International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

DNN-Powered Pose Detection for Action Detection

Geetha Sree Priya Narayana[1] , N Naveen Kumar[2]
[1]Post-Graduate Student, Department of Information Technology, Software Engineering, Jawaharlal Nehru Technological University, Hyderabad, India.

[2]Professor, Department of Information Technology, Jawaharlal Nehru Technological University, Hyderabad, India.

Abstract- *This paper presents a real-time human action recognition system based on pose detection using deep neural networks. A pre-trained TensorFlow model integrated with OpenCV's DNN module is used to extract keypoints from the human body, forming a skeletal structure for action classification. Rule-based logic, derived from the spatial relationships among joints, enables recognition of actions such as standing, sitting, waving, and hands up. The system processes images, video streams, and webcam feeds without requiring additional training, making it lightweight and efficient. Its modular design allows easy extension, making it suitable for surveillance, gesture control, and human-computer interaction.*

Keywords— Human Pose Detection, Action Recognition, Deep Neural Networks, OpenCV DNN, Rule-Based Classification.

1. INTRODUCTION

The ability to accurately recognize human actions from visual data has become a critical requirement in various real-world applications, including intelligent surveillance systems, gesture-controlled interfaces, healthcare monitoring, and human-computer interaction. Action recognition enables machines to interpret human behaviour and respond accordingly, thus enhancing automation and decision-making processes in smart environments. This work presents a pose-based approach to human action recognition, utilizing deep learning techniques to detect body keypoints and applying rule-based logic to classify actions. The system uses a pre-trained TensorFlow model integrated with OpenCV's deep neural network (DNN) module to estimate human poses from input images, video streams, or real-time camera feeds. The extracted keypoints are connected to form a skeletal representation, which is then analysed based on geometric relationships to identify actions such as standing, sitting, raising hands, or waving. The primary objective of this study is to design a lightweight, efficient, and real-time human action recognition system that does not rely on large-scale datasets or high-end computational resources. By combining pose detection with rule-based classification, the system ensures low latency and high interpretability, making it suitable for edge-based and embedded applications. The significance of this work lies in its simplicity, extensibility, and practical relevance. Unlike data-driven classifiers that demand extensive training and tuning, this method offers a modular, transparent solution that can be readily adapted for specific use cases, thereby contributing to the advancement of real-time vision-based activity recognition systems.

2.1. PROJECT FEATURES

The proposed human action recognition system is designed to be lightweight, modular, and capable of performing in real-time environments. It leverages a pre-trained TensorFlow pose detection model integrated with OpenCV's DNN module to detect body keypoints from images, videos, or live webcam feeds. Based on the spatial configuration of these keypoints, the system applies interpretable rule-based logic to classify basic human actions such as standing, sitting, waving, and hands up. The architecture is modular, with separate units for input handling, pose detection, and action recognition, allowing for ease of debugging and future upgrades. The system's lightweight nature eliminates the need for GPU acceleration or large datasets, making it suitable for low-resource or embedded platforms. Furthermore, the design is scalable and supports future enhancements, such as adding new action categories or integrating learning-based classifiers to recognize more complex human behaviours.

2.2. SYSTEM ANALYSIS

Human action recognition is an essential component of computer vision applications that aim to interpret and respond to human behaviour in real-time. Traditional methods for action recognition, including background subtraction and motion-based analysis, often fall short in dynamic environments due to varying illumination, occlusion, and camera movement. Moreover, machine learning approaches that rely on large datasets and extensive training cycles can be computationally intensive and less interpretable. To overcome these limitations, pose-based action recognition offers a promising alternative. By focusing on the structural configuration of the human body rather than raw pixel or motion data, pose detection enables a more robust understanding of actions irrespective of the background or environment. Existing solutions, however, often depend on complex deep learning pipelines that require GPU resources and are difficult to interpret or debug in real-time applications. This system addresses these challenges by integrating a pre-trained TensorFlow pose detection model with OpenCV's DNN module to extract human body keypoints. Rather than using a data-driven classifier, it adopts a rule-based approach to classify actions based on geometric relationships among the keypoints. This not only ensures computational efficiency but also makes the system easier to maintain, extend, and deploy on low-power devices. By balancing accuracy, interpretability, and speed, the system is well-suited for use cases in surveillance, gesture control, healthcare monitoring, and human-computer interaction. Its modular design further allows for future enhancement by incorporating learning-based methods or additional action categories.

2.3. PROPOSED SYSTEM

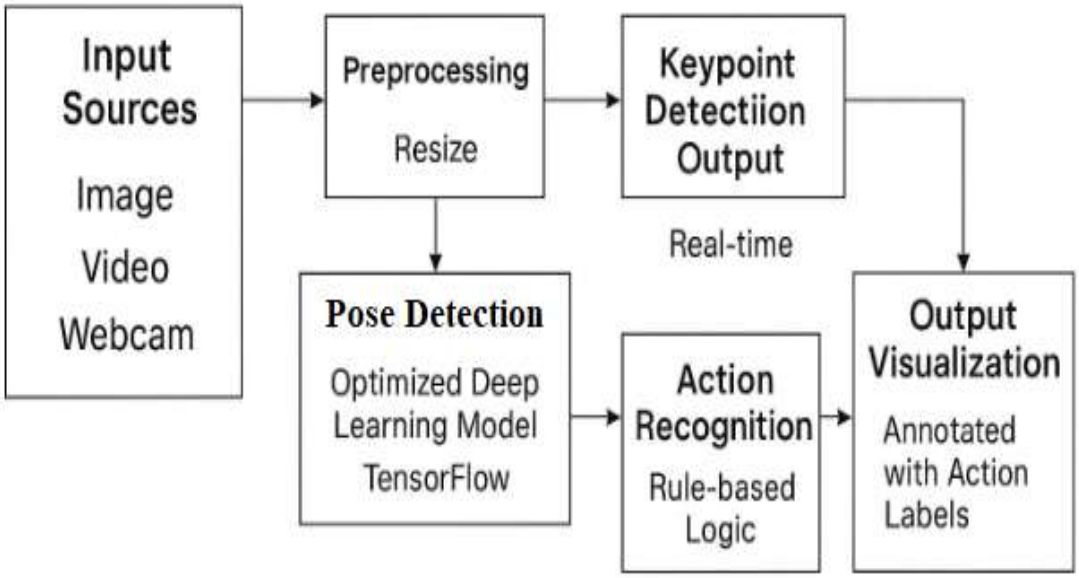
The proposed human action recognition system is designed to detect and classify human activities in real-time using pose detection techniques. It combines deep learning-based keypoint detection with rule-based action classification to provide an efficient and interpretable framework. The system is implemented using Python, OpenCV, and a pre-trained TensorFlow pose detection model, making it suitable for deployment on standard computing platforms without the need for GPU acceleration.

To process input data from static images, videos, or webcam streams in real-time. To extract human body keypoints using an optimized TensorFlow .pb model through OpenCV's DNN module. To generate a skeletal structure by connecting detected keypoints to visualize body posture. To classify actions such as Standing, Sitting, Hands Up, and Waving based on spatial relationships between key joints (e.g., wrist, shoulder, neck). To display and annotate the detected actions on output frames and enable storage or logging if required.

This rule-based approach reduces computational overhead and eliminates the need for training data, making the system lightweight and adaptable. Its modular design allows for future integration with machine learning classifiers to support more complex or ambiguous action recognition tasks.

2.4. PROJECT ARCHITECTURE

The architecture of the proposed human action recognition system is modular and optimized for real-time performance, as illustrated in Fig. 1. It consists of sequential processing stages designed to efficiently extract body pose information from input sources and infer human actions based on spatial keypoints. The system begins with the Input Sources module, which accepts three types of input: static images, video files, and live webcam feeds. These inputs are then passed to the Preprocessing unit, where each frame is resized and normalized to meet the input requirements of the pose detection model. The preprocessed frame is forwarded to the Pose Detection block, which utilizes an optimized deep learning model loaded from a pre-trained TensorFlow .pb file via OpenCV's DNN module. This model outputs a set of body Keypoints (e.g., nose, neck, elbows, wrists) in real time. Next, the Action Recognition module applies rule-based logic to the spatial arrangement of detected keypoints. This logic is designed to classify basic human actions such as standing, sitting, waving, and hands up, based on geometric relationships among joints (e.g., wrist height relative to neck or shoulders). Finally, the Output Visualization unit overlays the detected skeletal structure and predicted action label onto the frame.



.

Fig 1: Project Architecture

2.5. EVALUATION OF MODEL

Output from a static image showing pose keypoints and skeletal connections. The system correctly classifies the action as **"Standing"** based on the vertical alignment of the torso and limb keypoints.

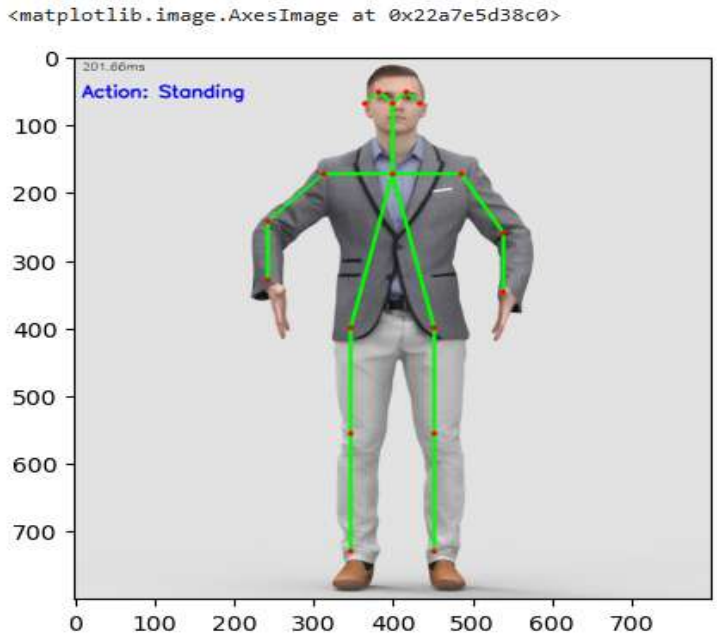


Fig2: output of image input

Output from a video input frame. The system detects human pose keypoints and correctly classifies the action as **"Sitting"** using rule-based spatial analysis.

Fig3: output of the video input

2.6. CONCLUSIONS

This paper presents a real-time human action recognition system that combines pose detection using deep neural networks with rule-based classification logic. By extracting skeletal keypoints from images, video, and webcam inputs through a pre-trained TensorFlow model, the system accurately identifies basic human actions such as standing, sitting, hands up, and waving. The use of OpenCV's DNN module ensures compatibility and performance on standard hardware without the need for GPU acceleration or large-scale training datasets.

The key contribution of this work lies in its lightweight, interpretable framework, which achieves high recognition accuracy with minimal computational requirements. The rule-based approach allows transparent decision-making and easy debugging. Despite its effectiveness, the current implementation is limited to a predefined set of actions and may not generalize well to more complex or dynamic human behaviours. Additionally, environmental variations such as occlusion, poor lighting, or fast motion may affect pose detection quality.

ACKNOWLEDGMENT

The author would like to express their sincere gratitude to Sri N. Naveen Kumar, Professor, Department of Information Technology, JNTUH, for valuable guidance and constant support throughout the duration of this work. We also acknowledge the Department of Information Technology, JNTUH, for providing the necessary facilities and resources to conduct this research.

REFERENCES

[1] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 7291–7299.

[2] A. Toshev and C. Szegedy, "Deeppose: Human pose estimation via deep neural networks," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 1653–1660.

[3] G. Papandreou, T. Zhu, N. Kanazawa, A. Toshev, J. Tompson, C. Bregler, and K. Murphy, "Towards accurate multi-person pose estimation in the wild," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4903–4911.

[4] OpenCV.org, "OpenCV: Open Source Computer Vision Library." [Online]. Available: https://opencv.org/. [Accessed: June 2025].

[5] TensorFlow, "TensorFlow: An End-to-End Open Source Machine Learning Platform." [Online]. Available: https://www.tensorflow.org/. [Accessed: June 2025].

[6] COCO Dataset, "Common Objects in Context." [Online]. Available: https://cocodataset.org/. [Accessed: June 2025].

[7] A. Addison, "How to Load a TensorFlow Model Using OpenCV," *Automatic Addison*, Jul. 2020. [Online]. Available: https://automaticaddison.com/how-to-load-a-tensorflow-model-using-opencv. [Accessed: June 2025].

[8] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 5693–5703.

[9] H. Fang, S. Xie, Y.-W. Tai, and C. Lu, "RMPE: Regional multi-person pose estimation," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2017, pp. 2334–2343.

[10] A. Kanazawa, M. J. Black, D. W. Jacobs, and J. Malik, "End-to-end recovery of human shape and pose," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 7122–7131.

[11] D. Tran, H. Wang, L. Torresani, J. Ray, Y. LeCun, and M. Paluri, "A closer look at spatiotemporal convolutions for action recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 6450–6459.

[12] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2D human pose estimation: New benchmark and state of the art analysis," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 3686–3693.