in healthcare to guarantee the quality and reliability of their performance.

Active research was devoted to the enhancement of ac- tivity recognition using deep learning approaches, including studies on problems of dataset variability, scalability, and generalizability. Vision Transformers (ConViT) were stud- ied concerning their application in still-image analysis for action recognition [11], and multimodal AI-based concepts were explored to overcome these limitations and increase accuracy. Activity recognition trends, limitations, and future directions obtained from benchmark datasets were discussed and compared with CNN and the other deep learning methods [10]. Examples such as Wi-Fi signal-based Sensor presence detection and resource-constrained IoT models [17] highlight the breadth of applicability of deep learning in addressing practical real-world problems.

Apart from human activity recognition, the work also inves- tigated novel systems, for instance, human detection on a UAV in a dynamic setting, as well as student attention monitoring systems. Surveillance video anomaly detection architectures and crowd-counting models were surveyed [18], with consid- eration of the limitations and areas for refinement. In general, this series of studies highlights the promise of deep learning to provide human detection and activity recognition capabilities in a wide range of application areas, from video surveillance and healthcare to the Internet of Things and industrial safety, and points out important challenges and future directions of research.

V. METHODOLOGIES

A. Deep Learning for Human Detection and Presence Recog- nition

Deep learning has enhanced human detection and presence recognition, enabling a wide range of applications across vari- ous scenarios. Studies such as DPDnet, Deep Learning-Based Human Presence Detection, Real-Time Human Detection by UAVs, and Human Position Detection [1], [2], [9], [11], [15], [16] depict the effectiveness of deep learning in detecting in- dividuals and recognizing their presence. Deep learning plays a prominent role in gesture recognition. These studies help in enhancing various technologies, including overhead depth cameras, neural networks, images, and drones, to enhance ro- bustness, showcase real-time capabilities, and ensure safety in mechanical environments, introducing innovative applications in surveillance, monitoring, smart homes, and healthcare.

B. Human Activity Recognition (HAR)

Human Activity Recognition (HAR) is an area of re- search that heavily depends on Convolutional Neural Networks

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

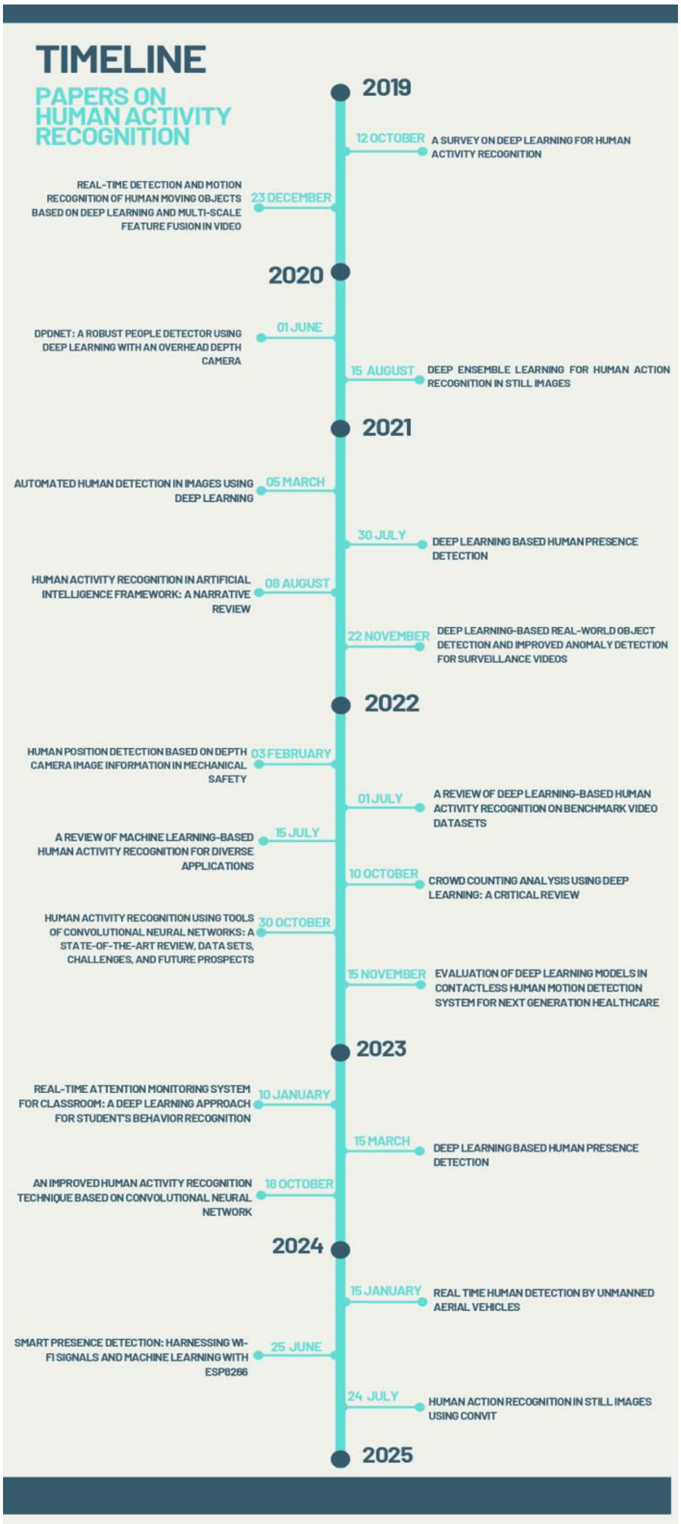


Fig. 1. Timeline of surveyed papers

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

(CNNs) and other deep learning techniques to identify and classify human activities accurately. The objective of HAR is to intensify the quality of life by making the device aware of users' needs. Surveys like A Survey on Deep Learning for Human Activity Recognition [2] provide valuable information on existing techniques and trends, while advanced CNN-based approaches, which are proposed in the paper An improved HAR technique [14]strengthen accuracy and expands appli- cation scope. Also, reviews like A Review of a Machine Learning-Based HAR [7] highlight the different applications of HAR, underscoring the importance of video datasets and broader AI integration to explore new possibilities and inno- vative applications in fields like healthcare, surveillance, and smart environments.

C. Ensemble and Multi-Model Approaches

These articles explore multimodal approaches for detection and recognition tasks. Ensemble learning and multi-model approaches enhance detection accuracy, particularly in action recognition and anomaly detection. Research such as Deep En- semble Learning for Human Action Recognition [5] describes the utilization of ensemble models to effectively analyze still images and combine multiple deep learning tools to improve generalization purposes. Similarly, deep learning-based real- world object detection [4], [7], [15] combines real-world detection with anomaly recognition, demonstrating innovative methods across various domains, including surveillance and safety.

D. Real-Time Detection and Motion Recognition

According to the papers [4], [7], [15], real-time motion detection and recognition research focuses on practical ap- plications in healthcare, education, and surveillance. Real- time motion detection and recognition have an important role in applications like target detection, recognition, posi- tion, tracking, and navigation. Techniques like multi-scale feature fusion enable efficient detection in videos. Behavior recognition systems like the Real-time Attention Monitoring System monitor student behavior in classroom environments. Contactless motion detection underscores the importance of being fast in critical real-world scenarios.

E. Applications in Surveillance, Safety, and Smart Environ- ments

Applications of deep learning innovations include surveil- lance, safety, and smart environments. Studies like crowd-counting analysis [11] describe various methods of crowd density analysis using video analytic datasets, and smart presence detection using hardware and IoT helps in detecting human presence. Safety-based approaches with depth cameras play a major role in ensuring mechanical and social safety. These studies show the impact of deep learning in a vast number of fields in addressing the key challenges for future developments.

VI. COMPARATIVE STUDY

A. Architecture

Some researchers (Article [16]) use DPDnet, and using overhead depth cameras, deep learning of human detection can be done. Real-time detection and motion recognition are used in research (Article [7]). Reviews of Deep Learning Techniques, especially Human Movement Detection and Hu- man Recognition, are provided by CNN ID (Articles [2], [6]). Deep ensemble learning for human recognition (Article [5]) explains ensemble learning with deep networks. Highlighting how networks can act upon human recognition in still images is depicted (Article [8]) as ConViT. An improved human activity recognition focuses on CNN improvements for activity recognition. Articles [1], [10], [11], [15] explore the detection of human presence and human motion recognition. Real-time and context-specific applications are in Articles [3], [4], and they explore real-time human detections and behavior analysis.

B. Focus of the articles

Human detection using depth cameras, UAVs, and real- world applications and Techniques like deep learning models for presence detection in safety or real-time environments are discussed in Human presence/detection (Article [1], [16], [18]). Human activity recognition contains survey and review articles analyzing deep learning and machine learning tech- niques for HAR (Articles [1], [2], [10]). Motion detection and behavior recognition refer to real-time motion detection and recognition in videos using multiscale features and deep learning (Articles [4], [15], [16]). Action recognition in images explores deep learning models such as ConViT to identify actions in still images and emphasizes ensemble models for robustness (Articles [5], [8]). Crowd counting and analysis explain the focus on crowd analysis using deep learning and applications in public safety and resource management (article [11]). Emerging technologies for detection innovatively use Wi-Fi signals and ESP8266 for presence detection (Article [10]).

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

C. Accuracy

Detection accuracy and human presence are handled and focused on precise human detection under varying conditions (Articles [1], [11], [15], [16]). Human activity recognition highlights the state of the techniques and the accuracy bench- mark for human resource recognition using CNNs, RNNs, or hybrid models (Articles [2], [6], [9], [12]–[14]). Crowd anal- ysis and non-visual methods explore alternative approaches like WiFi signals or models designed for crowd scenarios (Articles [10], [11]). Still, image recognition accuracy focuses on image-based action recognition using ensemble learning and transformer CNN hybrids (Articles [5], [8]). Behavior and motion recognition emphasizes motion recognition and behavior analysis in real-time surveillance or classroom set- tings (Articles [4], [7], [15]). Precision, recall, and F1score are commonly used to assess detection and classification tasks.

D. Description

Human activity recognition discusses the use of depth cameras, UVs, and WiFi signals . CNNs, transformers, and ensemble models are nowadays used for deep learning on human movement recognition in applications of videos, still images, and AI frameworks (Articles [2], [5], [6], [9], [12]– [14]).Crowd analysis and human movement analysis using deep learning for tasks like crowd counting and anomaly de- tection (Articles [11], [15]). Behavior and motion recognition mainly focus on the recognition of motion and behavioral pat- terns for real-time applications (Article [3], [4], [7]). Automa- tion and surveillance act on automation in detection systems and anomaly recognition for surveillance videos (Articles [15], [**?**]).

VII. CONCLUSION

HAR is a popular system in different domains like sports, healthcare, robotics, human-computer interaction, security, and surveillance. Various papers and surveys have been produced on the subject of human activity recognition, specifying different architectures, methodologies, and datasets showing different accuracy. This review paper [3], [16] presents an overview of the latest advancements in HAR, anchoring the applications in various fields. It explores various ranges of human activities through datasets obtained using sensors, videos, images, and specialized techniques. Out of the datasets, video-based datasets proved to have above 95% accuracy in the system. Sensor-based datasets and image-based datasets have been reported to have an accuracy range of 85% to 95%.These reviews [3], [4], [11] provides various approaches to human activity recognition, including CNN, YOLO, DPDnet, and ConVit applications in health care, safety, education, and en- vironmental sustainability, proving how human activity recog- nition can be used to ensure efficiency, safety, and ecological sustainability. These studies showcase the important role of artificial intelligence in advancing real-world solutions that can be a huge benefit for the industry and society at large.

REFERENCES

1. V. Balachandran, M. Shapiee, A. F. Ab. Nasir, A. Razman, and A. P P Abdul Majeed, “Deep learning based human presence detection,”MEKATRONIKA, vol. 2, pp. 55–61, 12 2020.
2. F. Gu, M.-H. Chung, M. Chignell, S. Valaee, B. Zhou, and X. Liu, “A survey on deep learning for human activity recognition,” ACM Computing Surveys (CSUR), vol. 54, no. 8, pp. 1–34, 2021.
3. Y. Song, W. Taylor, Y. Ge, M. Usman, M. A. Imran, and Q. H. Abbasi, “Evaluation of deep learning models in contactless human motion detection system for next generation healthcare,” Scientific Reports, vol. 12, no. 1, p. 21592, 2022.
4. Z. Trabelsi, F. Alnajjar, M. M. A. Parambil, M. Gochoo, and L. Ali, “Real-time attention monitoring system for classroom: A deep learning approach for student’s behavior recognition,” Big Data and Cognitive Computing, vol. 7, no. 1, p. 48, 2023.
5. X. Yu, Z. Zhang, L. Wu, W. Pang, H. Chen, Z. Yu, and B. Li, “Deep ensemble learning for human action recognition in still images,” Complexity, vol. 2020, no. 1, p. 9428612, 2020.
6. M. M. Islam, S. Nooruddin, F. Karray, and G. Muhammad, “Human activity recognition using tools of convolutional neural networks: A state of the art review, data sets, challenges, and future prospects,” Computers in biology and medicine, vol. 149, p. 106060, 2022.
7. M. Gong and Y. Shu, “Real-time detection and motion recognition of human moving objects based on deep learning and multi-scale feature fusion in video,” IEEE Access, vol. 8, pp. 25811–25822, 2020.
8. S. R. Hosseyni, S. Seyedin, and H. Taheri, “Human action recognition in still images using convit,” in 2024 32nd International Conference on Electrical Engineering (ICEE), pp. 1–7, 2024.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

9. V. Sharma, M. Gupta, A. K. Pandey, D. Mishra, and A. Kumar, "A review of deep learning-based human activity recognition on benchmark video datasets," Applied Artificial Intelligence, vol. 36, no. 1, p. 2093705, 2022.
10. R. Khan and D. Khan, "Smart presence detection: Harnessing wi-fi signals and machine learning with esp8266," Preprints, June 2024.
11. A. Patwal, M. Diwakar, V. Tripathi, and P. Singh, "Crowd counting analysis using deep learning: A critical review," Procedia Computer Science, vol. 218, pp. 2448–2458, 2023.
12. N. Gupta, S. K. Gupta, R. K. Pathak, V. Jain, P. Rashidi, and J. S. Suri, "Human activity recognition in artificial intelligence framework: a narrative review," Artificial intelligence review, vol. 55, no. 6, pp. 4755– 4808, 2022.
13. F. Kulsoom, S. Narejo, Z. Mehmood, H. N. Chaudhry, A. Butt, and A. K. Bashir, "A review of machine learning-based human activity recognition for diverse applications," Neural Computing and Applications, vol. 34, no. 21, pp. 18289–18324, 2022.
14. R. Raj and A. Kos, "An improved human activity recognition technique based on convolutional neural network," Scientific Reports, vol. 13, no. 1, p. 22581, 2023.
15. J. Arunnehru et al., "Deep learning-based real-world object detection and improved anomaly detection for surveillance videos," Materials Today: Proceedings, vol. 80, pp. 2911–2916, 2023.
16. D. Fuentes-Jimenez, R. Martin-Lopez, C. Losada-Gutierrez, D. Casillas- Perez, J. Macias-Guarasa, C. A. Luna, and D. Pizarro, "Dpdnet: A robust people detector using deep learning with an overhead depth camera," Expert Systems with Applications, vol. 146, p. 113168, 2020.
17. W. Guettala, A. Sayah, L. Kahloul, and A. Tibermacine, "Real time human detection by unmanned aerial vehicles," in 2022 International Symposium on iNnovative Informatics of Biskra (ISNIB), pp. 1–6, IEEE, 2022.
18. C. Zhou, D. Ren, X. Zhang, C. Yu, and L. Ju, "Human position detection based on depth camera image information in mechanical safety," Advances in Mathematical Physics, vol. 2022, no. 1, p. 9170642, 2022.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING